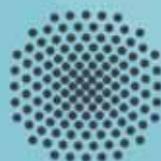




# Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V.

**Band 29**

**2020**



**Universität Stuttgart**  
Institut für Photogrammetrie

## **Vorträge**

### **40. Wissenschaftlich-Technische Jahrestagung der DGPF**

4. – 6. März 2020  
in Stuttgart

ISSN 0942-2870

Thomas P. Kersten, Hrsg.

Publikationen  
der Deutschen Gesellschaft für  
Photogrammetrie, Fernerkundung  
und Geoinformation e.V.

---



**Band 29**

**2020**

**Beiträge**

40. Wissenschaftlich-Technische  
Jahrestagung der DGPF

4. – 6. März 2020  
in Stuttgart

ISSN 0942-2870

Thomas P. Kersten, Hrsg.

ISSN 0942-2870

Publikationen der  
Deutschen Gesellschaft für Photogrammetrie, Fernerkundung  
und Geoinformation (DGPF) e.V.  
Band 29, 543 S., Hamburg 2020  
Hrsg.: Thomas P. Kersten

© Deutsche Gesellschaft für Photogrammetrie, Fernerkundung  
und Geoinformation (DGPF) e.V.  
München 2020

Geschäftsstelle der DGPF  
c/o Technische Universität München  
Institut für Geodäsie, GIS und Landmanagement  
Lehrstuhl für Geoinformatik  
Arcisstraße 21  
D-80333 München  
Tel.: 089 289-22578, E-Mail: [geschaeftsstelle@dgpf.de](mailto:geschaeftsstelle@dgpf.de)

Redaktion:

Thomas P. Kersten  
HafenCity Universität Hamburg  
Labor für Photogrammetrie & Laserscanning  
Überseeallee 16, 20457 Hamburg  
E-Mail: [Thomas.Kersten@hcu-hamburg.de](mailto:Thomas.Kersten@hcu-hamburg.de)

# INHALTSVERZEICHNIS

## FORST

- SIMON KÖNIG, JOHANNES SCHULTZ, OLENA DUBOVYK & FRANK THONFELD:  
Assessment of Drought Effects on Forests using Non-Parametric Methods  
and Satellite Imagery. .... 1
- TILMAN BUCHER, MATTHIAS GESSNER, DANIEL HEIN, HENRY MEIßNER,  
THOMAS KRAFT, AGNIESZKA SOSZYŃSKA & SEBASTIAN PLESS:  
Case study: Near real-time thermal mapping to support firefighting and crisis  
management. .... 13

## SENSOREN - FLUGZEUGE/SATELLITEN

- K. JACOBSEN & ROLF RÜDIGER LUDWIG:  
CORONA-Höhenmodelle und Orthofotos. .... 22
- MICHAEL CRAMER, SHUHANG ZHANG, HENRY MEIßNER & RALF REULKE:  
Quality assessment of high-resolution UAV imagery and products. .... 33
- HEINZ-JÜRGEN PRZYBILLA & MANFRED BÄUMKER:  
Untersuchungen zur Qualität des Realtime Kinematic GNSS Systems  
der DJI Phantom 4 RTK. .... 47

## METHODIK RADAR/LIDAR

- ROBERT WEIß & GOTTFRIED MANDLBURGER:  
Erprobung des Leica SPL100 Laserscanners zur Erfassung der  
Wasseroberfläche an deutschen Binnengewässern. .... 62
- SYLVIA SCHMITZ, MARTIN WEINMANN, UWE WEIDNER, HORST HAMMER & ANTJE THIELE:  
Automatic Generation of Training Data for Land Use and  
Land Cover Classification by Fusing Heterogeneous Data Sets. .... 73
- NICO ADAM, RALF REULKE & VOLKER SPRECKELS:  
Neues Normungsverfahren „InSAR – Radarinterferometrie  
für die Bodenbewegungserfassung“. .... 87

## **VIRTUAL REALITY + AUGMENTED REALITY**

THOMAS P. KERSTEN & ALEXANDER P. WALMSLEY: Entwicklung einer interaktiven, immersiven Virtual-Reality-Applikation des historischen Stadtmodells Stade 1620. ....	99
FLORIAN HRUBY & RAINER RESSL: Satellitenbild-basierte Geovisualisierung mittels immersiver VR. ....	114
CLAUDIA LINDNER, JOHANNES SCHULTZ, ANNETTE ORTWEIN, KILIAN STAAR, HENRYK HODAM, CARSTEN JÜRGENS, FABIAN SELG & ANDREAS RIENOW: Herausforderungen bei der Integration von 3D: Daten, Devices und Dimensionen im AR-VR-Kontinuum. ....	126

## **GEOLOGIE UND HYDROLOGIE**

KATJA RICHTER, DAVID MADER, PATRICK WESTFELD & HANS-GERD MAAS: Analyse der Auswirkung der Wasseroberflächenmodellierung auf die Genauigkeit von Gewässerbodenpunkten in der Laserbathymetrie. ....	143
HANNES NÜBEL, GOTTFRIED MANDLBURGER, MICHAEL KÖLLE & UWE SÖRGEL: Bathymetrieableitung aus multispektralen Luftbildern über Convolutional Neural Networks. ....	156

## **GRÜNLAND**

MICHAEL WACHENDORF, THOMAS ASTOR & JAYAN WIJESINGHA: Remotely sensed information for the protection and management of species-rich grasslands. ....	172
ANDREAS SCHAUMBERGER & MANUEL ADELWÖHRER: SpectroAnalyst – Management von Hyperspektraldaten für die Grünlandforschung. ....	184

## **STANDARDISIERUNG**

W. KRESSE: Neue ISO-Norm für Langzeitarchivierung von Fernerkundungsdaten. ....	197
--	-----

## **BILDANALYSE UND COMPUTER VISION**

DOMINIK LAUPHEIMER, MOHAMAD HAKAM SHAMS EDDIN & NORBERT HAALA: The Importance of Radiometric Feature Quality for Semantic Mesh Segmentation. ....	205
JOHANNES WOLF, RICO RICHTER & JÜRGEN DÖLLNER: Verwendung künstlicher neuronaler Netzwerke zur Bilderkennung für die Analyse von Straßenmarkierungen in Mobile-Mapping-3D-Punktwolken. ....	219

## **UMWELTANALYSE**

- GÖRRES GRENZDÖRFFER:  
Ermittlung urbaner Überwärmungsbereiche auf der Basis  
multitemporaler Satellitendaten am Beispiel der Hansestadt Rostock. .... 230
- MICHAEL H. WIMMER, MARKUS HOLLAUS, GOTTFRIED MANDLBURGER,  
ANDREAS BUTTINGER-KREUZHUBER, JÜRGEN WASER, JÜRGEN KOMMA  
& NORBERT PFEIFER:  
Prozessierung bundesweiter Geodaten für hydraulische Modellierung. .... 242

## **MOBILE MAPPING/BILDANALYSE UND COMPUTER VISION**

- JONAS MEYER, STEFAN BLASER, STEPHAN NEBIKER, LEA FRICKER & DANIEL WEBER:  
Leistungsfähiges portables Mobile Mapping im Außenraum –  
Genauigkeitspotential und Grenzen unterschiedlicher Georeferenzierungs-  
ansätze in der Stadt und im Wald. .... 254
- ARASH JAVANMARD-GH, CHARLES TOTH & DOROTA IWASZCZUK:  
Seamless Outdoor and Indoor Mapping using a LiDAR-based  
Multi-Sensor System: Case Study on Kalman-Filter Fusion. .... 270
- LINA E. BUDDE, STEFAN SCHMOHL & U. SÖRDEL:  
Unsicherheitsauswertung von semantischer Segmentierung  
mittels Neuronaler Netze. .... 280

## **METHODIK GEOINFORMATIK**

- HEINRICH LOREI, BERNHARD HÖFLE & RENE WESTERHOLT:  
Spatial Structure as an Element of Motivation in Location- Based Games. .. 290
- MICHAEL KÖLLE, VOLKER WALTER, STEFAN SCHMOHL & UWE SÖRDEL:  
Evaluierung der Leistungsfähigkeit der Crowd für das Labeln  
von 3D-Punktwolken im Kontext von Active Learning. .... 299
- KATHARINA ANDERS, LUKAS WINIWARTER, HUBERT MARA,  
RODERIK LINDENBERGH, SANDER E. VOS & BERNHARD HÖFLE:  
Einfluss der zeitlichen Auflösung auf die raumzeitliche Segmentierung  
geomorphologischer Änderungsprozesse in 3D-Punktwolken. .... 312
- MIKE SIPS, VALERY LAVRENTIEV, DANIEL EGGERT & JULIA NEELMEIJER:  
Assessing automatically-detected changes in the post-classification  
change-detection of Sentinel-2 data with Visual Analytics. .... 317

## **GEOLOGIE UND HYDROLOGIE / METHODIK RADAR/LIDAR**

- GOTTFRIED MANDLBURGER, MARTIN PFENNIGBAUER & ROLAND SCHWARZ:  
Genauigkeits- und Performanceuntersuchung eines neuen  
bathymetrischen UAV-Laserscanners. .... 321
- GOTTFRIED MANDLBURGER:  
Möglichkeiten und Grenzen der Bathymetrieableitung  
mittels Single Photon LiDAR. .... 335

## **OPTISCHE 3D MESSTECHNIK**

- ANSGAR GREIWE, RAINER BRECHTKEN, MICHAEL LÖSLER, CORNELIA ESCHELBACH  
& RÜDIGER HAAS:  
Erfassung der Hauptreflektordeformation eines Radioteleskops  
durch UAV-gestützte Nahbereichsphotogrammetrie. .... 346
- LARS S. OBROCK & EBERHARD GÜLCH:  
Automatisierte semantische Modellierung von Innenräumen aus Bildern  
und abgeleiteten Punktwolken basierend auf Deep Learning Methoden. .... 358
- LUKAS WINIWARTER, KATHARINA ANDERS & BERNHARD HÖFLE:  
Herausforderungen in der Fehlerfortpflanzung von Laserscandaten  
für multitemporale Analysen zur verbesserten Quantifizierung  
des Level of Detection. .... 373
- AGNES R. DINKEL, LUDWIG HOEGNER, ADRIAN EMMERT , LUKAS RAFFL & UWE STILLA:  
Änderungsdetektion in photogrammetrischen Punktwolken  
für das Monitoring hochalpiner, gravitativer Massenbewegungen –  
Beispiel Hochvogel. .... 381

## **AK AR/VR UND AK AUS- UND WEITERBILDUNG**

- MICHAEL WEINMANN, PATRICK STOTKO, STEFAN KRUMPEN & REINHARD KLEIN:  
Immersive VR-based Live Telepresence for Remote Collaboration  
and Teleoperation. .... 391
- MICHAEL MAYER, KATRIN HEß & JAN CERMAK:  
Gelingendes Zusammenspiel von Einzelmaßnahmen zur Weiterentwicklung  
des Lehr-Lernsystems im BSc-Studiengang „Geodäsie und Geoinformatik“  
am Karlsruher Institut für Technologie (KIT). .... 400
- JÜRGEN SELENSKI, VALENTIN WICH & ANSGAR BRUNN:  
Erfahrungen mit einem Regelbetrieb der Erstellung von  
E-Learning-Angeboten für Photogrammetrie und Mobile 3D-Vermessung. .. 411

## **MULTI- UND HYPERSPEKTRAL**

- MARTIN WEINMANN, FELIX EHMER & UWE WEIDNER:  
Analyse von multi-temporalen Sentinel-2-Daten zur Klassifizierung  
von Landbedeckung und Landnutzung am Kaiserstuhl. .... 422

## **3D-STADTMODELLE**

- VLADETA STOJANOVIC, MATTHIAS TRAPP, RICO RICHTER & JÜRGEN DÖLLNER:  
Comparison of Deep-Learning Classification Approaches  
for Indoor Point Clouds. .... 437

## **METHODIK RADAR/LIDAR**

- PHILIPP J. SCHNEIDER & UWE SÖRGEL:  
Monitoring einer Schleuse mittels  
satellitengestützter DInSAR-Techniken. .... 448
- TOBIAS BOLZ, SYLVIA SCHMITZ, ANTJE THIELE, UWE SOERGEL & STEFAN HINZ:  
Analysis of Airborne SAR and InSAR Data for Coastal Monitoring. .... 457

## **POSTER**

- J.-H. BRAATZ-NEBEL, H. GERIGHAUSEN & H. LILIENTHAL:  
Methodische Untersuchungen zum Einsatz von Copernicus Satellitendaten  
für das Monitoring im Rahmen der EU Agrarförderung. .... 462

## **BEITRÄGE KARL-KRAUS-NACHWUCHSPREIS**

- LAURA CROCETTI & WOUTER DORIGO:  
Impacts of Climatic Oscillations on Precipitation in  
an Extended Mediterranean Area. .... 469
- YIFEI YIN, MICHAEL KÖLLE & VOLKER WALTER:  
Entwicklung und Evaluierung eines Web-Tools zur bezahlten  
crowd-basierten Erfassung von Bäumen aus 3D-Punktwolken. .... 478
- ANNIKA BONERATH, BENJAMIN NIEDERMANN & JAN-HENRIK HAUNERT:  
Dynamic Aggregation of Geo-Objects for the  
Interactive Exploration of Research Data. .... 488
- DANIEL SCHERER, CHRISTIAN SCHWATKE & PETER KRZYSZEK:  
Estimation of River Discharge Using Satellite Altimetry  
and Optical Remote Sensing Images. .... 497
- CLAUDIO NAVACCHI, BERNHARD BAUER-MARSCHALLINGER & WOLFGANG WAGNER:  
Monitoring of Alpine Snow Conditions Using C-Band SAR. .... 505
- DOMINIK TRAU, THOMAS P. KERSTEN & FELIX TSCHIRSCHWITZ:  
Die Viermastbark Peking in Virtual Reality als neue Form  
der Wissensvermittlung. .... 514
- ARIANE DROIN, MICHAEL WURM & WOLFGANG SULZER:  
Semantic labelling of building types. A comparison of two approaches  
using Random Forest and Deep Learning. .... 527

**AUTORENVERZEICHNIS** ..... 539

**PROGRAMMKOMITEE** ..... 543

# Assessment of Drought Effects on Forests using Non-Parametric Methods and Satellite Imagery

SIMON KÖNIG<sup>1</sup>, JOHANNES SCHULTZ<sup>1,2</sup>, OLENA DUBOVYK<sup>1,3</sup> & FRANK THONFELD<sup>4,5</sup>

*Abstract: Temperate forests are a vital part of the natural environment and important for human existence. Yet, they are increasingly threatened by changing disturbance regimes, especially more severe droughts. Here, we combine Landsat and Sentinel-2 time series and use a kernel density estimation (KDE) to assess the baseline phenology of a mixed forest area in Germany. We compared this baseline to the forest's development during the drought year of 2018 and the following growing season of 2019. Our results show that parts of the examined forest were negatively impacted to an unprecedented degree. With the use of satellite remote sensing analyses and the KDE, we can assess the impact of drought on vegetation development without relying on parametric methods.*

## 1 Introduction

Forests are a key component of the natural environment in tropical, temperate and boreal regions as well as a major foundation for human existence in these areas. Locally and regionally, they provide a variety of ecosystem services. This includes the provision of wood-related products, the filtering of water or regulation of climate as well as the potential for recreation and tourism. In Europe, forestry is among the most important economic sectors. Besides being biodiversity hosts, forests are an integral part of global land-atmosphere interactions (BONAN 2008). Temperate forests serve as major net carbon sinks. Contrasting tropical and boreal forests, the amount of carbon stored inside the temperate forest biome increased between 2000 and 2007 compared to the decade before (PAN et al. 2011), partly outweighing anthropogenic CO<sub>2</sub> emissions. Consequently, temperate forests play a key role in mitigating climate change.

Natural disturbances, such as windthrow, fires and insect outbreaks, are crucial for ecosystem functions and the dynamics of forests. As they change the composition, structure and function of forests, they contribute to their heterogeneity, facilitate biodiversity and stimulate succession, reorganization and renewal (SEIDL et al. 2017). During the last decades however, disturbance regimes have changed globally. For many regions, disturbances are increasingly prevalent and becoming more frequent and severe. This includes fires, insect outbreaks and droughts. So-called

---

<sup>1</sup> Rheinische Friedrich-Wilhelms-Universität Bonn, Geographisches Institut, Meckenheimer Allee 166, D-53115 Bonn, E-Mail: [simonkoenig, odubovyk]@uni-bonn.de

<sup>2</sup> Ruhr-Universität Bochum, Geographisches Institut, Universitätsstraße 150, D-44801 Bochum, E-Mail: johannes.schultz@rub.de

<sup>3</sup> Rheinische Friedrich-Wilhelms-Universität Bonn, Zentrum für Fernerkundung der Landoberfläche, Genscherallee 3, D-53113 Bonn

<sup>4</sup> Deutsches Fernerkundungsdatenzentrum, Münchner Straße 20, 82234 Wessling, E-Mail: frank.thonfeld@dlr.de

<sup>5</sup> Julius-Maximilians-Universität Würzburg, Institut für Geographie und Geologie, Oswald-Külpe-Weg 86, D-97074 Würzburg

‘megadisturbances’ (MILLAR & STEPHENSON 2015) are emerging, having the potential to increase tree mortality to a level surpassing historical recordings and alter forests beyond their ecological resilience. Consequently, there are increasing rates of canopy mortality being observed across Europe’s temperate forests (SENF et al. 2018). Many of these changes in disturbance regimes are attributed to climate change – SEIDL et al. (2017) state that temperature and water availability are the most important drivers of disturbances in the temperate forest biome. There is strong evidence that disturbance activity is going to increase with both warmer and wetter as well as warmer and drier climatic conditions; and that climate change will facilitate large-scale disturbance episodes across Europe’s temperate forests (SENF & SEIDL 2018).

In this context, drought is an especially important disturbance agent. Climate observations and projections suggest that droughts are increasing in both more frequency and magnitude in the wake of a generally hotter climate in Europe. ALLEN et al. 2015 and MILLAR & STEPHENSON 2015 refer to these emerging, more severe droughts that are coupled with high temperature events as ‘hotter droughts’ or ‘global-change-type droughts’. These affect forests both directly and indirectly. Their direct effects include the increase of evaporation and heat damage due to higher temperatures, with a simultaneous increase in water stress due to a lack of precipitation. Additionally, a generally warmer climate leads to less precipitation in the form of snow that serves as a water reservoir (MILLAR & STEPHENSON 2015). Indirectly, droughts increase the susceptibility of trees to other disturbances, especially pathogens, fires and insect outbreaks while facilitating the activity of insects and intensify the risk of fires – there is strong evidence that an increase in droughts also leads to increased fire and insect disturbances (ALLEN et al. 2015; SEIDL et al. 2017).

As an example event to assess drought effects on forests, we use the major drought that took place in large parts of Central, Northern, and Western Europe in the summer of 2018. The drought was characterized by considerable rainfall lacks, combined with exceptionally high temperatures, especially during June and July. Hence, this event can probably be characterized as one of the global-change-type droughts mentioned above. Summer temperatures in Germany were only exceeded by temperatures in 2003, and the lack of rainfall in some areas was unprecedented in the historical recordings (IMBERY et al. 2018). The drought had severe consequences for vegetation health and crop yield (REINERMANN et al. 2019) and especially forests, with widespread low remote sensing vegetation index values and a strongly reduced vegetation productivity. Increased tree mortality during 2018 and 2019 is also reported, but partly may also be attributed to legacy effects and not only the drought itself (BURAS et al. 2019).

Remote sensing has been a widely used tool to appraise the impacts of drought events on ecosystems for decades (AGHAKOUCHAK et al. 2015). However, the impacts of droughts on forests have often been underestimated (ALLEN et al. 2015) and there are various challenges to the remote sensing of these effects, including the choice of proper datasets with regards to resolution and coverage as well as the appropriate assessment method (ZHANG et al. 2013). It is feasible to compare the impacts of droughts or other disturbances to forest’s typical phenology to get a sense of how severe it is. These comparisons however have often been based on parametric methods that may not actually well explain forest’s phenological cycles (CHÁVEZ et al. 2019). Additionally, there is a need for approaches that allow a near-realtime monitoring of drought events (ALLEN et al. 2015) as well as working with multi-sensor data to be more resilient towards cloudy pixels (ZHANG et al. 2013). Furthermore, some drought effects exhibit delays and take place in the years following the drought

events, so it is often sensible to consider a study period of multiple years. Here, we try to meet these requirements by estimating the impact of the drought of 2018 in on a mixed forest in Germany. This is done in a non-parametric way that allows for the comparison to a baseline phenology. We use all data from the Landsat and Sentinel-2 sensors available since the year 2000 and compare the forests' development for the years 2018 and 2019 to their phenological baseline, computed using a kernel density estimation (KDE).

## 2 Data & Methods

### 2.1 Study Area

The proposed procedure of this study is tested for a mixed forest area in Germany: the *Kottenforst* area (Fig. 1a), located directly in the south west of the city of Bonn in the state of North-Rhine-Westphalia, is a cohesive forest area of about 40 km<sup>2</sup> size. Though there are a variety of tree species abundant, *Quercus robur* L. (Common oak), *Quercus petraea* MATT. (Sessile Oak), *Picea abies* L. (Norway Spruce) and *Fagus sylvatica* L. (European Beech) prevail in most areas (BOSCH 1997). The area is relatively humid, typically receiving around 650 mm of rainfall annually (WACHTER 2009). The forest is protected to a diverging degree. Namely, large areas are strictly protected, whereas some allow for forestry to some degree and contain plantations.

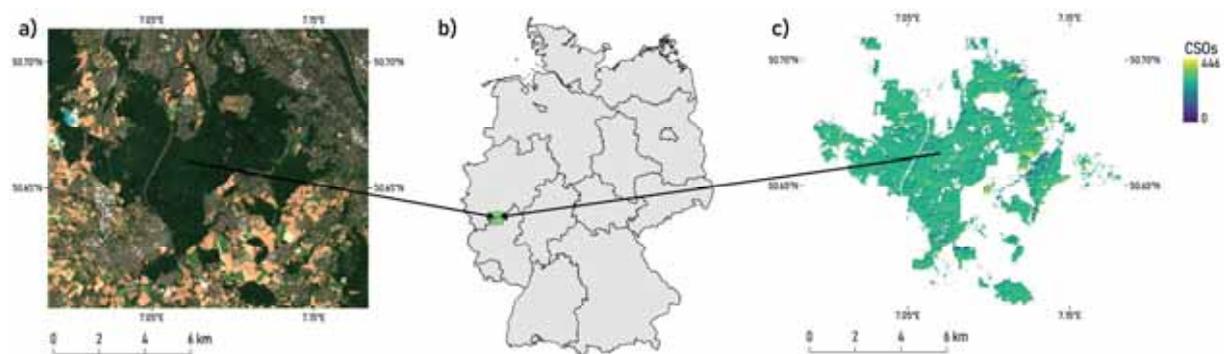


Fig. 1: Study area Kottenforst. A cloud-free Sentinel-2 composite (a) and the position of the study area inside Germany are displayed (b). In b), the green box indicates the tile of the common Landsat/Sentinel-2 gridding system in which the area is located. c) displays the forest pixels of the area that did not experience a stand-replacing disturbance between 2000 and 2017 with the number of clear sky observations (CSOs), i.e. the number of pixels that were not contaminated by clouds, snow, aerosols etc. based on each image's quality flags

### 2.2 Data

For this study, data from different moderate-resolution multispectral sensors were used. This includes Landsat TM, ETM+ and OLI as well as Sentinel-2 MSI imagery. With their similar spectral characteristics, they can be used as a 'virtual constellation' (WULDER et al. 2015) whose data can be combined to create a denser time series and be more independent from cloud coverage. We regarded all available imagery of these sensors from January 01, 2000 to September 30, 2019. Data before the year 2000 was not included for two reasons: first, the shorter period reduced the prevalence of trends in the data and thus reduces their effect on the fitting of the baseline phenology. Second, we used secondary data to remove areas that are not suitable for this study whose

data range however only includes the period since 2000. In total, we acquired 1823 images: 927 Landsat and 896 Sentinel-2 scenes.

### 2.3 Methods

The multi-sensor satellite imagery had to be properly radiometrically corrected to ensure a sensible inter-calibration between the sensors. Here, we used the *Framework for Operational Radiometric Correction for Environmental monitoring* (FORCE; FRANTZ 2019). FORCE can generate analysis-ready data (ARD) from both Sentinel-2 MSI as well as Landsat TM, ETM+ and OLI data by applying one consistent pre-processing scheme (FRANTZ et al. 2016). We applied the FORCE built-in topographic, atmospheric and BRDF corrections. Both Landsat and Sentinel-2 data were projected into one common, tiled grid to ease further processing; whereby Sentinel-2 was resampled to the 30 m resolution that matches the Landsat sensors. During processing, the FORCE software also created a variety of quality flags for each image, including a cloud and cloud shadow mask based on the *Fmask* algorithm (ZHU et al. 2015). However, the algorithm implemented in FORCE is adjusted for the lack of a thermal infrared band in the Sentinel-2 data – it considers parallax effects in the imagery to better detect clouds (FRANTZ et al. 2018). The output of the FORCE preprocessing steps was one consistent data cube of ARD at 30 m resolution, with cloudy and otherwise contaminated pixels being excluded and both Landsat and Sentinel-2 imagery projected into one common gridding system (FRANTZ 2019).

To ensure that only forest pixels were considered that did not experience any stand-replacing disturbance which would distort the fitting of the baseline, we applied one final preprocessing step. Using the global forest change data by HANSEN et al. (2013) implemented in Google Earth Engine (GEE; GORELICK et al., 2017), we generated a mask of all pixels that were classified as forest in 2000 and were not regarded as ‘lost’ until 2017. This mask was exported from GEE and applied to the ARD output of FORCE. The utilization of this procedure limited the suitable Landsat data to the period since 2000. From the resulting data we computed two multispectral indices to assess the drought effects: first, the Normalized Difference Vegetation Index (NDVI) as the most common remote sensing index that is sensible to plant health was regarded, which calculates as follows:

$$NDVI = \frac{NIR-Red}{NIR+Red} \quad (1)$$

where *NIR* corresponds to the near infrared and *Red* to the red band in the imagery. The second used index, the Normalized Difference Moisture Index (NDMI) is more sensitive to leaf water content and consequently especially suitable to detect the effects of droughts (cf. ZHANG et al. 2013). It is calculated similarly to the NDVI:

$$NDMI = \frac{NIR-SWIR1}{NIR+SWIR1} \quad (2)$$

where *SWIR1* corresponds to the first shortwave infrared band of both Landsat and Sentinel-2. The NDVI and NDMI data were used as the input for the drought-related analysis itself, i.e. the fitting of a phenological baseline of the years 2000-2017 and the comparison of the drought year of 2018 and the follow-up year of 2019 to this baseline. The fitting of the phenological baseline is done with a non-parametric procedure by using a bivariate kernel density estimation (KDE). This approach was first introduced by CHÁVEZ et al. (2019) to assess insect defoliation in deciduous

forests in South America based on MODIS Enhanced Vegetation Index (EVI) imagery. It was implemented into the `npphen` package (CHÁVEZ et al. 2017) for the R programming language (R CORE TEAM 2019) and modified for this study.

The procedure is graphically documented in Fig. 2 and consisted of five distinct steps: First, each pixel's NDVI/NDMI time series was divided into a baseline period – the years of 2000 to 2017 – and the target period in which the drought and its immediate aftermath took place – January 01, 2018 to September 30, 2019 (Fig. 2a). The dates of the baseline period were then converted to Day of Year (DOY) values, resulting in a transformed feature space where DOY is the independent and NDVI or NDMI, respectively, is the dependent variable (Fig. 2b). Mathematically, the feature space can be described as  $X_i = (DOY_i; IND_i)^T, i = 1, \dots, n$  where  $IND$  is the respective remote sensing index (ESTAY & CHÁVEZ 2018). This feature space served as the input for the KDE. The goal of the KDE is now to approximate the probability density function  $f(x)$ :

$$\hat{f}(x; H) = \frac{1}{n} \sum_{i=1}^n K_H(x - X_i) \quad (3)$$

where  $x$  corresponds to a generic point in the feature space  $X_i$ .  $K$  denotes the used kernel, which is a Gaussian kernel in this case (DUONG et al. 2019); and  $H$  is the estimated smoothing bandwidth that serves as an input for the kernel (WAND & JONES 1995, p. 91). The resulting density estimate (Fig. 2c) is standardized to sum 1 and consequently transformed into a 2-dimensional probability distribution, allowing for the derivation of confidence intervals around the most probable index value per DOY. This most probable index value forms the phenological baseline for each year (Fig. 2d). When the phenological baseline is compared to the observations of the target years (likewise converted to DOY-Index value pairs), these confidence intervals can be used to assess whether the observed values are actually a ‘true’ anomaly (Chávez et al., 2019). In addition to the difference in index values between baseline phenology and actual observations, the likelihood of these differences being an anomaly is an additional output (Fig. 2e). While the whole procedure is exemplified here for one pixel, it was applied to all forest pixels without stand-replacing disturbance between 2000 and 2017 in the Kottenforst area. Thereby, the raster-based function of the `npphen` package returns data cubes of the difference in index values between the fitted baseline and the observed values.

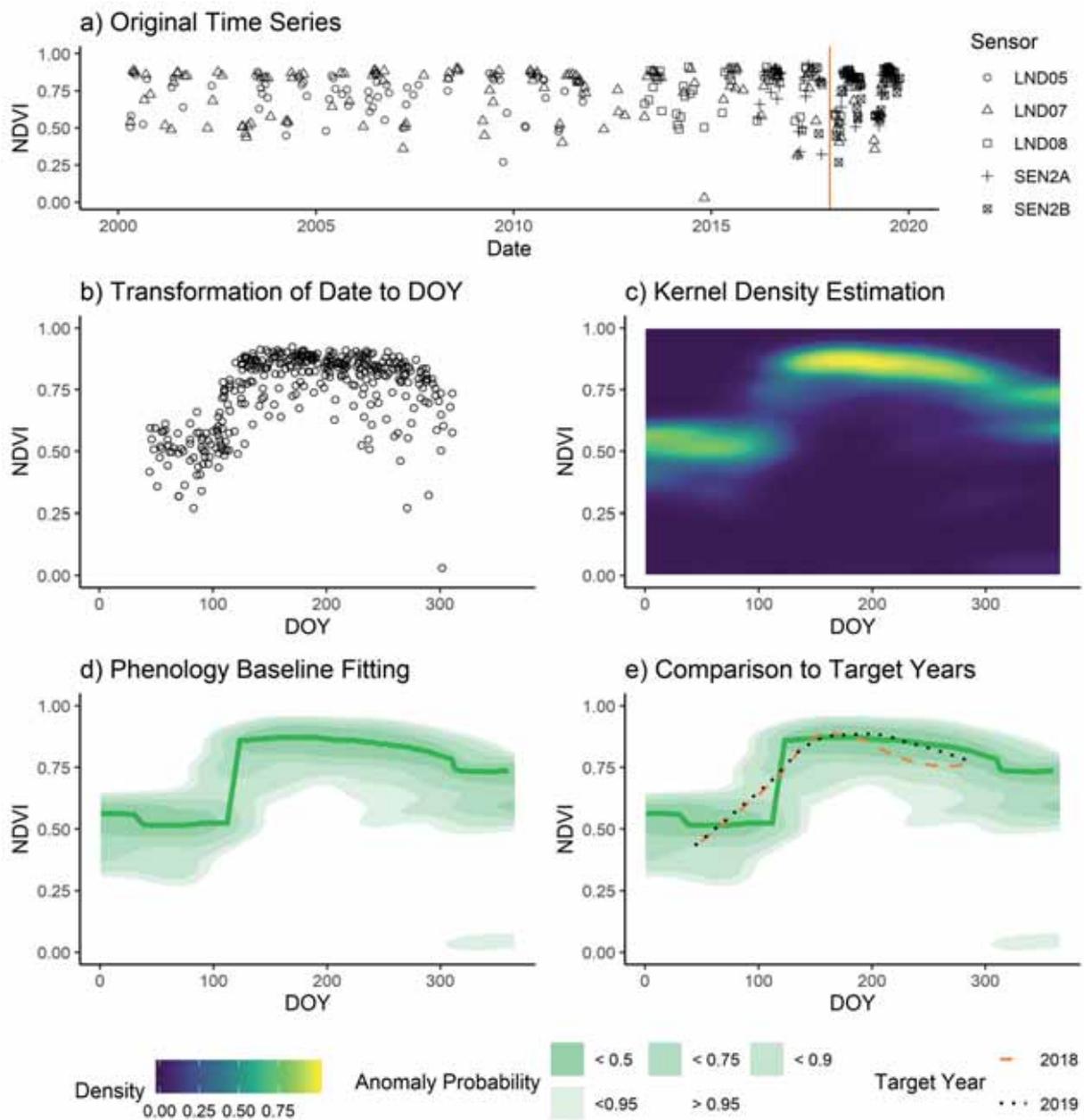


Fig. 2: Analysis procedure exemplified for the NDVI values of one pixel in the Kottenforst area. The original time series (a) is divided into a baseline and a target period, as indicated by the orange line. The values of the baseline period are transformed into a DOY-NDVI feature space (b), for which a KDE is applied (c). From this KDE, the phenological baseline (dark green line) as well as the probability of an observation being an anomaly (light green shades) can be derived (d). Both can then be compared to the target year observations, likewise converted to DOY-NDVI values (e)

### 3 Results & Discussion

As there were two years being observed using two different vegetation health indices, a total of four data cubes were output. Fig. 3 illustrates the difference between baseline and observation for one example timestep of 2018 and one of 2019 for both indices. Fig. 4 depicts the complete vegetation development of 2018 and 2019 compared to the phenological baseline for one pixel.

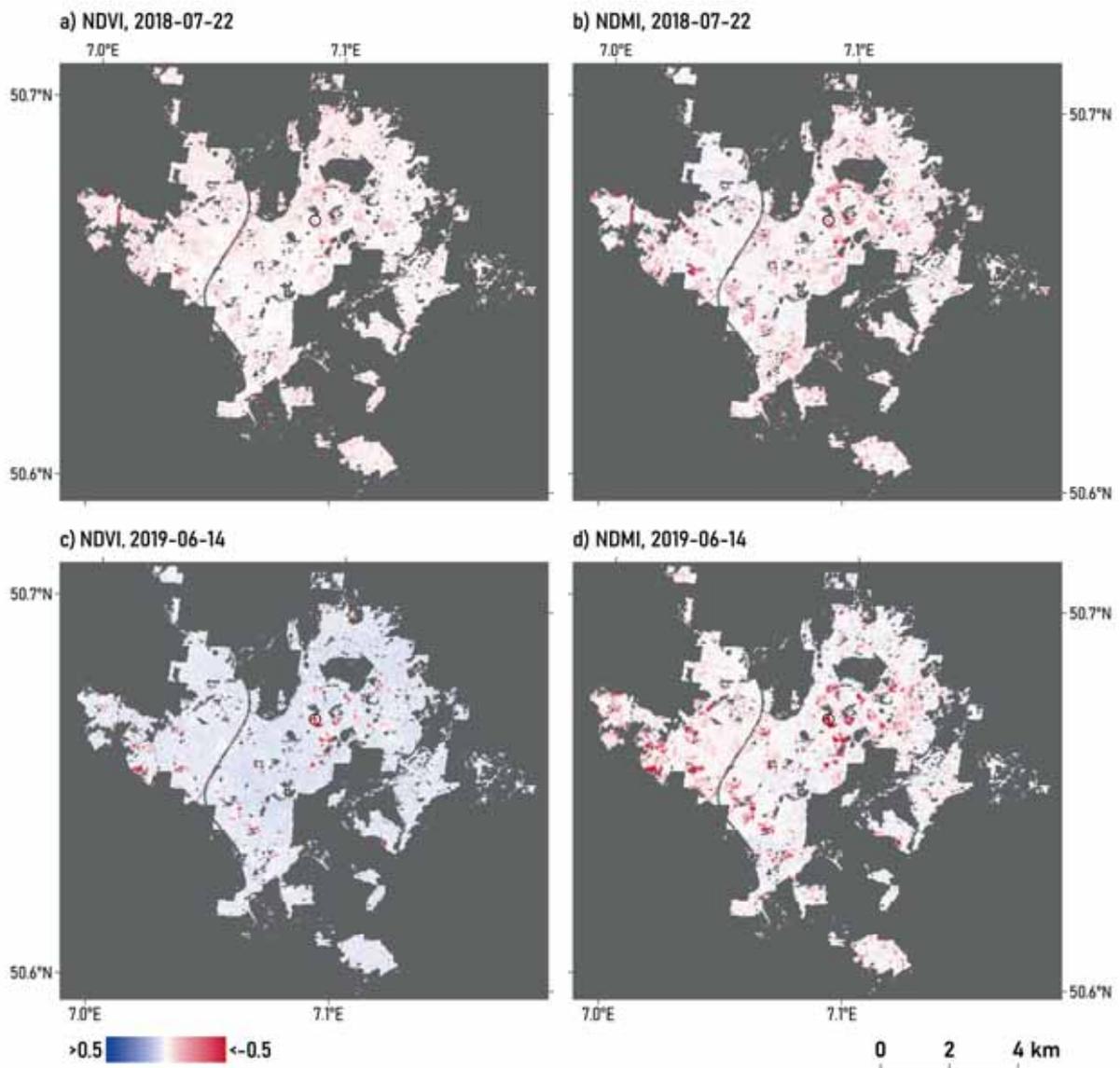


Fig. 3: Example results of the Kottenforst area for one day in 2018 (a, b) and 2019 (c, d), both chosen due to low cloud coverage. NDVI (a, c) and NDMI (b, d) differences from the baseline are displayed for all available pixels. White indicates a difference of zero, blueish colors illustrate positive deviations, red indicates negative differences. Grey areas are either cloud-contaminated, not covered by forest or did experience a stand-replacing disturbance during the reference period. The black circle is centered around an example pixel (see Fig. 4)

As Fig. 3 depicts, the drought did indeed influence the two examined forests, not only in the drought year of 2018, but also during the following vegetation period in 2019. In the Kottenforst area, NDVI values did not differ much from their phenological baseline during the peak of the drought at the end of July (Fig. 3a) but were still lower typically. In terms of the NDMI, negative differences were not extreme, but still visible, indicating a reduced leaf water content (Fig. 3b). By mid-2019, NDVI values were restored to baseline values in most areas, even being higher than regular in some. Yet, there are several small patches with very high negative anomalies – the negative difference in NDVI values between baseline and observation typically amounted to 0.5 or greater (Fig. 3c). At the same time, during mid-June of 2019, NDMI values were lower for large parts of the forest, implying a longer-term consequence of the drought: a reduced leaf water content even if the NDVI values are regular or above in the consecutive growing season. This may possibly be caused by an ongoing lack of soil moisture.

As the example pixel marked in Fig. 3 is part of a Norway spruce plantation, it did not experience much yearly phenological variation during the reference period, as illustrated by the bold green lines in Fig. 4a and 4b. Yet there has been some general variability, indicated by the relatively wide confidence intervals, displayed as lighter green shades. NDVI and especially NDMI values for 2018 were below the baseline, especially during the second half of the year. In 2019, both indices diminished and the probability of the respective values being an anomaly is very high, with NDMI exceeding 0.95 in some parts of the year. The plantation suffered from water stress and bark beetle attacks in 2018 and was thus clear-cut in 2019, which likely explains these results.

The procedure applied here has both advantages and limitations. A strong advantage is its non-parametric nature. The approach neither relies on any thresholds nor predefined functions, e.g. harmonic functions (ZHU et al. 2012) to represent seasonality. Instead, it uses the observed vegetation community and its dynamics itself (cf. ESTAY & CHÁVEZ 2018). The system is consequently easily applicable to different ecosystems. Beneficial is also its high robustness to missing and faulty values as well as outliers. This is visible in Fig. 4b where there apparently were some outliers that caused the ribbons in the bottom center. These however had no influence on the fitted phenology baseline. In contrast to other methods that have similar targets, this procedure not only quantifies anomalies but also clarifies the degree of deviation from the regular phenology in terms of confidence intervals. This is highly important, especially if in the case of a drought event, management measures have to be planned and applied. Using the medium-resolution Landsat and Sentinel-2 sensors results in an increased spatial precision compared to coarser-resolution sensors like MODIS, while the combination of both leads to a temporal resolution that is still comparably high. Thus, the procedure meets several of the demands to forest drought impact assessments mentioned in the introduction.

Among the limitations of our approach is the fact that, while the alterations of vegetation development during a drought year can be assessed, there is no direct proof that these alterations were actually caused by the drought. Possible biases in this context include the interplays of drought with other disturbances and legacy effects of past influences of the forest (BURAS et al. 2019; SEIDL et al. 2017). Furthermore, outliers in the data, as exemplified by Fig. 4b, often may not influence the phenological baseline, but its confidence intervals. Including proper measures in our workflow to remove these outliers and compute more accurate confidence intervals would be a meaningful

enhancement. An additional improvement may be the application and comparison of different remote sensing indices. Using the Vegetation Condition Index for example that scales each pixel's NDVI time series with its respective maximum and minimum and that is commonly applied in drought impact assessments (AGHAKOUCHAK et al. 2015; KOGAN 1990) or the Disturbance Index (HEALEY et al. 2005), commonly used in forest disturbance research, may further increase the validity of the results.

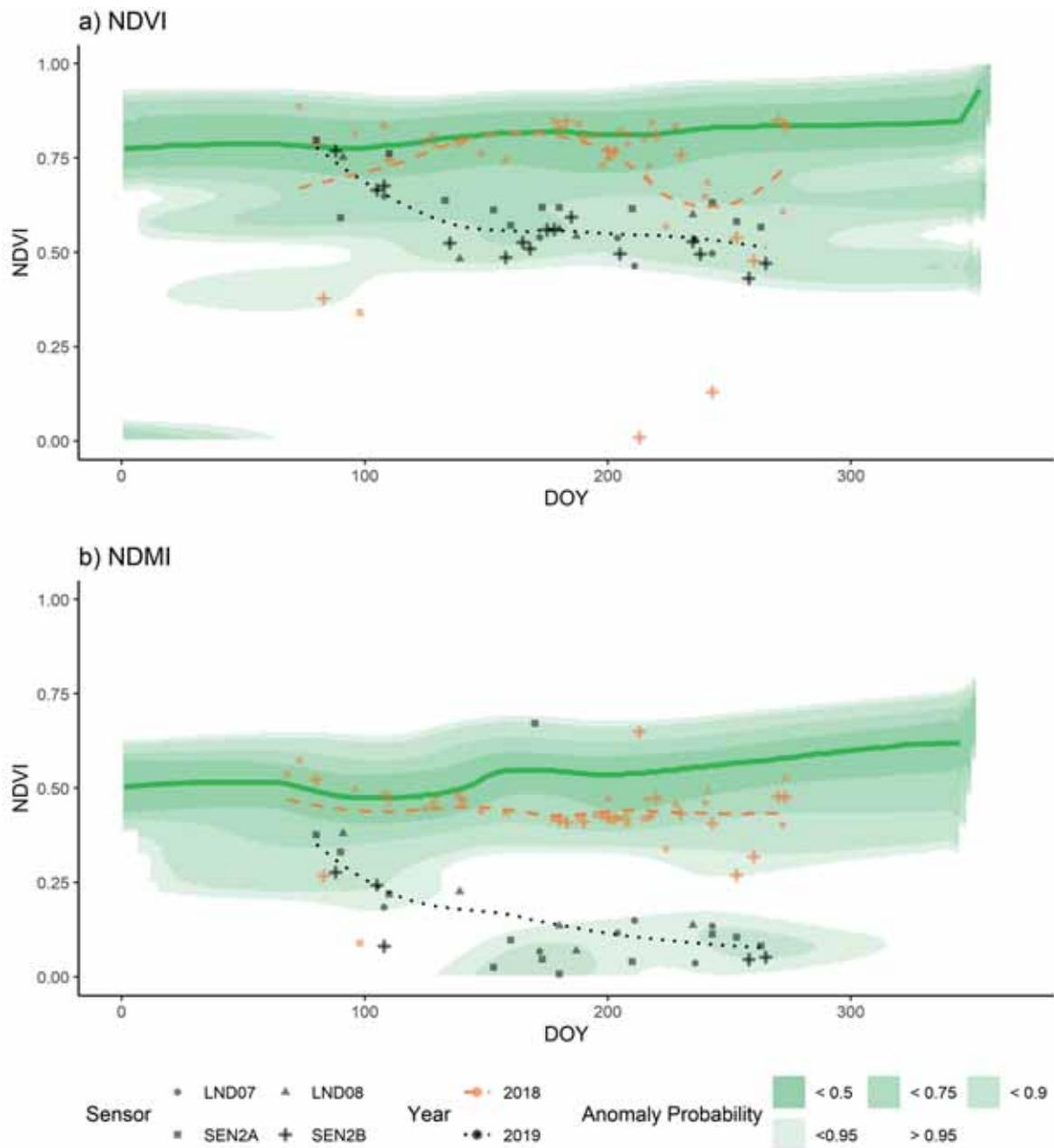


Fig. 4: Phenological baseline in terms of NDVI (a) and NDMI (b) for the pixel indicated in Fig. 3. The bold green line is the baseline, the lighter shaded areas indicate the anomaly probability. Superimposed are the respective DOY-Index values for 2018 (orange) and 2019 (black) as points, the lines in the re-spective colors are smoothing splines that are fit using these points

## 4 Conclusion

We applied a non-parametric procedure as well as multi-sensor multispectral satellite imagery to estimate the effects of droughts a mixed forest area in Germany. This study serves as a proof of concept and showed that the original idea of using a kernel density estimation and satellite imagery to analyze forest disturbances is feasible. It cannot only be performed for MODIS data with relatively coarse resolutions and uniform observation cycles (cf. CHÁVEZ et al. 2019), but is also transferable to higher-resolution, multi-source data with variable spatial, spectral and temporal characteristics. While there are some issues and problems remaining, solving which would considerably improve the approach and resolve several pitfalls, the idea works well and may further be elaborated. Combining it with in-situ data could be a benefit, e.g. to compare the drought response of different tree species. The method may be not only for drought effects on forests, but also for other ecosystems as well as disturbances or other phenomena that require the comparison of a vegetation community's yearly development to their baseline phenology. The approach can complement other techniques that serve the same purpose and can technically be even used for operational monitoring, where the phenological baseline is updated if a season without a major disturbance takes place and each new satellite observation is compared to it. To summarize, the method can be highly beneficial to researchers and users that are concerned with investigating and monitoring the temperate forests of Europe where disturbance regimes, especially the prevalence of droughts, are changing rapidly.

## 5 References

- AGHAKOUCHAK, A., FARAHMAND, A., MELTON, F.S., TEIXEIRA, J., ANDERSON, M.C., WARDLOW, B.D. & HAIN, C.R., 2015: Remote sensing of drought: Progress, challenges and opportunities. *Reviews of Geophysics* **53**, 452-480.
- ALLEN, C.D., BRESHEARS, D.D. & MCDOWELL, N.G., 2015: On underestimation of global vulnerability to tree mortality and forest die-off from hotter drought in the Anthropocene. *Ecosphere* **6**, 129.
- BONAN, G.B., 2008: Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests. *Science* **320**, 1444-1449.
- BOSCH, F., 1997: Die Nutzung des Lebensraumes durch Reh (*Capreolus capreolus* Linné, 1758) und Damhirsch (*Cervus dama* Linné, 1758) im Naherholungsgebiet Kottenforst bei Bonn. *Zeitschrift für Jagdwissenschaft* **43**, 15-23.
- BURAS, A., RAMMIG, A. & ZANG, C.S., 2019: Quantifying impacts of the drought 2018 on European ecosystems in comparison to 2003. *Biogeosciences Discussions*, 1-23.
- CHÁVEZ, R.O., ESTAY, S.A. & RIQUELME, C.G., 2017: npphen: Vegetation Phenological Cycle and Anomaly Detection using Remote Sensing Data. <https://CRAN.R-project.org/package=npphen>.
- CHÁVEZ, R.O., ROCCO, R., GUTIÉRREZ, Á.G., DÖRNER, M. & ESTAY, S.A., 2019: A Self-Calibrated Non-Parametric Time Series Analysis Approach for Assessing Insect Defoliation of Broad-Leaved Deciduous *Nothofagus pumilio* Forests. *Remote Sensing* **11**, 204.

- DUONG, T., WAND, M., CHACON, J. & GRAMACKI, A., 2019: ks: Kernel Smoothing. <https://CRAN.R-project.org/package=ks>.
- ESTAY, S.A. & CHÁVEZ, R.O., 2018: npphen: an R-package for non-parametric reconstruction of vegetation phenology and anomaly detection using remote sensing. *bioRxiv*, 301143.
- FRANTZ, D., 2019: FORCE—Landsat + Sentinel-2 Analysis Ready Data and Beyond. *Remote Sensing* **11**, 1124.
- FRANTZ, D., HAß, E., UHL, A., STOFFELS, J. & HILL, J., 2018: Improvement of the Fmask algorithm for Sentinel-2 images: Separating clouds from bright surfaces based on parallax effects. *Remote Sensing of Environment* **215**, 471-481.
- FRANTZ, D., RÖDER, A., STELLMES, M. & HILL, J., 2016: An Operational Radiometric Landsat Pre-processing Framework for Large-Area Time Series Applications. *IEEE Transactions on Geoscience and Remote Sensing* **54**, 3928-3943.
- GORELICK, N., HANCHER, M., DIXON, M., ILYUSHCHENKO, S., THAU, D. & MOORE, R., 2017: Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* **202**, 18-27.
- HANSEN, M.C., POTAPOV, P.V., MOORE, R., HANCHER, M., TURUBANOVA, S.A., TYUKAVINA, A., THAU, D., STEHMAN, S.V., GOETZ, S.J., LOVELAND, T.R., KOMMAREDDY, A., EGOROV, A., CHINI, L., JUSTICE, C.O. & TOWNSHEND, J.R.G., 2013: High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **342**, 850-853.
- HEALEY, S.P., COHEN, W.B., ZHIQIANG, Y. & KRANKINA, O.N., 2005: Comparison of Tasseled Cap-based Landsat data structures for use in forest disturbance detection. *Remote Sensing of Environment* **97**, 301-310.
- IMBERY, F., FRIEDRICH, K., KOPPE, C., JANSSEN, W., PFEIFROTH, U., DABLER, J. & BISSOLLI, P., 2018: 2018 wärmster Sommer im Norden und Osten Deutschlands. [https://www.dwd.de/DE/leistungen/besondereereignisse/temperatur/20180906\\_waermstersonmer\\_nordenosten2018.html](https://www.dwd.de/DE/leistungen/besondereereignisse/temperatur/20180906_waermstersonmer_nordenosten2018.html).
- KOGAN, F.N., 1990: Remote sensing of weather impacts on vegetation in non-homogeneous areas. *International Journal of Remote Sensing* **11**, 1405-1419.
- MILLAR, C.I. & STEPHENSON, N.L., 2015: Temperate forest health in an era of emerging megadisturbance. *Science* **349**, 823-826.
- PAN, Y., BIRDSEY, R.A., FANG, J., HOUGHTON, R., KAUPPI, P.E., KURZ, W.A., PHILLIPS, O.L., SHVIDENKO, A., LEWIS, S.L., CANADELL, J.G., CIAIS, P., JACKSON, R.B., PACALA, S.W., MCGUIRE, A.D., PIAO, S., RAUTIAINEN, A., SITCH, S. & HAYES, D., 2011: A Large and Persistent Carbon Sink in the World's Forests. *Science* **333**, 988-993.
- R CORE TEAM, 2019: R: A language and Environment for Statistical Computing. Vienna, Austria. <https://www.r-project.org/>.
- REINERMANN, S., GESSNER, U., ASAM, S., KUENZER, C. & DECH, S., 2019: The Effect of Droughts on Vegetation Condition in Germany: An Analysis Based on Two Decades of Satellite Earth Observation Time Series and Crop Yield Statistics. *Remote Sensing* **11**, 1783.
- SEIDL, R., THOM, D., KAUTZ, M., MARTIN-BENITO, D., PELTONIEMI, M., VACCHIANO, G., WILD, J., ASCOLI, D., PETR, M., HONKANIEMI, J., LEXER, M.J., TROTSIUK, V., MAIROTA, P., SVOBODA, M., FABRIKA, M., NAGEL, T.A. & REYER, C.P.O., 2017: Forest disturbances under climate change. *Nature Climate Change* **7**, 395-402.

- SENF, C., PFLUGMACHER, D., ZHIQIANG, Y., SEBALD, J., KNORN, J., NEUMANN, M., HOSTERT, P. & SEIDL, R., 2018: Canopy mortality has doubled in Europe's temperate forests over the last three decades. *Nature Communications* **9**, 1-8.
- SENF, C. & SEIDL, R., 2018: Natural disturbances are spatially diverse but temporally synchronized across temperate forest landscapes in Europe. *Global Change Biology* **24**, 1201-1211.
- WACHTER, H., 2009: Eichenheisterpflanzung – Eichensaat, ein Vergleich nach 100 Jahren. *Allgemeine Forst- und Jagdzeitung* **180**, 177-184.
- WAND, M.P. & JONES, M.C., 1995: Kernel smoothing, 1st ed. ed, Monographs on statistics and applied probability, Chapman & Hall, London & New York.
- WULDER, M.A., HILKER, T., WHITE, J.C., COOPS, N.C., MASEK, J.G., PFLUGMACHER, D. & CREVIER, Y., 2015: Virtual constellations for global terrestrial monitoring. *Remote Sensing of Environment* **170**, 62-76.
- ZHANG, Y., PENG, C., LI, W., FANG, X., ZHANG, T., ZHU, Q., CHEN, H. & ZHAO, P., 2013: Monitoring and estimating drought-induced impacts on forest structure, growth, function, and ecosystem services using remote-sensing data: recent progress and future challenges. *Environmental Reviews* **21**, 103–115.
- ZHU, Z., WANG, S. & WOODCOCK, C.E., 2015: Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sensing of Environment* **159**, 269-277.
- ZHU, Z., WOODCOCK, C.E., OLOFSSON, P., 2012. Continuous monitoring of forest disturbance using all available Landsat imagery. *Remote Sensing of Environment* **122**, 75-91.

## Case study: Near real-time thermal mapping to support firefighting and crisis management

TILMAN BUCHER<sup>1</sup>, MATTHIAS GESSNER<sup>1</sup>, DANIEL HEIN<sup>1</sup>, HENRY MEIßNER<sup>1</sup>,  
THOMAS KRAFT<sup>1</sup>, AGNIESZKA SOSZYŃSKA<sup>1</sup> & SEBASTIAN PLESS<sup>1</sup>

*Abstract: Hot and dry summers have led to an increase in forest fires both concerning numbers and intensity in north-eastern Germany in the last years. In the project FireSense the German Aerospace Center (DLR) has adapted its sensor system MACS (Modular Airborne Camera System) with a set of thermal mid- and long wave infrared (MWIR and LWIR) cameras to detect, monitor and quantify high temperature events (HTE) like forest fires. Ground-based, airborne and spaceborne measurements over fire-experiments are synchronized for cross-validation of the systems and to test the developed workflows.*

*In summer 2019 gas flaring tests were conducted in cooperation of DLR and the Federal Institute for Material Research and Testing (BAM), parallel several large forest fires in Brandenburg (Lieberose) and Mecklenburg-Vorpommern (Lübtheen) developed. In coordination with the crisis management group (local authorities, firefighters, armed forces, federal police) to get the permits MACS conducted 3 flights over the fires in altitudes between 6000 (sunny) down to 3500 ft (under clouds), Lübtheen was covered twice, on July 2 and July 4, when the fire was already under control. Synchronously firefighting helicopters operated close to ground, also delivering videos of the fires for visual interpretation.*

*To get both background temperatures for orientation and landscape features and also information about the fires within one data set, a broad calibration range for the LWIR camera was commanded. Using synchronized position- and orientation data of MACS with given calibration data and a Digital Terrain Model, direct geocoding and the processing of near real-time mosaics was possible using the DLR workflow even without post-processing. The accuracy was sufficient for planning purposes. Geo-tiff maps were delivered shortly after landing within less than three hours. The real-time capabilities of the system could not be used as the flights were conducted on very short notice and the radio link was not installed.*

*The thermal data were delivered as false color heat maps. They show the thermal anomalies very well, clearly discriminating burning area, recently burnt area and unaffected forest. In the RGB data the ground fires are rarely visible as they are covered by and almost did not affect the closely standing crowns. The spread of the fires can be seen in the overlapping regions of adjacent flight lines.*

*Data exchange and use of the data proved to be difficult due to limited data rates and IT infrastructure in the command and situation center in the field, sometimes taking more time than the acquisition and processing. This reduces the practical benefit for the data in the field. For future planned experiments for real-time mapping of forest fires this will be one of the main points to improve the latency of the data transfer to the control center ideally by using a live data link and to optimize the coordination with the control center. Further activities will be coordinated by the Helmholtz Innovation Lab OPTSAL (Optische Technologien für Situationserfassung im Sicherheitsbereich), which was started at DLR in 2020. In OPTSAL hard- and software solutions are developed and activities concerning situational awareness for safety and security are coordinated with industry and authorities.*

---

<sup>1</sup> Deutsches Zentrum für Luft- und Raumfahrt, Institut für Optische Sensorsysteme, Rutherfordstr. 2, D-12489 Berlin, E-Mail: tilman.bucher@dlr.de

## 1 Introduction

Forest fire activity is a continuous threat worldwide. This is to some extent attributed to the climate change, cumulating in the extreme forest fire season in Australia in the summer 2019-2020. In Brandenburg (North-Eastern Germany) the summer 2018 has been a record year concerning number of fires and the size of burnt area (LANDESKOMPETENZZENTRUM FORST EBERSWALDE 2019); this tendency has continued in 2019.

In order to detect and observe fire from space, the German Aerospace Center (DLR) developed and launched two small satellites (TET-1, BIROS) in the context of the FireBIRD (Fire Bispectral InfraRed Detector) earth observation mission (REILE et al. 2013). The mission aims to significantly improve detection, mapping and analysis of high temperature events (HTE) compared to currently existing sensor systems. One main application of FireBIRD is the detection and monitoring of forest fires. (BORG et al. 2017). On a local to regional scale the modular airborne camera system MACS (LEHMANN et al. 2011) is used for 2D and 3D-mapping and monitoring purposes on a variety of different carriers in different spectral ranges, focusing on security and research application. Examples are the projects ‘EMSEC’ for maritime real-time (RT) monitoring (BRAUCHLE et al. 2018), ‘ACHILLES’, with the first responders ISAR (International Search and Rescue) Germany and ‘Live-Lage’, with the Duisburg fire department, all using airborne carriers like planes or UAVs for rapid mapping applications (HEIN et al. 2017). In the ZIM project FireSense ground based, airborne and spaceborne sensors are coordinated to synchronously record fires for cross-validation of the different systems and to evaluate and improve the benefit of different sensors for fire detection and monitoring. In the course of this project the observation of gas flaring tests at BAM could be coordinated with overflights of several contemporary forest fires located in north-eastern Germany to test DLR’s rapid mapping technology under emergency conditions for forest fires. This paper gives an overview of these activities as a case study. A detailed overview over wildfire remote sensing is not goal of this paper, a comprehensive review is given by ALLISON et al. (2016).

## 2 Experiment Description

In June 2019 several experiments were planned and carried out at the BAM to closely evaluate the properties of gas flares using ground based and remote sensing data for the FireSense project. In the same time period several forest fires in north-eastern Germany emerged, including the largest ever recorded fire in Mecklenburg-Vorpommern in Lübtheen. In order to obtain both high temperatures and background land surface temperatures to quantify the gas-flaring tests, a MACS-system with a set of three aligned thermal cameras and a RGB sensor (MACS-Micro) for a true color image were installed into the research airplane of the Free University Berlin. Several flights were conducted with this instrumentation to measure gas flare tests at the BAM and to test the workflow for multi-temporal forest fire monitoring. Details concerning the gas flare experiment, focusing on quantifying radiometry, flame temperatures and their relation to the gas flow can be found in SOSZYNSKA et al. (this volume), the overflights of the forest fires are covered below, focusing on security related aspects.

## 2.1 Camera System and Carrier

Four cameras were aligned and synchronously operated in an experimental setup to observe both HTEs and background temperatures in high detail (see Fig. 1). The cameras used during aerial survey have following characteristics:

Tab. 1: Overview sensors in the MACS-camera system.

	MWIT_High_T	LWIR_Low_T	LWIR_FullRange_T	MACS-Micro Truecolor
<b>Camera type</b>	Infratec Imager IR 8300	Infratec ImagerIR 8800	Infratec VarioCam HD	Industrial High Speed Camera
<b>Sensor type</b>	Actively cooled photon detector InSb	Actively cooled photon detector MCT	Uncooled microbolometer focal plane array	CCD
<b>Wavelength</b>	3.4–4.2 $\mu\text{m}$	8–12 $\mu\text{m}$	7.5–14 $\mu\text{m}$	RGB
<b>Pixel size</b>	15 $\mu\text{m}$	15 $\mu\text{m}$	17 $\mu\text{m}$	7.4 $\mu\text{m}$
<b>Image size [pix]</b>	640×512	640×512	1024×768	4864×3232
<b>Working range used</b>	200 to 500 and 300 to 800 °C	-20 to 20, 0 to 60° and 20 to 100° C	0 to 500 and 250 to 1200 °C	Not applicable

As a major goal of the forest fire flights was to create a near real-time map of the fires, we restricted the near real-time processing on two systems, the RGB MACS-Micro and the LWIR VarioCam HD, which offers a wide calibration range (0 – 500°C used) and sufficient radiometric resolution to cover both background and high temperatures. The cameras were mounted in an experimental setup in a Cessna 207 (Fig. 2) on a stabilized platform (Fig. 1).



Fig. 1: Operation of 4-head camera system on stabilizing platform



Fig. 2: Cessna 207, FU Berlin, cloudy weather

## 2.2 Flight Operation

Due to the fire the airspace over the forest fires was blocked. To receive the flight permits the overflight was coordinated with the local operation teams consisting of representatives of the federal police, local authorities, firefighters and armed forces and air traffic control (Deutsche Flugsicherung, DFS). Weather conditions for Lübben were cloudy with strong winds, therefore the flight altitude was lowered to fly under the clouds. Due to the low flight altitude the flight

plan was adjusted, which resulted in a fairly high spatial resolution but reduced coverage per hour. A lower limit for the flight altitude was given due to the operation of the firefighting helicopters, which were simultaneously operating. This excluded the operation of UAVs, which could have been an alternative for smaller areas. For the Lübtheen flight on 2019/07/02 parts of the fire could not be covered due to clouds below the lower flight limit. There were also helicopters of the federal police operating with thermal- and visual video systems (Fig. 3). These are used for visual inspection of the fires, but don't deploy geocoded image maps for offline planning purposes, which are especially helpful to delineate and track small features.

Tab. 1: Overview of forest fires covered with the MACS-camera system.

Target	Lieberose	Lübtheen_1	Lübtheen_2
Date	2019/06/28	2019/07/02	2019/07/04
Local Time	11:30 – 11:45	16:45 – 17:30	14:00 – 14:45
Altitude	6000 ft	4000 ft	3500 ft
Weather	sunny	cloudy	cloudy
Number Flight-	2 (small area flown as	7 (norther part not com-	9

### 2.3 Forest Fires

Two forest fires were covered: Lieberose (Brandenburg) and Lübtheen (Mecklenburg-Vorpommern), both on former army training areas with limited access due to unexploded ammunition, restricting ground-based firefighting to cleared areas. Due to the larger extent and the multi-temporal coverage the paper concentrates on the data of Lübtheen. The Lübtheen fire started on 2019/06/30, it has been the largest forest fire recorded in history in Mecklenburg-Vorpommern. Local firefighters, Federal Agency for Technical Relief (THW), federal police and armed forces cooperated to fight the fire, coordinated by the district authority. In the main phase 3000 people were activated. The fire affected about 1300 ha and forced the evacuation of three villages. DLR coordinated its activities directly with armed forces and federal police, data distribution was managed through the district authorities.



Fig.3: Oblique view of the Lübtheen fire seen from a federal police helicopter. Data are recorded as side-by-side videos, georeferenced maps are not provided



Fig.4: First overflight Lübtheen forest fire 2019/07/02

### 3 Data Processing and distribution

The data was stored on hard disks during the flight. A real-time radio transmission interface is available and can be operated on planes, UAVs or any other carrier; it was not installed as the flight was initially planned as regular post-processing overflight of the BAM test site only. Data processing started after delivery of the hard disks at DLR in Berlin with a delay of about 90 minutes caused by the flight back to Berlin and the data transfer by car.

For a georeferenced rapid mapping product the parameters for the interior (calibration) and exterior orientation are needed. Calibration data for all sensors are automatically selected in the processing chain from a data-base, exterior position and orientation are measured and stored in the MACS image headers. Therefore an initial rapid mapping product can be generated right after receiving the imagery using the workflow described in HEIN & BERGER (2018). Using drag-and-drop the images are projected on a coarse DTM using an adaptive projection. Systematic offsets due to misalignments and rotations can be interactively adjusted; remaining offsets are mainly caused by the missing post-processing of orientation data and the lack of an aero-triangulation, which is also part of the workflow but was not performed to save time (about 2 hours for this data set). As the image acquisition was continued from the BAM test overflight, which was operated in a different acquisition mode, the thermal data needed one conversion step to apply the rapid mapping tool-chain. Overall the process for the generation of image mosaics from the raw data including the production of color coded maps and export to geo-tiff took about 2.5 hours. If all cameras are operated in the rapid mapping mode, faster response times below one hour are possible, with radio data link real-time fire maps can be realized. In the following figures examples for different data products are given:

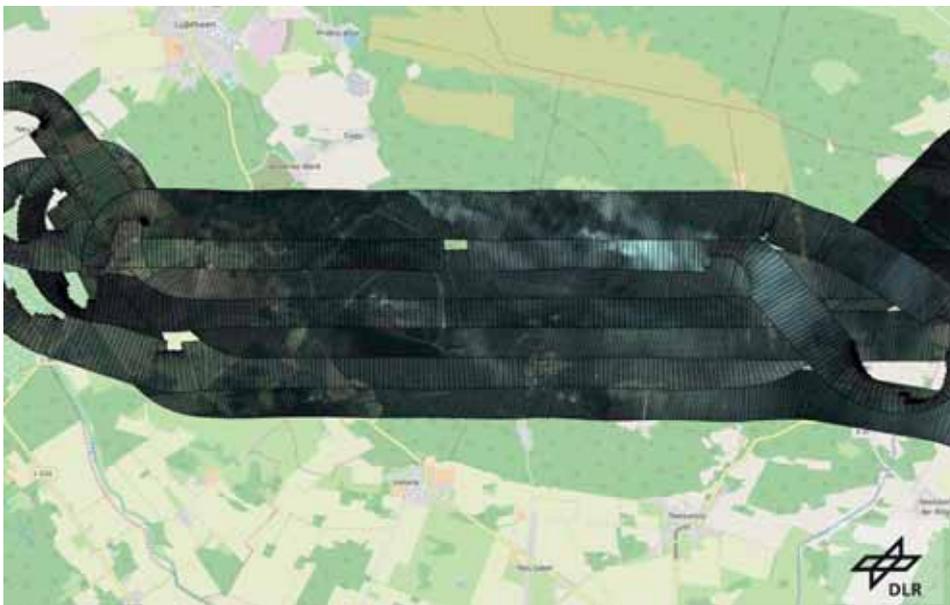


Fig. 5: Rapid mapping of the first overflight Lübbtheen, MACS-Micro RGB, as seen directly after drag-and-drop import into the workflow. Cloudy images in the center (small hole) were removed. High overlap of images of about 80%, black lines outlining the center parts of the images taken for the mosaic. Smoke covers parts of the area

## 4 Results

The prototypic multi-sensor camera system designed to map hot temperature events was successfully applied for tests over forest fires. The offer by DLR to map the forest fires was quickly approved by the local authorities in charge. Coordination and management of the three flights worked well on very short notice.

Within less than three hours after landing GeoTiff map products could be delivered consisting of RGB mosaics (30 / 25 cm resolution) and false color IR data (1 m / 30 cm resolution) showing the affected areas. The data were delivered via ftp-service and to a cloud-server. The operational use of the data on-site was limited mainly due to time/manpower and local IT restrictions. As the data was acquired and delivered on short notice, the IT infrastructure was not well prepared for such data.

In the maps of the thermal imagery the main hotspots as well as many – certainly not all – small glowing nests and the burnt area can easily be identified. In RGB data the burning and burnt areas are almost indiscernible from undisturbed forest as the fire was restricted to ground vegetation and not immediately affecting the crowns which hide most of the fire activity (see Fig. 6). On the other hand RGB data gives explanations to effects like the localization of firefighting infrastructure or precautionary measures which cannot be resolved in thermal data (see Fig. 7).

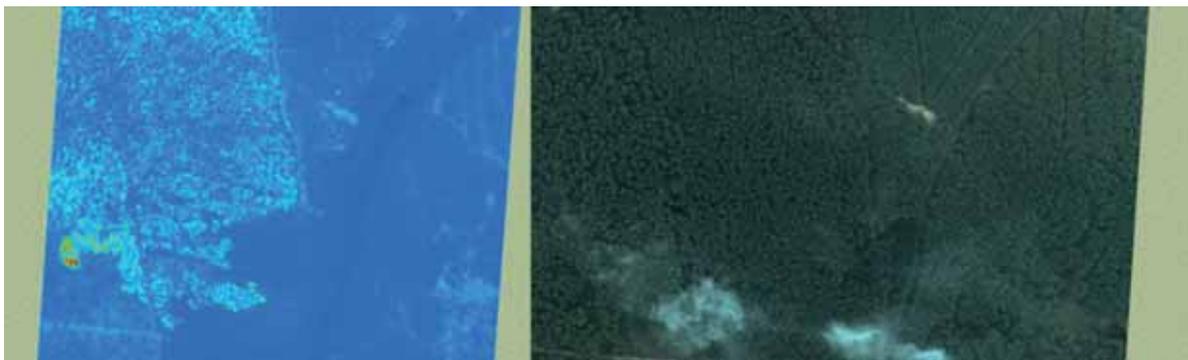


Fig. 6: Individual thermal (false color, left) and RGB images, map-projected: Thermal images nicely show burning and burnt areas, which do not show up in the RGB

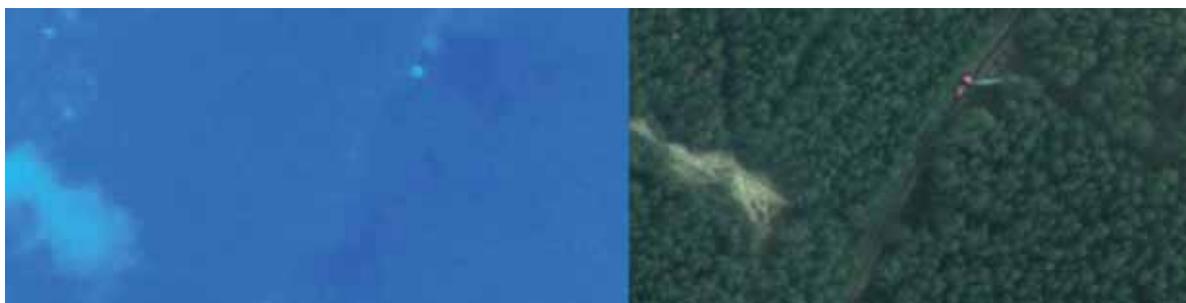


Fig. 7: Zoom into Fig. 6: Colder areas (dark blue) can be nicely explained using in the high resolution RGB image, they are caused by watering the adjacent forest by fire trucks

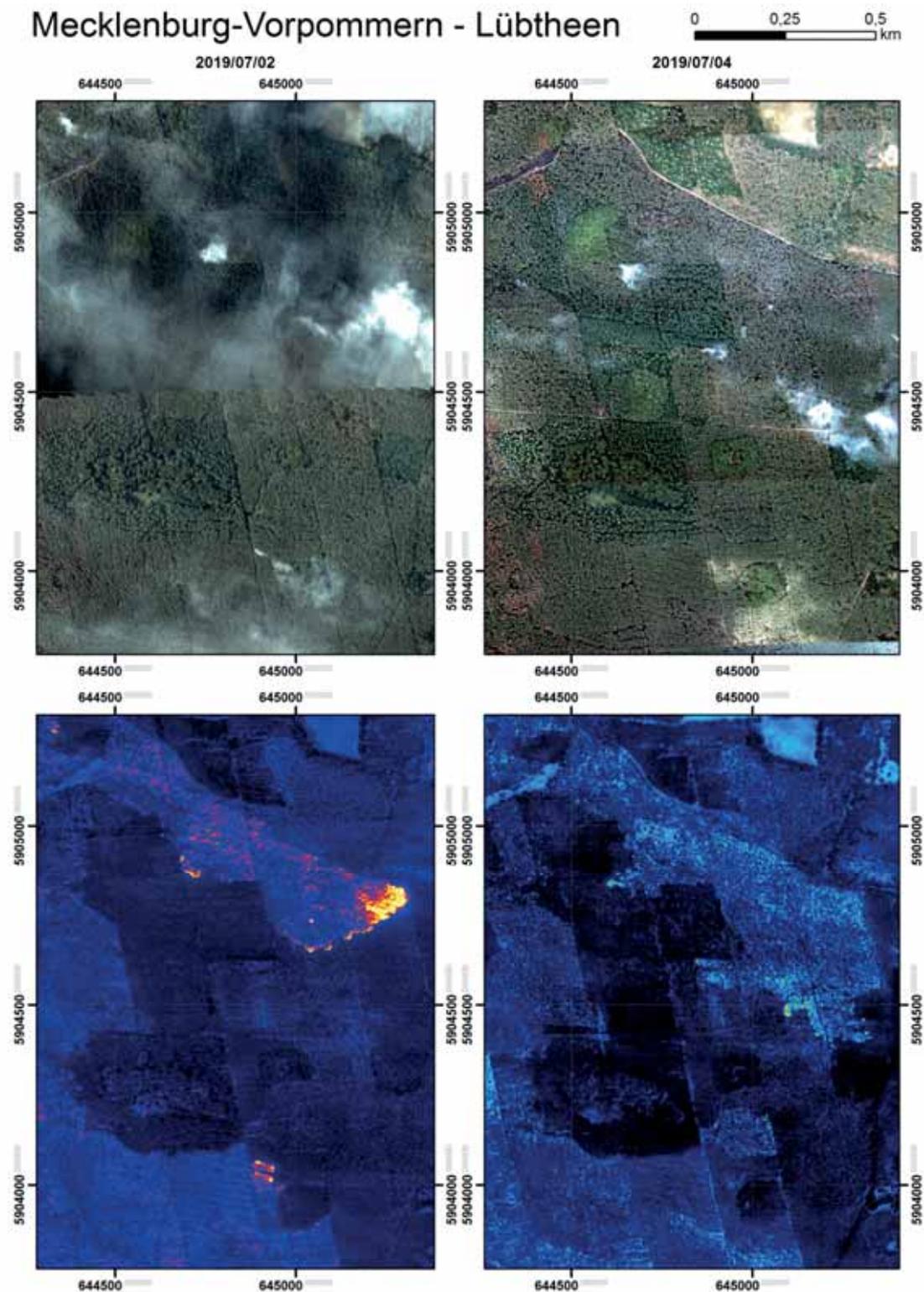


Fig. 8: Single projected thermal (false color, left) and RGB images, zoom: Cool areas (dark blue) caused by spraying the forest nicely show up in the geometrically high resolution RGB image

In the multi-temporal comparison (see Fig. 8) the spread of the fire and the burnt area can be documented. On July 4<sup>th</sup> the fire was under control, only very small active fires remaining. But exactly the monitoring of these small glow nests over a large area is a task which could be ideally suited for thermal infrared remote sensing systems.

The observed temperatures in the thermal imagery are considerably lower than measurements on the ground. This is caused by many factors, including atmosphere and size-of-source effects. Details on these effects are in the focus of the gas-flaring experiments at BAM (see SOSZYŃSKA et al., this volume). For the on-site planning of fire-fighting measures the qualitative false-color maps resulted to be an efficient base for interactive interpretation in geo-information systems.

## 5 Discussion

The results of the overflights of the forest fires are encouraging. Essential information about size and location of burning areas can be provided within few hours. For the near future a real-time monitoring experiment using radio transmission of images to a ground station situated close to the command and situation center seems feasible, the technology is existing. Such an experiment needs to be well prepared especially concerning the interfaces to the planning software in the control center. Starting 2020 the Helmholtz Innovation Lab OPTSAL (Optische Technologien für Situationserfassung im Sicherheitsbereich) was started, which develops hard- and software and coordinates activities concerning situational awareness especially for safety and security with industry and authorities. There is a strong need for operational systems in this field so that the reaction on the growing number of forest fires worldwide can be improved and accelerated.

## 6 Acknowledgements

We thank the partners (Landkreis Dahme-Spreewald, Landkreis Ludwigslust-Parchim, Bundespolizei, Bundeswehr, Feuerwehr, DFS) for the fast response and constructive collaboration. These activities were financed in the course of the ZIM project FireSense.

## 7 References

- ALLISON, R. S., JOHNSTON, J. M., CRAIG, G., & JENNINGS, S., 2016: Airborne Optical and Thermal Remote Sensing for Wildfire Detection and Monitoring. *Sensors* (Basel, Switzerland), **16**(8), 1310. doi:10.3390/s16081310.
- BORG, E., FRAUENBERGER, O., LORENZ, E., PAPROTH, C., SÄUBERLICH, T., FISCHER, C., KLEIN, D., FICHELTMANN, B., VAJEN, H., & HALLE, W., 2017: Validating FireBIRD hot spot detection and fire recognition parameter. In: *Small Satellites for Earth Observation 11th Symposium of the IAA*, Wissenschaft und Technik Verlag, 231-234.
- BRAUCHLE, J., BAYER, S., HEIN, D., BERGER, R. & PLESS, S., 2018: MACS-Mar: a real-time remote sensing system for maritime security applications. *CEAS Space Journal*, 1-10, Springer, DOI: 10.1007/s12567-018-0207-7.

- HEIN, D. & BERGER, R., 2018: Terrain Aware Image Clipping for Real-Time Aerial Mapping. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **IV(1)**, 61-68. DOI: 10.5194/isprs-annals-IV-1-61-2018.
- HEIN, D., KRAFT, T., BRAUCHLE, J. & BERGER, R., 2019: Integrated UAV-Based Real-Time Mapping for Security Applications. *ISPRS International Journal of Geo-Information* **8(5)**, 1-16. Multidisciplinary Digital Publishing Institute (MDPI). DOI: 10.3390/ijgi8050219.
- HEIN, D., BAYER, S., BERGER, R., KRAFT, T. & LESMEISTER, D., 2017: An Integrated Rapid Mapping System for Disaster Management. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **42(W1)**, 499-504, DOI: 10.5194/isprs-archives-XLII-1-W1-499-2017.
- LANDESKOMPETENZZENTRUM FORST EBERSWALDE, 2019: Waldbrandstatistik 2018. <https://forst.brandenburg.de/sixcms/media.php/9/wbra2018.pdf> (letzter Zugriff 15.01.20).
- LEHMANN, F., BERGER, R., BRAUCHLE, J., HEIN, D., MEIBNER, H. & PLESS, S., 2011: MACS - Modular Airborne Camera System for generating photogrammetric high-resolution products. *Zeitschrift der Deutschen Gesellschaft für Geowissenschaften*, (6), 435-446, E. Schweizerbart'sche Verlagsbuchhandlung.
- REILE, H., LORENZ, E. & TERZIBASCHIAN, T., 2013: The FireBird Mission – A Scientific Mission for Earth Observation and Hot Spot Detection. *Small Satellites for Earth Observation. Digest of the 9th International Symposium of the International Academy of Astronautics*. Wissenschaft und Technik Verlag.

# CORONA-Höhenmodelle und Orthofotos

KARSTEN JACOBSEN<sup>1</sup> & ROLF RÜDIGER LUDWIG<sup>2</sup>

*Zusammenfassung: Für Stadtentwicklungsprojekte in Bangladesh werden Orthofotos und Höhenmodelle aus der Zeit vor etwa 50 Jahren benötigt als Grundlage einer geomorphologischen Kartierung und um die frühere Lage der Wasserläufe berücksichtigen zu können. Beides ist zur Einschätzung der Baugrundstabilität und Gebietsdrainage erforderlich. Alte Unterlagen sind nicht vorhanden, weswegen auf CORONA-Satellitenbilder zurückgegriffen werden musste. CORONA Bilder wurden mit einer Panoramakamera auf Dünnschichtfilm im Format von ~75cm x 5,5cm aufgenommen, womit die Bildgeometrie schwierig ist. Aus diesem Grund wird in der Literatur nur von der Auswertung von Bildausschnitten berichtet. Für das Projekt war aber die Auswertung der ganzen Filme erforderlich, was eine intensive Analyse und Berücksichtigung von systematischen Bildfehlern erforderte. Die absolute Standardabweichung der Orthofotos konnte so mit etwa 10m bis 15m realisiert werden. Um die gute relative Höhenbestimmungsgenauigkeit von etwa 3m auch absolut zu erreichen wurden die Höhenmodelle auf das frei verfügbare TDM90 DSM eingepasst.*

## 1 Einleitung

Bangladesch weist weltweit eine anhaltend sehr hohe Urbanisierungsrate auf. Dhaka ist beispielsweise seit 1975 um das 35-fache gewachsen und nimmt weltweit eine Spitzenposition ein. Neubaugebiete entstehen zunehmend in Niederungen auf mehrere Meter mächtigen künstlichen Sandaufschüttungen um die Hochwassergefahr zu minimieren. Hinweise auf ehemalige Gewässer und Niederungen werden damit verdeckt. Für die Stadtplanung ist deren ehemalige Lage, die sich in dem Hochwassergebiet seit 1972 stark verschoben hat, jedoch von großer Bedeutung, da hier erhöhter Gründungs- und Konstruktionsaufwand zu erwarten ist. Unzureichend Planung von Straßendämmen behindert die natürliche Drainage. Alte Luftbilder und Höhenmodelle sind nicht verfügbar. Es bieten sich vorhandene CORONA-Aufnahmen an (MCDONALD 1997). CORONA KH-4B Stereopaare vom April 1972 decken den für ein Vorhaben der bangladeschisch-deutschen technischen Zusammenarbeit interessierenden Bereich ab. Die amerikanischen CORONA-Aufnahmen aus der Anfangszeit der Satellitentechnik sind eine häufig genutzte Möglichkeit um Informationen über alte Zustände preisgünstig zu erhalten. CORONA-Aufnahmen wurden ursprünglich primär für Objektinterpretationen eingesetzt, mit etwas Aufwand sind aber auch geometrische Auswertungen möglich.

## 2 CORONA Kamera und Bildaufnahme

Die CORONA KH-4B-Aufnahmen mit einer Panoramakamera mit 610mm Kammerkonstante (Abb. 1) aus etwa 154 km Höhe haben für die damalige Zeit eine hervor ragende nominelle Objektauflösung entsprechend der vom USGS verwendeten 7 µm Pixelgröße der gescannten Fotos von 1.77 m in der Bildmitte bis 2.18 m in Y-Richtung und 2.69 m in X-Richtung an den Bildenden, entsprechend im Flächenmittel etwa 2,0 m. Eine Kantenanalyse (JACOBSEN 2008) ergab eine effektive Objektauflösung,

---

<sup>1</sup> Leibniz Universität Hannover, Nienburger Str. 1, D-30167 Hannover; E-Mail: jacobsen@ipi.uni-hannover.de

<sup>2</sup> Bundesanstalt für Geowissenschaften und Rohstoffe; D- 30655 Hannover E-Mail: r.ludwig@bgr.de

die etwa 10% über der nominellen liegt. Nur die letzten etwa 2 cm an den Bildenden sind wegen deutlich unscharfer Abbildung nur eingeschränkt brauchbar.

CORONA KH-4B hat eine vorwärts und eine rückwärts gerichtete Panoramakamera mit einem Nadirwinkel von jeweils etwa  $15^\circ$  (Abb. 1), was zu einem Basis-Höhenverhältnis des Stereomodells von 1: 1.85 führt.

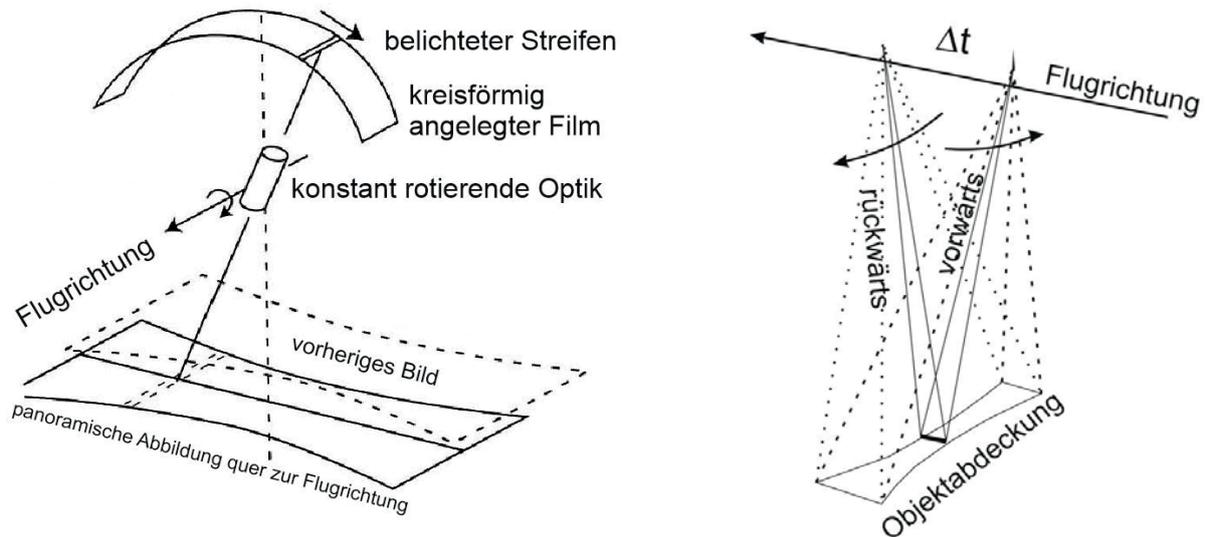


Abb.1: Abbildungs- und Aufnahmeprinzip von CORONA KH4B

Aus etwa 154 km Höhe wird durch eine Panoramaaufnahme von  $75\text{cm} \times 5,5\text{cm}$  eine Fläche von 220 km Länge  $\times$  14 km Breite in der Bildmitte und etwa 17km am Bildende erfasst. Benachbarte Bilder überdecken sich geringfügig in der Bildmitte und bis zu 40% an den Bildenden. Die zu einem Stereopaar zusammengefügte, korrespondierenden Vorwärts- und Rückwärtsaufnahmen überdecken sich in Flugrichtung nur zu etwa 90% (Abb. 2), womit im mittleren Bereich benachbarter Bilder eine Lücke zwischen den erstellten Höhenmodellkombinationen verbleibt, die durch eine schmale Stereokombination des Vorwärtsbildes des vorangegangenen Stereopaars und des aktuellen Rückwärtsblickes geschlossen werden kann.

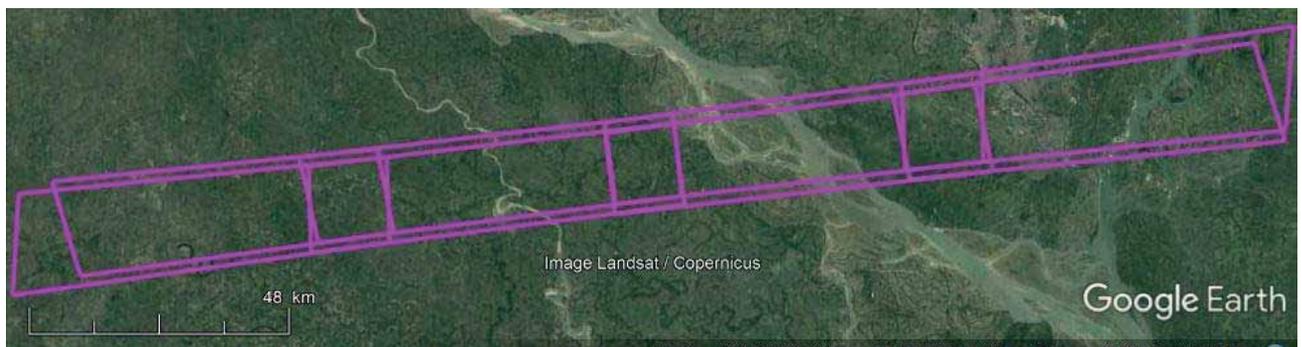


Abb. 2: Durch ein Stereopaar mit den vier Bildteilen der Aufnahmen erfasste Fläche

CORONA-Bilder werden mit  $7\ \mu\text{m}$  Pixelgröße in 4 überlappenden Teilen gescannt (Abb. 2). Es ist erforderlich, dass die Bildteile mit Hilfe von Verknüpfungspunkten durch eine Ähnlichkeitstransfor-

mation zusammengesetzt werden da auch leichte Rotationen durch die mechanische Anlage der Bilder im Scanner vorhanden sind. Die Auswirkung der geneigten Aufnahmerichtungen ist besonders an den nicht parallelen Bildenden in Abb.2 erkennbar.

Wegen des Filmformats von etwa 75 cm quer zur Flugrichtung und 5,5 cm in Flugrichtung sind Bildverbiegungen zu erwarten. Bildverbiegungen lassen sich durch die Seitenlinien in Filmlängsrichtung kontrollieren (Abb. 3).

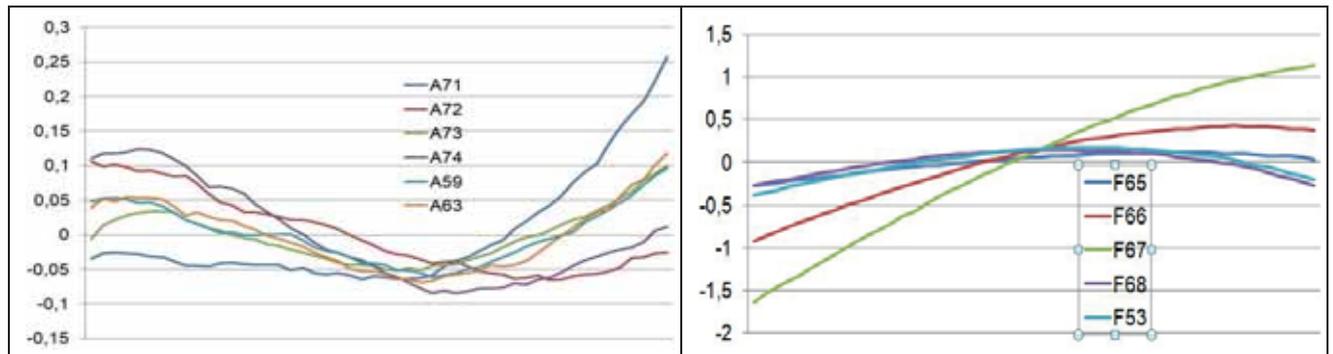


Abb. 3: Seitenlinienkorrektur von Rückwärtskamera- (A) und Vorwärtskameraaufnahmen (F) [mm]

Die Bildverbiegungen haben einen klaren Trend. Mit einer Pfeilhöhe von etwa 0,2 mm bzw. 0,5 mm, entsprechend etwa 30 bis 70 Pixeln, sind sie nicht vernachlässigbar. In Abb. 3 fällt auf, dass die Verbiegungen ähnlich, aber nicht identisch sind; sie unterscheiden sich deutlich durch eine Rotation obwohl die Bildkoordinaten auf die beiden Markierungen der Bildmitte an den Bildenden transformiert wurden. Solche leichten Rotationen werden durch die äußere Orientierung ausgeglichen. In einigen Bildern fehlten diese Markierungen, die Rahmenmarken entsprechen, die dann mittels der Bildbegrenzungen in y-Richtung rekonstruiert wurden. Die Seitenlinien kamen für die Rekonstruktion nicht infrage da sie an den Bildenden nicht mehr erkennbar sind.

### 3 Bildorientierung

Panoramaaufnahmen (Abb. 1) haben nur quer zur Scanrichtung eine perspektivische Geometrie. Da viele Programme zur Auswertung auf zentralperspektiver Geometrie basieren, ist es sinnvoll, die Panoramageometrie auf zentralperspektive Geometrie umzurechnen. Eine Projektion vom mittleren Projektionszentrum über die kreisförmige Lage des Bildes auf eine Tangentialebene ist wesentlicher Teil der Umrechnung, berücksichtigt aber nicht die Bewegung des Projektionszentrums über die 0,39 sec Scanzeit entsprechend 3050m im Orbit. Wegen des kleineren Bildmaßstabs an den Bildenden führt dieses zu einer leicht S-förmigen Deformation der in den Objektraum projizierten Bildachse (DOWMAN et al. 2012: 103ff; SCHNEIDER et al. 2001). Zwar ist die Umdrehungsgeschwindigkeit der CORONA-Optik bekannt, CORONA-KH-4B war jedoch mit einer Vorwärtsbewegungskompensation durch eine kontinuierliche Aufnahmerichtungsänderung ausgestattet (AUELMANN 2015), die den Einfluss der Bewegung des Projektionszentrums im Objektraum verringerte und praktisch eine Berechnung der leicht S-förmigen Deformation nicht ermöglichte. Die exakte Bildgeometrie der auf eine Tangentialebene projizierten CORONA-Bilder konnte somit nur durch Selbstkalibrierung bestimmt werden. Durch die Aufnahmezeit von 0,39 sec ist eine Auswirkung einer Satellitenrotation auf die

Bildgeometrie nicht auszuschließen, außerdem kann die kreisförmige Lage des Films zu individuellen Anlagefehlern führen, womit nicht vorausgesetzt werden kann, dass die Bildgeometrie unverändert bleibt. Für die vorwärts und die rückwärts blickenden Kameras wird sie ohnehin unterschiedlich sein. Die Bildgeometrie kann somit nur mittels einer ausreichenden Anzahl von Passpunkten bestimmt werden. Die Passpunkte wurden in Google Earth bestimmt. Aus der Auswertung einer Reihe von höher aufgelösten Satellitenaufnahmen durch einen der Autoren ist bekannt, dass bei Google Earth mit einer Objektpunktgenauigkeit von in der Regel besser als 2 m zu rechnen ist, was für die CORONA-Aufnahmen mit etwa 2 m Objektpixelgröße ausreichend ist. Das größere Problem ist ohnehin die Identifizierung von unveränderten Punkten in den CORONA-Aufnahmen, die in Google Earth erkennbar sind. Seit 1972 hat sich das Projektgebiet in Bangladesh stark verändert. Es haben sich nicht nur die Ströme teilweise um Kilometer verschoben, auch das Straßennetz hat sich stark verändert. Im Innenbereich von Städten gibt es noch die geringsten Veränderungen, allerdings sind heute wegen des stark angewachsenen Baumbestands viele Kreuzungen in den Satellitenbildern nicht mehr erkennbar. Andere Kreuzungen sind heute stark abgerundet und die Straßen haben sich verbreitert, wobei nichtmehr erkennbar ist, wo die Straße früher lag. Selbst bei Eisenbahnen gibt es das Problem, dass heute zwei Spuren, statt früher eine Spur, vorhanden sind und sich die Bahnübergänge stark verändert haben. Die wenigen Brücken, die früher vorhanden waren, sind häufig durch breitere ersetzt worden. Es ist somit sehr zeitaufwändig geeignete Passpunkte zu finden und die Genauigkeit der Identifizierung ist beschränkt.

Trotz dieser Probleme konnten zwischen 50 und 90 Passpunkte je Bild gefunden werden. Im Bereich der großen Ströme gibt es allerdings teilweise 5 km breite Bereiche in denen keine unveränderten Objekte vorhanden sind.

Die Orientierung und Berechnung von Objektkoordinaten erfolgte in einem tangentialen Koordinatensystem zum Erdellipsoid da bei 220 km abgedecktem Bereich eine Erdkrümmungskorrektur der Bildkoordinaten nicht akzeptable Restfehler zweiter Ordnung erzeugt.

Die Abbildungen 4 bis 7 zeigen die systematischen Bildfehler basierend auf den Passpunkten der Einzelbilder. Die Widersprüche an den Passpunkten wurden gefiltert um die Auswirkung von Ausreißern zu reduzieren. In y-Richtung erfolgte eine lineare Ausgleichung zur Füllung von Lücken da die Punkte mit gleichem x-Wert gleichzeitig aufgenommen wurden (Abb. 1, links) und somit dynamische Effekte ausgeschlossen werden können. In der x-Richtung wurden Lücken durch lineare Interpolation benachbarter Spalten gefüllt.

Die so ermittelten systematischen Bildfehler der Abbildungen 4 bis 6 zeigen eine gewisse Ähnlichkeit, wie auch zu den folgenden 3 Bildern, deren systematische Bildfehler hier nicht gezeigt werden. Ab dem sechsten Bild ändern sich die systematischen Bildfehler, wie auch Abb. 7 zeigt. Basierend auf 10 Bildern und insgesamt 720 Widersprüchen an Passpunkten ergeben sich die in Abb. 8 gezeigten Werte.

Ohne eine Vorkorrektur haben die räumlichen Rückwärtsschnitte einen Gewichtseinheitsfehler von etwa 63  $\mu\text{m}$  bis 90  $\mu\text{m}$ , im Mittel etwa 80  $\mu\text{m}$ . Bei 7  $\mu\text{m}$  Pixelgröße entspricht das 11,4 Pixeln oder etwa 23 m im Objektraum. Das ist nicht die Genauigkeit, die von anderen Satellitenbildern bekannt ist, bei denen mit sehr guten und klar identifizierbaren Passpunkten etwa eine Standardabweichung von einem Pixel erreicht werden kann. Diese Ergebnisse stimmen mit denen in (SCHNEIDER et al. 2001) in China erreichten überein, wo ohne o.a. Vorkorrektur an 16 Passpunkten Standardabweichungen in X von 27.6 m, in Y von 19.7 m und in Z von 14.2 m erzielt wurden. Es muss bei CORONA berücksich-

tigt werden, dass die Passpunktidentifizierung schwierig ist und es wegen des langen Filmformates, sowie der dynamischen Aufnahme geometrische Probleme gibt.

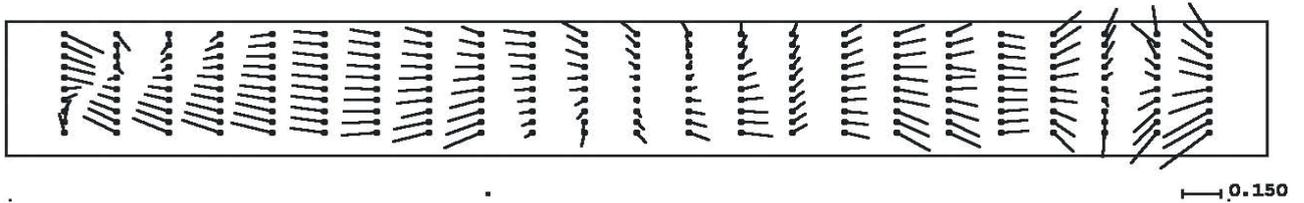


Abb. 4: Systematische Bildfehler, Rückwärtskamera, Bild 63, 49 Passpunkte, Vektormaßstab [ $\mu\text{m}$ ]

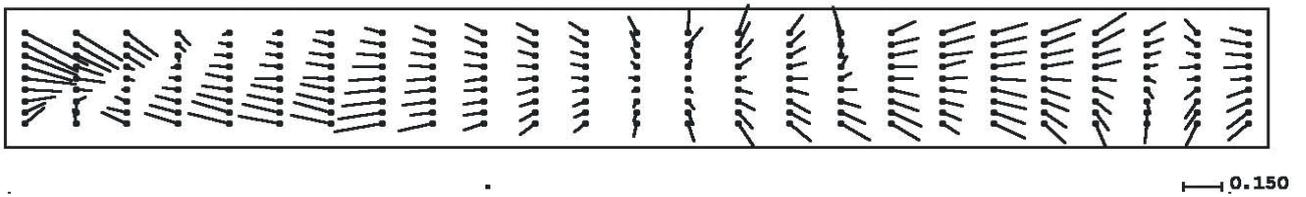


Abb. 5: Systematische Bildfehler, Rückwärtskamera, Bild 64, 80 Passpunkte, Vektormaßstab [ $\mu\text{m}$ ]

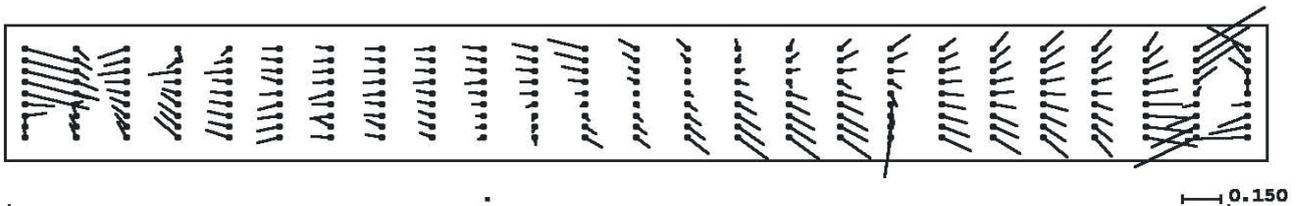


Abb. 6: Systematische Bildfehler, Rückwärtskamera, Bild 65, 79 Passpunkte, Vektormaßstab [ $\mu\text{m}$ ]

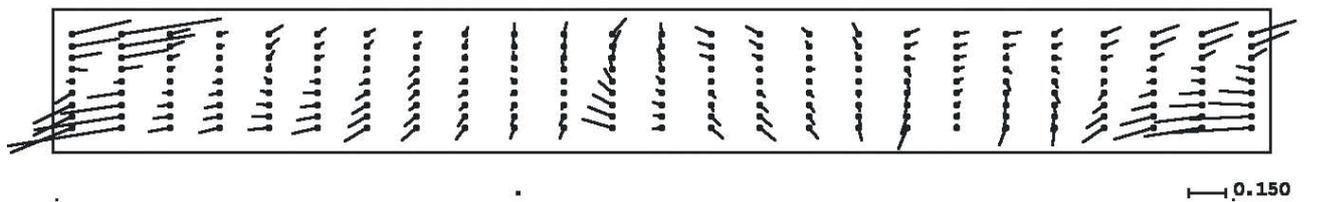


Abb. 7: Systematische Bildfehler, Rückwärtskamera, Bild 74, 72 Passpunkte, Vektormaßstab [ $\mu\text{m}$ ]

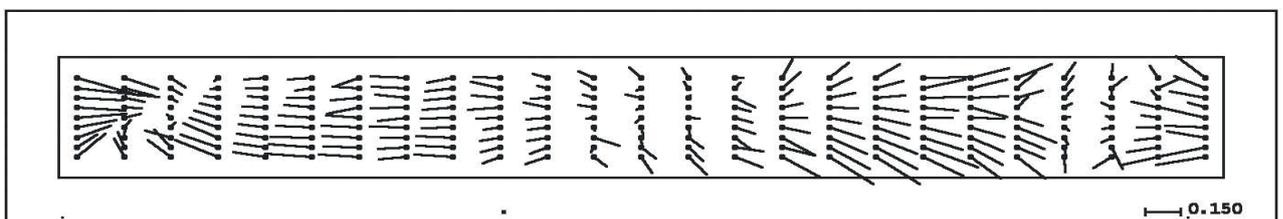


Abb. 8: Mittlere systematische Bildfehler, Rückwärtskamera, basierend auf 10 Bildern mit insgesamt 720 Passpunkten

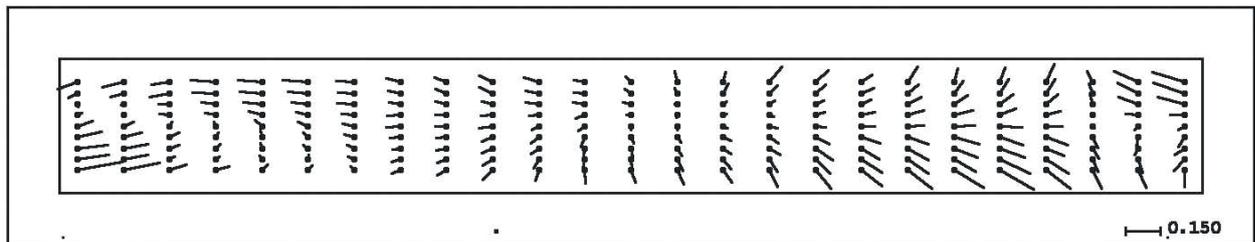


Abb. 9: Mittlere systematische Bildfehler, Vorwärtskamera, basierend auf 10 Bildern, 720 Passpunkte

Aus den Abbildungen 4 bis 8 zeigt sich klar, dass die systematischen Fehler in der x-Richtung etwa doppelt so groß sind wie in der y-Richtung, was sich durch die Seitenlinienkorrektur (Abb. 3) erklären lässt, die nur für die y-Koordinaten Korrekturwerte liefert. Besonders an den Bildenden zeigen sich Verdrehungen des Abbildungsschlitzes (Abb. 1 links) und in x-Richtungen systematische Fehler, die durch ungleichmäßige Scangeschwindigkeit oder nicht exakte kreisförmige Lage des Films erklärt werden können. Eine S-förmige Deformation ist nicht erkennbar.

Durch eine Vorkorrektur der um die Seitenlinienkorrektur und panoramische Umrechnung berechneten perspektiven Bildkoordinaten mit den in den Abbildungen 8 und 9 dargestellten mittleren Bildkoordinatenerbesserungen verbesserte sich die Genauigkeit des räumlichen Rückwärtsschnittes von im Mittel  $80\mu\text{m}$  auf etwa  $60\mu\text{m}$ , entsprechend 8.5 Pixel oder 17 m. Bei zusätzlicher perspektiver Anpassung auf die Passpunkte der Teilbilder konnte für die mittleren beiden Teilbilder mit einer Objektpixelgröße von 1.77 m eine Lagegenauigkeit von etwa 10 m erreicht werden. Für die äußeren Teilbilder, die größere Objektpixel von bis zu 2.18m in Y-Richtung und 2.69 m in X-Richtung haben und auch im Bereich der Bildenden eine schlechtere Bildqualität zeigen, reduziert sich die Standardabweichung der Lage auf etwa 15 m.

#### 4 Erstellung von digitalen Oberflächenmodellen

Das Basis-Höhenverhältnis des Stereomodells von 1: 1.85 ist für die Erstellung von Höhenmodellen gut geeignet. Da im Projektbereich im Jahr 1972, im Gegensatz zu heute, relativ wenige Bäume vorhanden waren, werden in dem digitalen Oberflächenmodell (DOM) die Geländeoberfläche und die Gebäude gut erfasst. Wegen der Grauwertvariationen durch das Filmkorn und der im Gelände nicht vorhandenen klaren Kanten ist eine semiglobale Bildzuordnung nicht sinnvoll und es wurde eine flächenbasierte Bildzuordnung nach der Methode der kleinsten Quadrate eingesetzt. Dieses war auch erforderlich da die zugeordneten Bildfenster durch die perspektive Abbildung gegeneinander geschert sind, was in Abb. 10 an der rechten Seite erkennbar ist. Die Höhen wurden für jedes dritte Pixel bestimmt, entsprechend einem Objektpunkt Abstand von etwa 6m. Die vorliegenden Bilder waren nicht immer optimal gescannt und erforderten für einzelne Teilbilder eine Wallis-Filterung um überhaupt noch den Bildinhalt erkennen zu können. Trotzdem waren die Bildzuordnungsergebnisse nicht schlecht. Mit Ausnahme der Wasserflächen hatten etwa 50% der zugeordneten Bildfenster einen Korrelationskoeffizienten von größer als 0.9 (Abb. 10).

Die landwirtschaftlich genutzten Flächen weisen teilweise keine hohen Kontraste auf, was sich in Abb. 10 durch die gelben Bereiche zeigt. Die nicht vollständige Überdeckung des Vorwärts- und des Rückwärtsbildes ist in Abb. 10 am unteren Bildrand erkennbar. Die linke Seite von Abb. 10 entspricht der

Mitte des vollständigen CORONA-Bildes, hier ist die nicht stereoskopisch überdeckte Fläche größer als auf der rechten Seite, wo die Objektpixel größer als in der Bildmitte sind. Durch Satellitenrotationen fällt dieser nicht überdeckte Bereich in einigen Stereobildern deutlich größer aus.

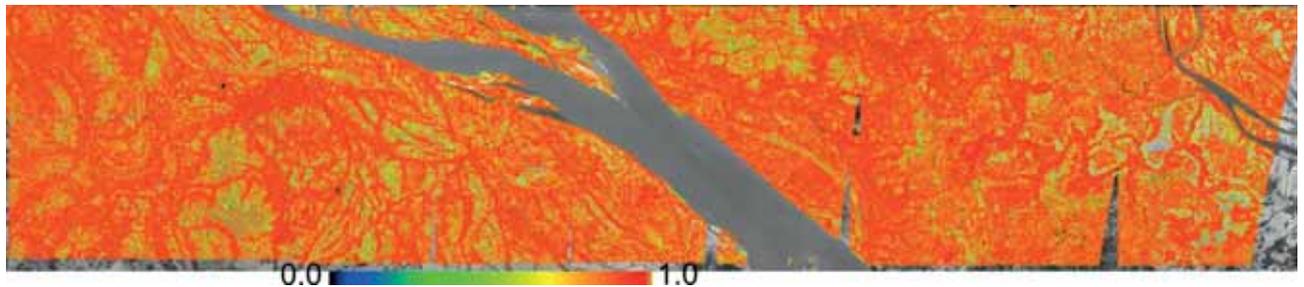


Abb. 10: Farbkodierte Korrelationskoeffizienten, einem CORONA-Teilbild überlagert; ~ 70 km × 20 km

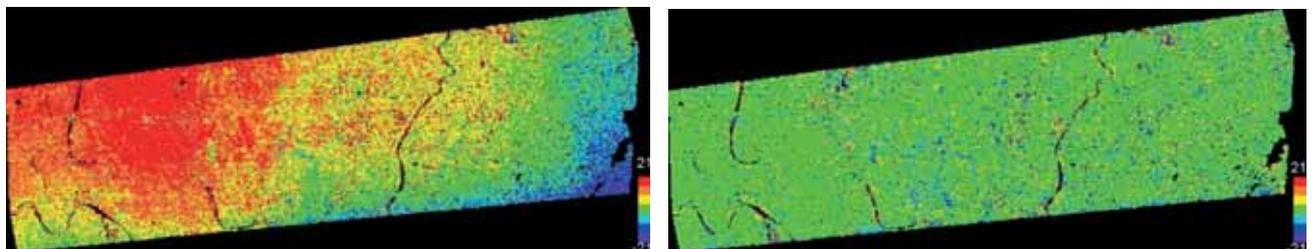


Abb. 11: Farbkodierte Höhendifferenzen CORONA-DOM gegen TDM90 im Bereich der Stadt Dhaka, Links: ursprüngliches CORONA-DOM, rechts: CORONA DOM nach Anpassung auf das TDM90; Farbskala von -21 bis +21m

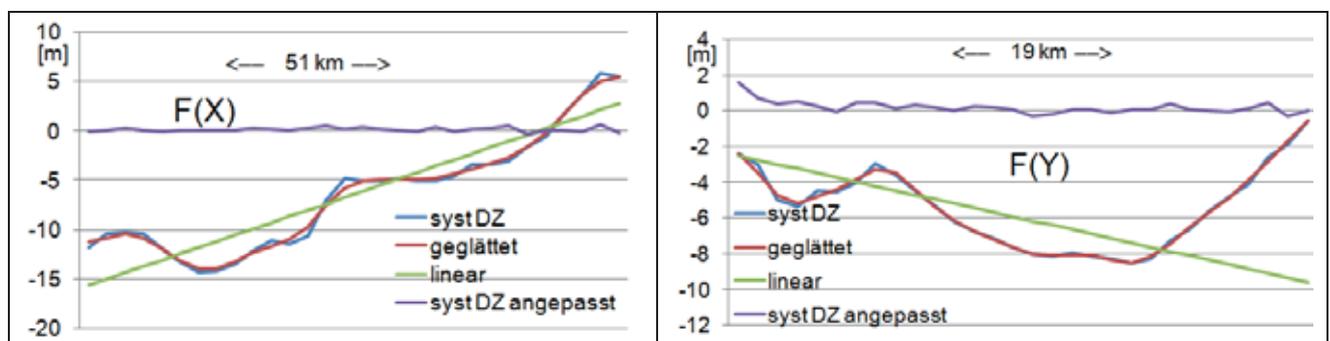


Abb. 12: Systematische Höhendifferenzen CORONA-DOM gegen TDM90 als Funktion von X und Y

Abb. 11 zeigt farbkodierte Höhendifferenzen zwischen dem CORONA DOM und dem TDM90 DOM. TDM90 ist das frei verfügbare TanDEM-X (WorldDEM) DOM, reduziert auf etwa 90m Rasterweite (ALDOSARI & JACOBSEN 2019). Es ist zur Zeit das frei verfügbare DOM mit der höchsten Genauigkeit und erreicht in flachen Gebieten eine Standardabweichung von bis zu 1m. Die linke Seite der Abb. 11 zeigt die ursprünglichen Höhendifferenzen, die in dieser Größenordnung nach den obigen Ergebnissen der Orientierungsuntersuchungen zu erwarten waren. Der Verlauf der systematischen Höhenfehler als Funktion von X und Y ist in Abb. 12 gezeigt. Neben einer Neigung - in Abb. 12 mit linear gekennzeichnet - sind lokale Abweichungen vorhanden. Die offensichtlichen lokalen systematischen Höhenfehler des CORONA DOM wurden durch Anpassung des CORONA DOM auf das TDM90 beseitigt.

Dafür wurden alle Höhendifferenzen zwischen CORONA und TDM90 berechnet und für die Anpassung das gleitende Mittel aller Höhendifferenzen im Bereich von  $\pm 500\text{m}$  Abstand als Korrektur an den jeweiligen CORONA-DOM-Punkt angebracht. So wurden die morphologischen Details des CORONA-DOM erhalten und die systematischen Fehler beseitigt, wie die farbkodierten Höhendifferenzen in Abb. 11 rechts und die angepassten systematischen Höhendifferenzen als  $F(X)$  und  $F(Y)$  in Abb. 12 zeigen.

Vor der Anpassung an TDM90 gab es im obigen Beispiel ein Bias von  $-7.67\text{m}$  und eine Standardabweichung von  $8.46\text{m}$ . Die relative Standardabweichung benachbarter Höhenpunkte im Abstand bis zu  $50\text{m}$  beträgt  $3.65\text{m}$ . Nach der Anpassung auf TDM90 reduzierte sich der Bias auf  $0.12\text{m}$ , die Standardabweichung auf  $4.64\text{m}$  und die relative Standardabweichung auf  $3,02\text{m}$ . In Relation zu etwa  $2\text{m}$  Objektpixelgröße und dem Basis-Höhenverhältnis von  $1:1.85$  entspricht das einer Standardabweichung der x-Parallaxe von  $1,2$  Pixeln bzw. für die relative Genauigkeit  $0,82$  Pixel, was als sehr gutes Ergebnis bezeichnet werden kann. Durch die Anpassung auf TDM90 war auch keine Höhenanpassung benachbarter CORONA DOM erforderlich.

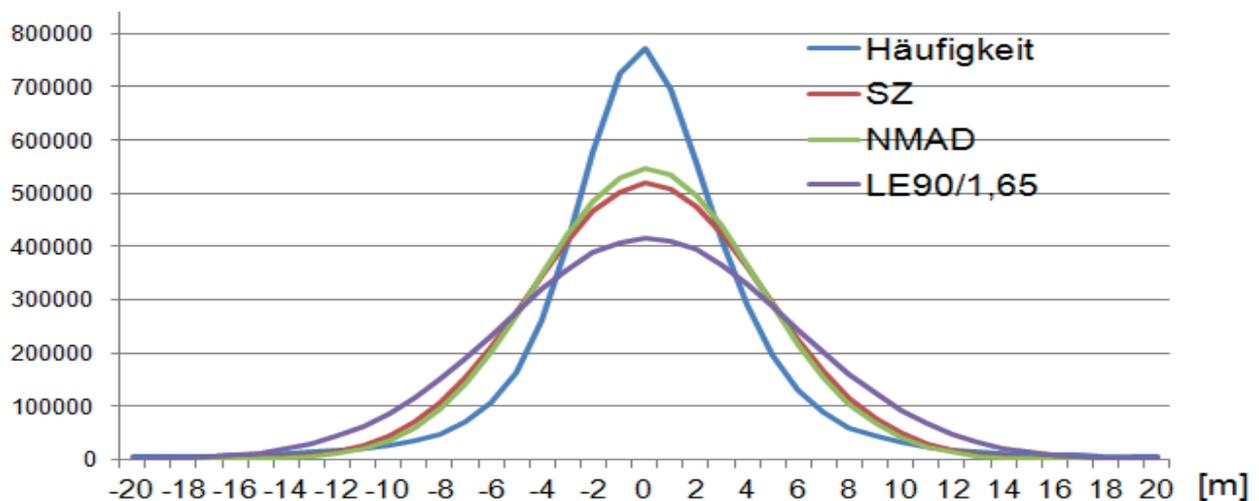


Abb. 13: Häufigkeitsverteilung der Höhenwidrsprüche CORONA DEM nach Anpassung auf TDM90 gegen TDM90 DOM, und Normalverteilung basierend auf NMAD, SZ und normiertem LE90

Die Häufigkeitsverteilung der Widersprüche des angepassten CORONA DOM gegenüber TDM90 zeigt keine Auffälligkeiten. Wie üblich ist die Häufigkeit im zentralen Bereich größer (ALDOSARI & JACOBSEN 2019), NMAD basierend auf der Normalverteilung passt besser mit der Häufigkeitsverteilung überein als die von SZ und besonders die des normierten LE90. LE90 ist nur durch die 10% größten Differenzen bestimmt, wodurch es stark von lokalen Abweichungen, hervorgerufen durch Gebäude und Bäume, bestimmt ist.

Wie oben angemerkt, überdecken sich die Bilder der Vorwärts- und der Rückwärtskameras nicht vollständig (Abb. 2). Dadurch gibt es Lücken zwischen benachbarten Höhenmodellen, die durch die kreuzweise, schmale Überdeckung des Bildes der Vorwärtskamera des nördlichen Nachbarbildes und des Rückwärtsbildes des aktuellen Bildes geschlossen werden können (Abb. 14).

Das in Abb. 14 gezeigte Beispiel liegt etwa in der Mitte von CORONA-Bildern und hat Objektpixelgrößen von unter  $2\text{m}$ . Hier waren die Lücken zwischen benachbarten Standardstereomodellen aus der

Vorwärts- und der Rückwärtskamera etwa 2km breit und mussten durch kreuzweise Bildkombinationen geschlossen werden. Die Flächen ohne Höhenwerte im oberen Teil von Abb. 14 sind auf Überflutungen zurückzuführen. Die DOM-Kombination in Abb. 14 überdeckt etwa 66 km × 34 km und enthält 67 Millionen Höhenpunkte.

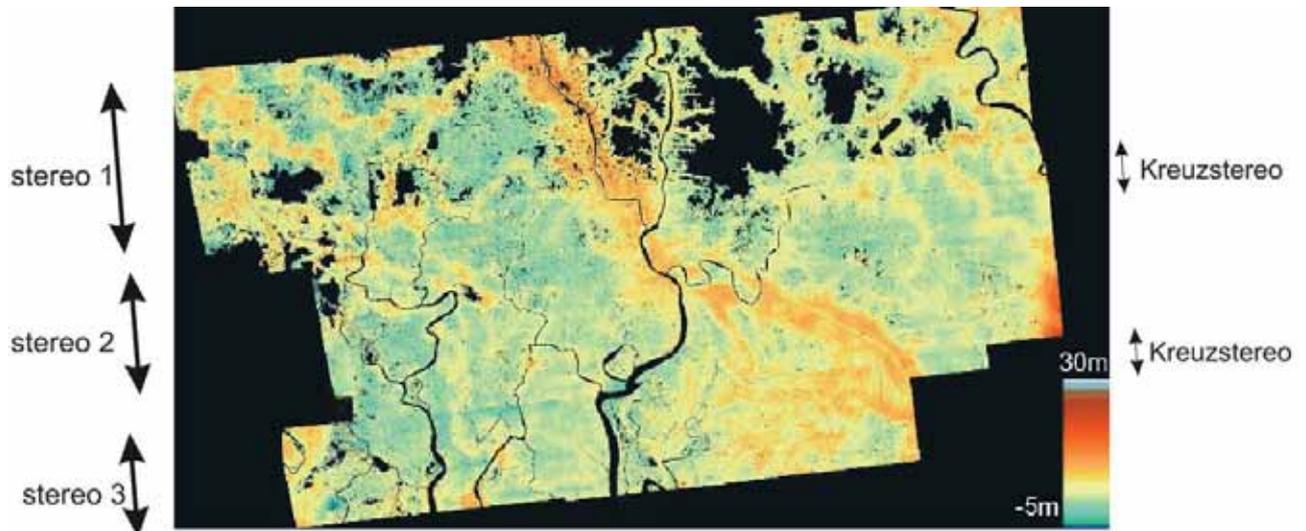


Abb. 14: Schließung von Lücken benachbarter Stereopaare durch kreuzweise Bildkombinationen – farbkodierte Höhendarstellung des Höhenbereichs von -5 m bis 30 m

## 5 Erstellung von Orthofotos

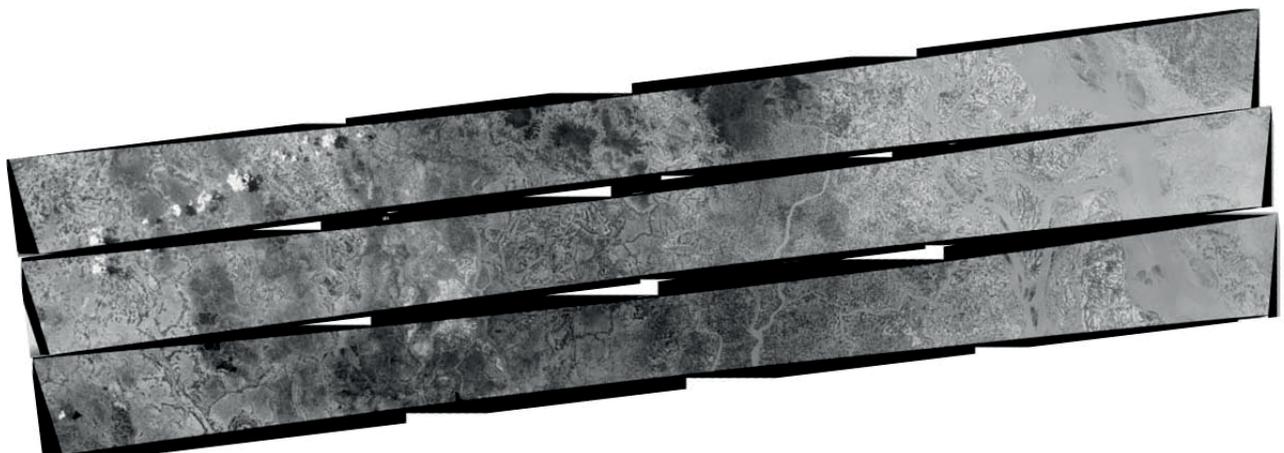


Abb. 15: Drei nebeneinander gelegte CORONA-Orthofotokombinationen – bestehend aus jeweils 4 Teilbildern

In den erstellten Höhenmodellen sind die aktuellen und früheren Wasserläufe gut erkennbar. Ergänzt wird es durch Orthofotos, die weitere Informationen liefern. Wie für die Erstellung der Höhenmodelle hat auch bei der Orthofotogenerierung berücksichtigt zu werden, dass die Bildorientierung in einem tangentialen Koordinatensystem vorliegt, das Ergebnis aber im nationalen Koordinatensystem gebraucht wird. Die Transformation vom vorgegebenen Orthofotopixel mit der Punkthöhe in geographi-

sche, geozentrische und tangentielle Koordinaten unter Berücksichtigung der systematischen Bildfehler, der panoramischen Umrechnung, der Seitenlinienkorrektur, mit der Bildorientierung in den korrespondierenden Bildpunkt zur Zuordnung der Grauwerte ist zeitaufwändig, aus diesem Grund wird mit Bildausschnitten gearbeitet deren Eckpunkte vollständig transformiert werden. Innerhalb der Bildausschnitte wird eine ebene Perspektivtransformation benutzt um die anderen Pixel umzurechnen. Auf diese Weise kann ein Orthofoto von einem CORONA-Teilbild mit  $33000 \times 12000$  Pixel auf einem normalen Laptop innerhalb von 2 Minuten erstellt werden. Das CORONA-Vorbereitungsprogramm schätzt die Eckkoordinaten des Bildes im nationalen Koordinatensystem ab um den Bereich für das zu erstellende Orthofoto wählen zu können.

Die Orthofotos werden separat für die 4 Bildteile eines CORONA-Bildes erstellt. Von Teilbild zu Teilbild gibt es keine Anpassungsprobleme da die Bildorientierung für das jeweils kombinierte CORONA-Gesamtbild berechnet wird. Da aber die Orientierung, einschließlich aller erforderlichen Korrekturen, für jedes Gesamtbild separat berechnet wird, gibt es Anpassungsprobleme zwischen Orthofotos benachbarter Bilder im Bereich der zuvor genannten, abgeschätzten Standardabweichung von etwa 10m im Mittelteil und 15m an den Bildenden. Diese Widersprüche werden durch lokale Nachbarschaftanpassung gelöst. In Abb. 15 ist erkennbar, dass die Orthofotostreifen mit etwa 14,9 km in der Mitte schmaler sind als mit etwa 18,3 km am Bildende entsprechend der Objektpixelgröße in Y von 1,77m bis 2,18m. Für die Orthofotos wurden 2 m Pixelgröße gewählt, dementsprechend ist die Bildqualität am Bildende kaum bemerkbar geringer, nur die extremen Bildenden des CORONA-Gesamtbildes sind wegen Unschärfen im Originalbild nicht so gut.

Es wurden Orthofotos für alle im Projektbereich vorhandenen CORONA-Bilder berechnet, aber Höhenmodelle nur für die engeren Bearbeitungsbereiche, somit konnten diese Höhenmodelle nicht für die Orthofotos verwendet werden. Das Gesamtgebiet ist allerdings sehr eben, so dass das TDM90 DOM ohne Probleme für die Orthofotoerstellung benutzt werden konnte. In weiten Bereichen reichte auch eine konstante Höhe aus.

## 6 Fazit

CORONA KH-4B Aufnahmen von 1972 stellen, mit etwa 2 m Objektpixelgröße und stereoskopischer Überdeckung, eine optimale Quelle für die Rekonstruktion des Zustands von vor 48 Jahren dar, wenn keine anderen Unterlagen vorhanden sind. Die Geometrie der CORONA Fotos von ca.  $75 \text{ cm} \times 5.5 \text{ cm}$  entspricht nicht unbedingt der von Messbildern, aber mittels der Seitenlinien und einer höheren Anzahl von Passpunkten, die von Google Earth stammten, konnte im mittleren Bereich der CORONA-Bilder eine Standardabweichung der Lage von etwa 10 m und im Bereich der Bildenden etwa 15 m erreicht werden, was für den Zweck des Projekts ausreichte. Die geometrischen Probleme stimmen mit denen in (SCHNEIDER et al. 2001) in China erreichten überein. Die Auflösung und das Basis-Höhenverhältnis von 1:1.85 führten zu einer relativen Standardabweichung der Höhe von etwa 3m, womit Flussläufe gut identifiziert werden konnten. Die von den Bilddeformationen beeinflussten absolute Höhenwerte konnte durch eine Anpassung auf das TDM90 DOM verbessert werden. Damit bleiben die morphologischen Details der CORONA-Höhenmodelle erhalten und die absolute Höhen Genauigkeit entspricht etwa 4 m bis 5 m. Mit diesem Verfahren konnten so Genauigkeiten der erstellten Produkte erreicht werden, die nicht weit von denen der mit Messbildern vergleichbarer Objektauflösung zu erzielen sind.

Die für die Handhabung der CORONA-Bilder erforderlichen Programme mussten, mit Ausnahme der Mess- und Bildzuordnungsroutinen, speziell erstellt werden, da kommerzielle Programme nicht die erforderlichen Schritte ermöglichten.

Mit 10 Stereopaaren konnten für etwa 25% der Landesfläche von Bangladesh Orthofotos erstellt werden, die für die heutige Stadtplanung von Bedeutung sind, da sie Bereiche problematischen Baugrunds anzeigen.

## 7 Literaturverzeichnis

- ALDOSARI, A. & JACOBSEN, K., 2019: Quality of Height Models Covering Large Areas. *PFG – Photogrammetrie, Fernerkundung, Geoinformation* **87**(4), 177-190, <https://link.springer.com/article/10.1007/s41064-019-00072-11>.
- AUELMANN, R.R., 2015: Corona KH-4B Satellites. <http://ftp.uni-koeln.de/pc/basp/Auelmann.pdf>, letzter Zugriff Dez. 2019.
- DASHORA, A., LOHANI, B. & MALIK, J., 2007: A repository of earth resource information – CORONA satellite programme. *Current Science* **92**(7), 2007.
- DOWMAN, I., JACOBSEN, K., KONECNY, G. & SANDAU, R., 2012: High Resolution Optical Satellite Imagery. Whittles Publishing, ISBN 978-184995-045-6.
- JACOBSEN, K., 2008: Tells the number of pixels the truth? – Effective Resolution of Large Size Digital Frame Cameras. ASPRS 2008 Annual Convention, Portland.
- MCDONALD, R.A., 1997: CORONA between the Sun & the Earth. ASPRS 1997, ISBN I-57083-041-X.
- SCHNEIDER, T., JACOBSEN, K., SEITZ, R. & FÖRSTER, B., 2001: Remote Sensing Based Parameter Extraction for Erosion Control Purposes in the Loess Plateau of China. *High Resolution Mapping from Space* 2001, Hannover, <https://pdfs.semanticscholar.org/1a86/53f9ff997c164e2f2572bdf6b6f8e7d33fe6.pdf>, letzter Zugriff Feb. 2019.

# Quality assessment of high-resolution UAV imagery and products

MICHAEL CRAMER<sup>1</sup>, SHUHANG ZHANG<sup>1</sup>, HENRY MEIßNER<sup>2</sup> & RALF REULKE<sup>2</sup>

*Abstract: The objective of this paper is to evaluate the performance of high-end and regular UAV-based camera systems. Different factors contribute to the overall accuracy. The evaluation partly relies on the methods, which are part of the new upcoming German standard DIN 18740-8 “Photogrammetric products – Part 8: Requirements for image quality (quality of optical remote sensing data)”. The image data quality in general is quantitatively evaluated at different processing levels e.g. uncorrected, corrected original image, influence of debayering, orthoimage processing, image restoration, etc. This requires pre-processing of the image data to produce a comparable data quality, the acquisition, provision and processing of reference data and additional information. This analysis includes the spatial resolution. Furthermore, the geometric camera stability and the influence of different image block constellations directly influences the overall 3D object point quality. This is typically evaluated from test sites. For empirical testing UAV-based images from the DJI Phantom 4 series with proprietary in-built cameras are compared to drone images taken with Phase One iXM 100 MPix camera.*

## 1 Introduction

Photogrammetry provides methods to obtain geometric and thematic information from data captured by different sensors i.e. cameras, radar and laser systems. The quality of the sensor data itself plays an important role for the quality of derived products. Defined quality requirements for the estimation of quality of remote sensing data are necessary. This is why, according to the German Standard series DIN 18740, a new standard for the estimation of optical remote sensing data is close to be published soon. This will be available as *DIN 18740-8, Photogrammetric Products – Part 8: Image quality requirements (quality of optical remote sensing data)*. Parameters describing image quality are of relevance in different application scenarios like

- sensor and mission design,
- comparison of sensors (using defined parameters and numbers),
- image quality estimation including sensor resolution,
- algorithm development,
- in-flight / in-orbit behaviour of the instrument,
- traceability to standardised parameters and instruments.

It is complex to find objective definitions for the parameters of image quality. This is why there is no general standard for describing image quality. In general, the requested quality of image data depends on the problem to be solved (e.g. point detection and co-registration, object recognition and tracking, classification in remote sensing, images related to the human visual system) and

---

<sup>1</sup> Universität Stuttgart, Institut für Photogrammetrie (ifp), Geschwister-Scholl-Str. 24D, D-70174 Stuttgart, E-Mail: [Michael.Cramer, Shuhang.Zhang]@ifp.uni-stuttgart.de

<sup>2</sup> Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Optische Sensorsysteme, Rutherfordstraße 2, D-12489 Berlin-Adlershof, E-Mail: [Henry.Meissner, Ralf.Reulke]@dlr.de

cannot be defined in the same way for every imaging system in general. It depends on factors like image content (structure), sensor performance, accuracy (geometric, radiometric, and spectral) and several other factors. The image quality can be determined at different processing levels. This requires pre-processing of the image data to produce a comparable data quality, the acquisition, provision and processing of reference data and additional information. With that, also the influence of image processing algorithms is evaluated. In order to objectively evaluate the quality of the sensor, representative image data is necessary. This is why defined artificial test patterns or suitable natural structures and features are imaged. For resolving power estimation sharp natural edges or artificial resolution charts like bar pattern or Siemens stars are typically in use. The image quality is maybe the most important part in any photogrammetric imaging and processing chain as it directly influences the tie point extraction and multi-stereo matching. This now can be evaluated following the definitions in the new upcoming standard DIN 18740-8.

The quality of 3D object points, as one main result in geometric reconstruction, is closely dependent on the individual image block geometry. This image acquisition fixes the overlap between neighbouring images and therefore is of importance for the final overall geometric accuracy. Any derived product, like dense point cloud and orthophotomosaic relies on this image block constellation and georeferencing. This also includes the quality of in-situ calibration to determine the internal sensor geometry containing any additional (optional) boresight parameters. This is not covered by the new standard, as it is, similar to the individual requests on image quality as mentioned above, application dependent. Each application will have its specific scenario and with that it is a problem to find standardized procedures to evaluate the performance of georeferencing and derived products. Within this context the availability and accessibility of official evaluation and calibration test sites is often mentioned and for sure, the use of highly accurate test sites is widely accepted and several campaigns already have been done to estimate individual sensors performance from test fields, i.e. the German DGPF Camera Test (CRAMER 2010). Still, this is typically only done for very few samples of a sensor series, in scientific context quite often only one single specific sensor is very deeply analysed. With that it is more or less a sensor type, i.e. sensor series than a specific individual sensor characterization. Even when flown in the same test site with same flight configuration, slightly different environmental conditions or different operators' expertise may already influence the processing result. This all does not consider the fact, that due to operational issues multiple defined evaluation test sites have to be installed, distributed all over the country to allow efficient overfly while systems are in air and on their way to operation. Due to such complexity, standardized test sites and test site evaluations are not yet considered in German standards, even though the US Geological Survey (USGS) some years ago started to put effort into such initiative and some freely accessible test sites have been installed and maintained in the US by now (USGS 2020). Three of these sites are designed for geometric evaluation of satellite and aerial sensors. The quality check is based on the evaluation of the final product, the orthophoto namely – this is different to the current methodology in DIN 18740-8, which focuses on the (original) image data. As mentioned in the objectives from USGS, “accurate and standardized high resolution orthophotos (reference image) will be used to determine the relative accuracy of other high resolution aerial and satellite images. Co-registration analysis will be performed between this image and other search images. The results of the analysis will provide insights into the accuracy

of the imaging system.” Another forth test site is especially prepared for the geometric in-situ calibration of satellites and aerial sensors.

Within the following sections, the evaluation of empirical UAV-based images will be done in two parts. The first one is following the concept as proposed in the DIN 18740-8 standard. A Siemens star pattern is captured in several overflights from which the spatial resolution is derived in a defined way. The software used here will become official part of the future standard. In a second step, the overall 3D object point performance is obtained from check point analysis. Highly precise reference points have been signalized in a local test field. In order to compare different level of UAV sensors, two different systems are selected. There is a large variety of operational camera UAVs in the market, many of them especially designed for survey and photogrammetric mapping applications. Additional RTK-GNSS functionality for the precise trajectory implementation and measurement is common for the professional systems and can help to obtain the geo-data in a more effective way. The DJI Phantom 4 series with its proprietary in-built cameras represents a very popular drone system increasingly used for 3D mapping. In addition to this, larger professional mapping drone platforms are able to lift up heavier sensors, like latest generation Phase One iXM cameras providing image images up to 150 MPix. Their performance will be characterized and compared in the following.

## 2 Sensors and test data

The DJI Phantom 4 and Phantom 4 RTK are all consumer-level quadcopters manufactured by DJI (DJI 2019b; DJI 2019c). The Phantom 4 model is designed for daily aerial image and video capture, while the Phantom 4 RTK is designed for aerial mapping. Compared with professional mapping drones, the Phantom quadcopters are lighter in weight and cheaper in price; both systems are all within 1.4 kg. The compact design and lightweight advantages of Phantom systems allow for one-man operation, which makes them more cost-effective. Both systems have ca. 25 minutes flying time in normal cases. The Phantom 4 is equipped with a consumer-grade GNSS receiver, inertial measurement unit (IMU), and barometer for performing autonomous flights. It can achieve accuracy up to  $\pm 1.5$  m vertical and  $\pm 0.5$  m horizontal positioning accuracy (DJI 2019c). The Phantom 4 RTK has a multi-frequency multi-system high-precision RTK GNSS that can achieve positioning accuracy up to 1 cm+1 ppm horizontal, 1.5 cm+1 ppm vertical (DJI 2019b).

Both DJI Phantom systems have integrated gimbal cameras with 3-axis stabilization that can compensate minor vibration or instability of the drone during mapping flights. For a detailed comparison, parameters of the cameras are listed in Table 1. Both cameras are proprietary and have fixed focal length, which leads to better stability of interiors and therefore more ‘metric-like’ (CRAMER et al. 2017). The camera of Phantom 4 RTK shares similar specifications compared with that of Phantom 4, but the increase in sensor dimensions and resolution is significant, which leads to higher resolving ability. In addition, the Phantom 4 RTK camera has a mechanical shutter that avoids the rolling shutter effect to further increase the imaging quality. The ground controlling software of these two systems differs. Although Phantom 4 is not designed for aerial mapping, it has an official flight planning software called DJI GS Pro (DJI 2019a). It can plan nadir/oblique-imaging missions according to given ground sampling distance (GSD) and overlap automatically. The Phantom 4 RTK has a built-in software DJI GS RTK, it is similar to DJI GS Pro but with

additional RTK management features. Additionally, the factory-calibrated interiors and measured exteriors are stored in the XMP data of images, they can be imported to data processing software like Agisoft Metashape or Pix4DMapper directly.

Different to the DJI integrated system solution, the Phase One cameras in principle can be integrated in any type of professional drone with sufficient maximum take-off mass (MTOM) and corresponding interfaces. As the growing UAV mapping market is foreseen as one of the most important field of application for this camera, the iXM-RS already provides fully integrated solutions for the integration with existing drone platforms like the DJI M600 Pro aerial platform. The camera offers different optional lenses with different focal length and characteristics. The selected 35 mm lens is a fixed focus lens, focussed to infinity. The corresponding depth of field starts at 30 m minimal distance with maximum opening of aperture. This has to be considered when setting the flying height above ground. However, these fix focus lenses can also be set to a fixed distance, non-equal to infinity, which is equivalent to the mean flying height above ground for example. This must be adjusted at the factory and it is not possible for the user to later change this setting. In addition Phase One also offers so-called ‘motorized’ lenses, where the focus can be changed and a precise mechanism allows to repeat defined focus positions with a repeatability better than 6  $\mu\text{m}$  (WIESER 2018). The iXM100-RS uses back-side-illuminated CMOS sensors, which should allow for higher dynamic range, improved sensitivity and less noise despite smaller pixels compared to the Phase One iXU predecessor models. Details can be found in Tab. 1 and PHASE ONE (2020).

Tab. 1: Camera parameters of DJI Phantom 4 / 4 RTK and Phase One iXM100-RS

Camera	Phantom 4	Phantom 4 RTK	iXM100-RS
Sensor format [mm] / [pix]	6.17 × 4.55 / 4000 × 3000	13.2 × 8.8 / 5472 × 3648	43.9 × 32.9 / 11664 × 8750
Pixel Count	12 Mpix	20 Mpix	100 Mpix
Pixel pitch [ $\mu\text{m}$ ]	1.56	2.41	3.76
Focal length [mm] / 35mm equiv. focal length [mm]	3.6 / 20	8.8 / 24	35 / 28 (other lenses available)
Field-of-View (FOV)	94°	84°	76°
ISO range	100-1600	100-12800	50-6400
Shutter speed [s]	8 - 1/8000	8 - 1/2000	Up to 1/2500
Shutter type	Rolling	Mechanical	Mechanical
Focus	fixed at $\infty$	AF/MF	Fixed at $\infty$ (for selected lens)
Max f-stop	1:2.8	1:2.8	1:5.6
Image format	DNG (raw), JPG	JPG only	IIQ (raw)

For data acquisition, the camera was integrated in the CS-SQ8 multi-copter drone provided by CopterSystems. The gimbal mount was especially designed and optimized for this iXM camera. In order to get precise GNSS/inertial trajectory information, the Applanix APX-15el UAV board was additionally integrated and time synchronized with the camera exposures. This unit contains a multi-frequency multi-system high-precision RTK GNSS and a dual IMU in order to provide support for the movements of the mount with respect to the copter airframe. The post-processed positioning accuracy is around 2-5 cm, the roll and pitch angle accuracy around  $0.025^\circ$  and the heading accuracy is about  $0.08^\circ$  according to the manufacturers' specifications.

### 3 Quality evaluation

#### 3.1 Geometrical resolution

As motivated above, the quality is analysed in two different ways, namely from original images and from derived products. For estimating the resolution of the original images, a software is used, which will become publicly available as part of the upcoming DIN 18740-8 standard. This software uses the modulation transfer function (MTF) which is the spatial frequency response of an imaging system to a given illumination and therefore estimates the effective image resolution or resolving power, which then can be transformed into object space to get so-called ground resolving distance (GRD) or true ground sampling distance (tGSD). Using the a priori knowledge of the original scene – in this case the well-known Siemens star target – contrast transfer function (CTF), MTF and point spread function (PSF) are approximated by a Gaussian shape function. The mathematical foundations and further details can be seen from MEIBNER et al. (2018, 2019). The parameter  $\sigma$  (standard deviation) of the PSF (assuming a Gaussian-shape function) is one criterion for spatial resolution. It directly refers to image space and can be seen as objective measure to compare different camera performances. Another criterion is the spatial frequency where the MTF reaches a certain (minimal-)value (i.e. 10%, MTF10). The reciprocal of that frequency is the approximation for the size of the smallest line per pixel. This MTF10 value will be used for quality estimation here. As a side note, values for MTF at 10% modulation contrast are obtained as piecewise linear interpolation between adjacent measurements. Alternatively, a polynomial- or Gaussian-fit may be applied but it can be shown that both fit-approaches use all measurements as input and if these are falsified (e.g. due to inhomogeneous target illumination or sharpening artefacts) approximated MTF10 values may vary substantially.

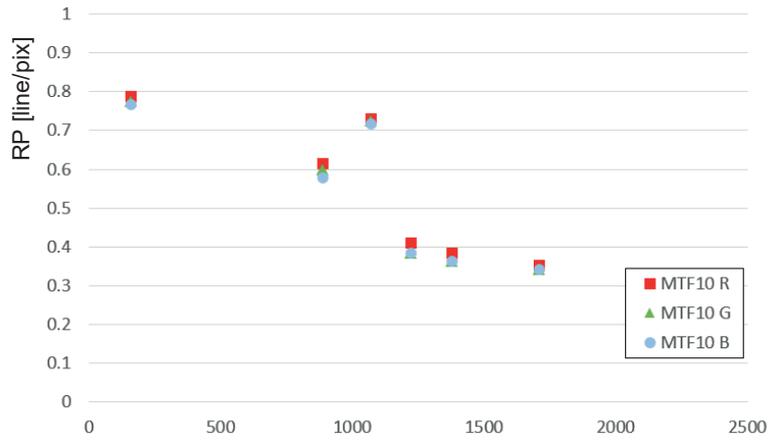
The Siemens star images have been taken during regular mapping flights. The target was fixed on ground while the drone was flying the planned test. Placing the target in a multiple-overlap area, due to several overlapping flight lines, the star will be imaged multiple times from different camera stations. Thus, the star will be located at different positions in image plane. Each imaged star allows for the estimation of the resolving power at this distinct position in image plane. The Fig. 1 shows three samples of imaged Siemens stars taken by the three different cameras. All samples were selected in a way that the Siemens star is almost in the centre of the image. This should give an estimation of best possible resolving power, which is supposed to decrease from centre to image corners. In all cases jpg-images have been analysed, as this is the mostly preferred image data format for most UAV scenarios, even though the possible influence of compression effects is well-known.

As shown from Fig. 1 the Siemens star has a special design. It consists of 32 segments (16 black and 16 white), 4 normalization patterns (2 black and 2 white), 4 rotor shaped external markers, 4 circular shaped external markers and a black square to extract the star's 2D-orientation in image plane. The four normalization patterns (each 2 black and white circles for redundancy if significant noise is present) are used to set maximum and minimum values during determination of spatial resolution. This is not only necessary but very helpful if the imaged Siemens star is too small. The 4 rotor shaped and / or circular shaped external markers are used to find the correct star-centre position which is crucial for correct results. Furthermore, these markers are used to obtain the homographic transformation matrix to correct test target inclination (especially in image corners) if required.

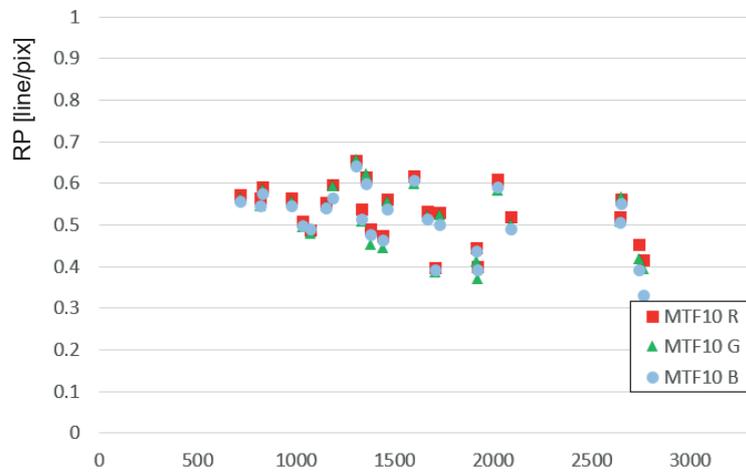
As the GSD is different for each flight, the size of the star in the image is different. The specific image from Phantom 4 shown here, is taken from 28 m flying height above the Siemens star pattern, corresponding to a local GSD of 1.20 cm. Different to this, the selected Phantom 4 RTK image has been captured from 25 m above Siemens star. This is the minimal flying height above ground, which can be realized in Phantom 4 RTK pre-planned automatic mission flights. Due to the smaller sensor pixel size the corresponding GSD is 6.9 mm, which is the minimal GSD due to the minimal flying height limits above. When flown in manual mode no such limits are available. Interesting to note, that the Phantom 4 does not have such limitation. The shown Phase One iXM-100 image was taken from 47 m above Siemens star, corresponding to a GSD of 5.0 mm. The results from MTF analysis are also shown in this figure. The MTF10 value given in line/pix refers to the image space. In order to transform this measure to the object space value the reciprocal of MTF10 is multiplied with the nominal GSD. The corresponding GRD values are given in the figure. If one compares the GRD to GSD, the factors are 128%, 178% and 130% for Phantom 4, Phantom 4 RTK and iXM100-RS respectively. Obviously Phantom 4 provides the same relative resolution as iXM100-RS for the centre image region, but the different absolute numbers and the fact, that Phantom 4 images has be sharpened in internal DJI image processing, has to be noticed.



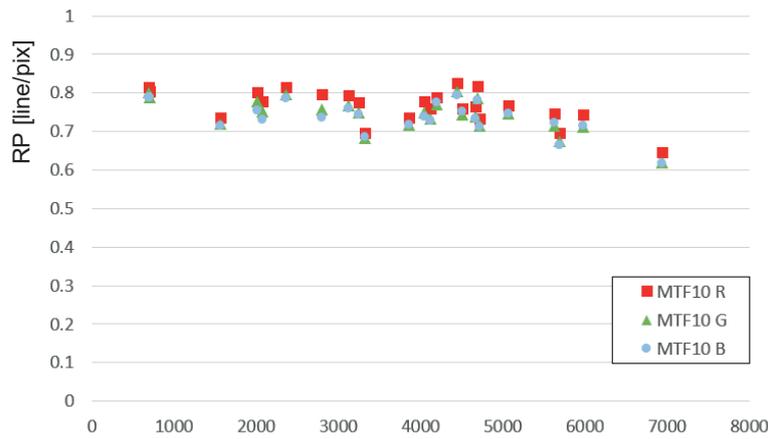
Fig. 1: Siemens star image and analysis from Phantom 4 (left), Phantom 4 RTK (middle) and Phase One iXM100-RS (right). Each image sample is sized 300 x 300 pix.



Phantom 4 – MTF10



Phantom 4 RTK – MTF10



iXM100-RS – MTF10

Fig. 2: Resolving power (RP) of tested cameras give for the three colour channels. MTF10 values [line/pix] are plotted against the radial distance from image centre to image corners given [pix]. Notice the different sensor formats.

A comparison of each camera regarding spatial resolution in relation to distance of image centre is given in Fig. 2. The iXM100-RS with its huge sensor format and 100 megapixels shows very low variation in resolving power over the complete image space (Fig. 2, bottom). Imaging performance of both DJI systems (Fig. 2, top and centre) is fairly homogenous but the loss of spatial resolution towards image corners is significantly higher in comparison to the iXM100-RS sensor-optic combination. Unfortunately, only six images are available for the Phantom 4 resolution determination.

Phantom 4 RTK images have been used without DJI internal pre-processing, i.e. without sharpening and pre-distortion corrections as implemented in the DJI raw image processing. This causes quite strong distortions in the image borders and especially those Siemens star targets imaged there show strong ellipsoidal deformation. The homographic transformation has been applied to correct the target's distortion. This transformation however is not bijective (HARTLEY & ZISSERMAN 2004) and thus, the transformed measurement coordinate system is not absolutely accurate and obtained values contain noise. While the overall measurement accuracy of spatial resolution is in the range between 5 % – 7 % (iXM100-RS), it rises to approximately 10 % - 12 % (Phantom 4 RTK) when homographic transformation is applied.

Two observations have to be mentioned at this point. First, exposure-time usually is set to obtain enough information (in form of signal strength) even in darker, shadowed areas. The Siemens star target with its (ideal) black and white segments then often is over-exposed which influences the measurement of spatial resolution. The Siemens stars in Phantom 4 RTK images show this effect, as well as some of the Phantom 4 images. In addition, images of Phantom 4 have been sharpened in the internal DJI image pre-processing. This is the reason, why spatial resolution in Phantom 4 RTK in general is lower compared to the Phantom 4 imagery.

Second, if the test target is too small in the corresponding images the selected target design is crucial to normalize the data during measurement process. Alternatively, the used target and corresponding diameter needs to satisfy the requirement that each segment (black or white) of the outermost ring (low spatial frequency) is covered of at least three pixel. This then ensures unmingled values for black and white. Both issues can be solved when planning and executing a dedicated test flight for spatial resolution determination. Especially the optimal setting of exposure parameter to get optimal Siemens star images is not feasible in many practical scenarios. Regarding the size of the Siemens star in the images, the resolution target diameter itself can be increased, but this limits the mobility and flexible use of the target.

### 3.2 3D object point determination

The performance of 3D object point determination is derived from extended aerial triangulation, which includes camera self-calibration and (optional) directly measured GNSS/inertial exterior orientation elements. Data is captured in a test site, which provides significant number of signalized and highly precise coordinated 3D object points. These points are considered as control and/or check points. The test field for the following experiments was the ship lock at Hessigheim, Neckar river, about 30 km north of Stuttgart, which is sized about  $200 \times 70 \text{ m}^2$ . The site was established as part of a research and development project initiated by the German Federal Institute of Hydrology (BfG) in Koblenz in partnership with the Office of Development of Neckar River Heidelberg

(ANH). The 3D monitoring of the ship lock infrastructure and its surrounding area is the underlying application. Different to traditional monitoring, where classical terrestrial survey methods are applied, the 3D structure should be precisely mapped from UAV data, in order to move from point to surface capturing. This is why the test site was established with extremely accurate 3D signalized object points. Checkerboard discs with approximately 30 cm diameter are used for signalization. Their 3D point accuracy is within 1 mm (std.dev.). The Fig. 3 depicts the test area and the distribution of signalized points.



Fig. 3: Test area of ship lock Hessigheim/Neckar with distribution of signalized points. The image on the right shows one of the signalized points. All points are fixed to the ground except three, which are mounted on top of approximately 140 cm high concrete pillars (red squares).

Several tests have been flown in this site, only a small sub-set will be presented in the following, the main flight parameters are given in Tab. 2. Only three control points have been introduced, all remaining points are for check only, in order to have efficient operational scenario. Regular blocks are compared to cross-pattern flight scenarios. In all cases the additional self-calibration parameters are estimated from zero initial values as part of the bundle adjustment. The corresponding accuracy as obtained from 15-17 check points (not all of the points were available in all the different flight campaigns) are given in Tab. 3. The accuracy is given in absolute numbers and is also referred to individual GSD of flight test.

In flight test #1, the result is only GCP-based, as the on-board GNSS of Phantom 4 only provides navigation grade initial camera exteriors. The horizontal RMS is consistent and all within one pixel, while the vertical RMS reached 2.2 pixel. The distribution of Z error of check points shows a ‘bowling effect’ – the error in the centre of the survey area was positive while it became negative at the border. This is the typical situation for minimal GCP constellations with no perspective centre coordinate measurements and additional self-calibration. Using a denser GCP distribution, the solution stabilizes. In an extra test, based on a dense control point distribution using eleven GCPs, the RMS Z dropped to 0.7 pixel and the RMS 3D is 1.17 pixel.

Tab. 2: Tested flight constellations, all versions using 3 GCPs only.

#	System	Block design	GSD (cm)	Number of images	Overlaps & Remarks
1	Phantom 4	Regular 4 NS lines	1.29	64	long-track 70%, cross-track 76%  GCP-based only
2	Phantom 4 RTK	Regular 8 EW lines	1.29	82	long-track 80%, cross-track 60%  with RTK-GNSS
3	Phantom 4 RTK	Regular 6 NS lines	0.68	298	long-track 80%, cross-track 60%  with RTK-GNSS
4	Phantom 4 RTK – CalibFlight	Calibration flight scenario  6 NS lines / 8 EW lines	0.68 NS / 1.29 EW	388	Overlaps as in test #2 & #3  with RTK-GNSS
5	iXM100-RS – Cal- ibFlight	Calibration flight scenario  4 NS lines / 11 EW lines	0.4 NS / 0.5 EW	148	NS: long 70%, across 50% EW: long 77%, across 57%  with post-processed GNSS/inertial positioning from APX-15el UAV board

Tab. 3: 3D object point performance from check point differences, all versions using 3 GCPs only.

#	System	RMS X [cm] / [pix]	RMS Y [cm] [cm] / [pix]	RMS Z [cm] [cm] / [pix]	RMS 3D [cm] / [pix]
1	Phantom 4	0.93 / 0.7	0.92 / 0.7	2.90 / 2.2	3.19 / 2.4
2	Phantom 4 RTK	0.62 / 0.4	0.93 / 0.7	4.46 / 3.4	4.59 / 3.5
3	Phantom 4 RTK	0.31 / 0.4	0.50 / 0.7	0.80 / 1.1	1.00 / 1.4
4	Phantom 4 RTK – CalibFlight	1.11 / 1.6	1.70 / 2.5	1.81 / 2.6	2.73 / 4.0
5	iXM100-RS – Cal- ibFlight	0.14 / 0.3	0.23 / 0.5	0.19 / 0.4	0.33 / 0.8

The processing of Phantom 4 RTK data in test #2 and #3 takes the initial exteriors from the RTK-GNSS. The two flights are with different strip direction and GSD. The flight #2 is ‘cross-strip’, i.e. with strips in east-west direction, perpendicular to the long side of the rectangular survey area. It has eight strips and nine images per strip; the GSD is equivalent to the Phantom 4 flight test #1. In test #3 the flight direction is in north-south axis, similar to test #1, but it has six strips and 47 images per strip, because of the smaller GSD. The ‘cross-strip’ pattern of #2 weakens the connection of image block, as more images were connected side-by-side and the side-lap is lower than

the front-lap. This mainly influences the vertical RMS, which is significantly larger in test #2 compared to test #3. However, flight test #3 shows consistent results in all axes. In terms of absolute numbers, the accuracy obtained here is best, as this test provides the smallest GSD. Relating the accuracy with respect to the pixel size, horizontal accuracy is comparable to the previous flight tests #2 and #3, still the vertical accuracy RMS is significantly better around 1.1 pixel. This proves the positive influence of high-quality GNSS perspective centre coordinates in combination with a sufficient image block geometry.

It is worth noting that the RMS 3D of flight test #4 is worse than #3 and #2. The two flying heights in cross-pattern should have strengthened the image block geometry, and additional RTK-GNSS observations should have increased the precision as well. However, both horizontal and vertical RMS do not increase. There may exist systematic error in the GNSS observation that disagree with the image block geometry. Additionally, when introducing ten more GCPs, the overall RMS decreased to 3.4 pixel, as more GCPs fixed the image block in a stronger way. Still, this result does not satisfy expectations and needs further investigations.

Flight test #5 with the Phase One iXM100-RS camera and post-processed GNSS/inertial positioning is the best in absolute numbers as well as in pixel. This in a way was expected, as the much larger sensor format allows less images to get the same project area captured with similar overlap, even though the GSD is again smaller than the smallest Phantom 4 RTK GSD. Interesting to see that, comparing test #3 with #5, the Phantom 4 RTK can achieve similar horizontal accuracy. Still, vertical accuracy exceeds the result in #3 by a factor of two.

## 4 Summary and Conclusions

This paper exemplarily highlights two methods to evaluate the quality of UAV-based or other airborne sensor image data. The methods start with the resolution testing of images using defined resolution charts, the overall geometric performance is then derived from test site with sufficient number of highly accurate check point information. The first step is using a software, which comes as part of an official DIN standard soon, and gives objective, comparable and transferable results as the testing scenario is clearly defined. Results have shown that the image pre-processing has significant influence on the estimated spatial resolution, in addition to the camera-lens combination itself. It therefore has to be mentioned, in which processing level the resolution testing has been done. In principle the resolution of derived products like orthoimage can also be analysed. The typical sharpening of raw imagery, implemented in the manufacturers' internal raw data processing has to be mentioned here. In our example the spatial resolution of Phantom 4 sharpened images was better than the Phantom 4 RTK unsharpened images, which has to be expected. In general, sharper images contain sharper details and more information, this is why many images are internally sharpened before the image is provided to the user. However, it has to be investigated how this sharpening influences the subsequent tie point transfer, especially when it comes to later dense image matching.

Additional environmental conditions, like platform vibrations, motion blur or atmospheric conditions also play a role. These are mission dependent, nevertheless, when resolution testing is done within the mission itself, this gives a quality measure from the same data used for the application.

The advantage of this method is the flexibility as the defined resolution target can be provided in each project area with minimal effort.

The testing of 3D object point quality from test sites is well established. Still, as discussed earlier, it is quite complex to define and install test sites, which provide very similar test environments. Special care has to be laid on the high quality of reference object data. In addition, the area should provide enough texture, which is helpful for the later tie point transfer. This is the essential part in bundle adjustment. From this point, the selected ship lock area is not optimal, even though it provides numerous reference points, as the surrounding water areas prevent reliable tie point transfer, especially when camera only has smaller image formats. Additional depth information in the test field support the in-situ camera calibration. This is easier to reach for UAV-applications as the factor between flying height above mean terrain level and height variations in object space is much smaller compared to classical large area airborne mapping. Assuming a mean flying height of 40 m, a height variation of only 4 m already corresponds to 10 % of the flying height. This is easy to reach in every test area. Flying in 2000 m above ground, the same 10 % height variation already corresponds to 200 m height difference, which already is a significant height jump.

It was intention of the paper to compare the performance of standard, more low-cost UAV-camera systems to the high-end camera based UAVs. As shown before, some differences are visible from the empirical check point accuracy. Although DJI systems are well-built and easy-to-use, some drawbacks still exist compared to professional unmanned aerial photogrammetric systems. First of all, the systems are close-sourced, which means that users cannot get access to the raw observations and logs to perform further processing to increase accuracy. For example, users are not allowed to get access to the raw image data but only JPG of Phantom 4 RTK. The JPG image may have already been ‘corrected’ by the firmware so that it does not fit the camera distortion model (PEPPA et al. 2019). Other limitations by the manufacturer may include the minimum flying height and the adjustable camera parameters. Users usually have to set the exposure settings empirically and adjust them during the flight to achieve an optimal image quality. However, the aperture of Phantom 4 RTK camera is not adjustable during the image capture, which may introduce additional problems. On the other hand, both cameras have large FOV and significant distortion. This may cause the actual image overlap being smaller than the planned one. Moreover, these cameras require careful in-situ calibration (CRAMER et al. 2017), despite that DJI provides the lab camera calibration result of each Phantom 4 RTK. The interaction of camera calibration parameters and directly measured GNSS perspective centre coordinates with the tie point observations and ground reference data is essential. Since DJI systems are compact, the possibility of hardware upgrading is zero — users cannot change any of the components of Phantom 4 or Phantom 4 RTK.

It was already expected that Phase One iXM100-RS performs better in several aspects. What should be highlighted is the consistent resolution of the tested camera-lens combination, with almost no decrease when it comes to the image borders. The true GSD is close to the nominal GSD throughout the whole image plane. Larger image format helps to cover given areas with less images per strip which gives a stronger geometry, especially for low number of GCPs. Even though the tested camera was lab-calibrated right after the empirical test, using these lab calibration parameters did not deliver highest object point accuracy. Therefore, similar to the DJI systems, the in-situ calibration is mandatory. In principle, an additional lab calibration is not of need at all, but

might be used to provide good initial values for the in-situ calibration. One of the biggest advantages is the accessibility of raw camera data. With the given software, the user has full access to the pre-processing of images. In general, the overall flexibility is much higher compared to the DJI integrated systems. The iXM offers different lens options and provides defined interfaces for other platforms and combinations with other sensors, like aiding navigation sensors. Still, in order to realize such multiple system integrations, the integration and fine-tuning of all the sensor-platform interactions is under the full control but also responsibility of the user, which requests for some hardware experience.

In the end, it is always up to the user to select between low-cost versus high-end, compact versus mid-format, close-sourced versus flexible UAV solutions according to the specific application. The proper selection of the UAV imaging system in high-resolution mapping highly relies on the quality and accuracy that users want to achieve. The in-field image resolution testing can help users to better understand the imaging quality, which is as necessary as the in-situ camera calibration. The current version of software to estimate the spatial resolution of imagery in a defined way according to DIN 18740-8 together with the appropriate template of the Siemens star can be made available to interested readers. Please contact the main author of this paper.

## 5 Acknowledgments

Special thanks has to be expressed to the German Federal Institute of Hydrology (BfG) in Koblenz financing the research project on area wide 3D UAV-based monitoring, which makes the test site Hessigheim possible. The Office of Development of Neckar River Heidelberg (ANH) kindly supports the terrestrial measurements and all on-site work. Applanix kindly provided the APX-15el UAV GNSS/inertial board for the Phase One camera flights and delivered the processed trajectory information for Phase One iXM100-RS flights. All this support is kindly acknowledged!

## 6 Bibliography

- CRAMER, M., 2010: The DGPF-Test on Digital Airborne Camera Evaluation – Overview and Test Design. *Photogrammetrie – Fernerkundung – Geoinformation (PFG)*, **2010(2)**, 75-84.
- CRAMER, M., PRZYBILLA, H.-J. & ZURHORST, A., 2017: UAV Cameras: Overview and Geometric Calibration Benchmark. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **42(2/W6)**, 85-92. doi: 10.5194/isprs-archives-XLII-2-W6-85-2017.
- DJI, 2019a: DJI GS Pro, Mission-Critical Flight Simplified. DJI Innovation, <https://www.dji.com/uk/ground-station-pro>.
- DJI, 2019b: Phantom 4 RTK Specs. DJI Innovation. <https://www.dji.com/uk/phantom-4-rtk/info#specs>.
- DJI, 2019c. Phantom 4 Specs. DJI Innovation, <https://www.dji.com/uk/phantom-4/info#specs>.
- HARTLEY, R. & ZISSERMAN, A., 2004: *Multiple View Geometry in Computer Vision*. Second Edition. 2<sup>nd</sup> edition, Cambridge University Press, ISBN: 0521540518.
- MEIBNER, H., CRAMER, M., & REULKE, R., 2018: Towards standardized evaluation of image quality of airborne camera systems. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **42(1)**, 295-300, doi: 10.5194/isprs-archives-XLII-1-295-2018.

- MEIBNER, H., CRAMER, M. & REULKE, R., 2019: Evaluation of structures and methods for resolution determination of remote sensing sensors. *Advances in Image and Video Technology, Proceedings of 9th Pacific-Rim Symposium on Image and Video Technology*, November 18-22, Sydney, Australia, Springer, Lecture Notes in Computer Science (LNCS) **11994**, 59-69.
- PEPPA, M. V., HALL, J., GOODYEAR, J. & MILLS, J. P., 2019: Photogrammetric Assessment and Comparison of DJI Phantom 4 Pro and Phantom 4 RTK Small Unmanned Aircraft Systems. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **42**(2/W13), 503-509, doi: 10.5194/is-prs-archives-XLII-2-W13-503-2019.
- Phase One (2020): Phase One iXM-100 specifications. [https://industrial.phaseone.com/iXM\\_Camera\\_Series.aspx](https://industrial.phaseone.com/iXM_Camera_Series.aspx).
- USGS, 2020: US Geological Survey Online Test Site Catalog. [https://calval.cr.usgs.gov/apps/test\\_sites\\_catalog](https://calval.cr.usgs.gov/apps/test_sites_catalog).
- WIESER, C., 2018: Personal correspondence, 26.6.2018.

# Untersuchungen zur Qualität des Realtime Kinematic GNSS Systems der DJI Phantom 4 RTK

HEINZ-JÜRGEN PRZYBILLA<sup>1</sup> & MANFRED BÄUMKER<sup>1</sup>

*Zusammenfassung: Das UAV DJI Phantom 4 (und Vorläufer) ist seit mehr als 10 Jahren im Markt verfügbar und ist in seiner aktuellen Version „RTK“ mit einem 2-Frequenz GNSS-Empfänger ausgestattet. In Verbindung mit einer Referenzstation oder alternativ z.B. durch die Nutzung des SAPOS-Dienstes ist eine präzise Positionsbestimmung im Realtime Kinematic Modus möglich. Da das System auch Rohdaten im RINEX-Format zur Verfügung stellt, kann grundsätzlich auch eine Positionsbestimmung mittels PPK (Post-Processed Kinematic) durchgeführt werden. Dem Anwender bieten sich somit erweiterte Möglichkeiten zur Georeferenzierung von UAV-Bildflügen.*

*Im Rahmen des Beitrags erfolgen Untersuchungen zur geometrischen Genauigkeit von UAV-Bildverbänden, die mit vier verschiedenen DJI Phantom 4 RTK Systemen unter fest definierten Bedingungen auf dem UAV-Testfeld „Zeche Zollern“ geflogen wurden. Die nachfolgenden Auswertungen erfolgten mit identischer Parametrisierung durch die Software Agisoft Metashape. Die vorliegenden Ergebnisse zeigen die zum Teil starken Variationen in der Qualität der gemessenen RTK-Positionen sowie deren Auswirkungen auf die Bildorientierung und beteiligte Parameter. Ein abschließender Vergleich zwischen der Nutzung der RTK-Messungen und solchen aus einem Post-Processing (PPK) zeigt für die untersuchten Bildverbände keinen erkennbaren Genauigkeitsgewinn.*

## 1 Einleitung

Unmanned Aerial Vehicle (UAV) erfreuen sich in der geodätisch-photogrammetrischen Community einer steigenden Beliebtheit. Der Markt – mit von diversen Herstellern angebotenen Komplettsystemen – wächst stetig. Hierbei kommt dem chinesischen Anbieter Da-Jiang Innovations Science and Technology Co., Ltd (DJI) die Markführerrolle zu, mit einem derzeitigen Anteil von ca. 70% des globalen Consumer-Drohnen Marktes (HANDELSBLATT 2020). Das Modell Phantom 4 (und Vorläufer) ist seit mehr als 10 Jahren verfügbar und ist in seiner aktuellen Version „RTK“ mit einem 2-Frequenz GNSS-Empfänger ausgestattet (Abb. 1).



Abb. 1: DJI Phantom 4 RTK

<sup>1</sup> Hochschule Bochum, Fachbereich Geodäsie, Lennershofstr. 140, D-44801 Bochum, E-Mail: [Heinz-Juergen.Przybilla, Manfred.Baeumker]@hs-bochum.de

In Verbindung mit einer Referenzstation oder alternativ durch die Anbindung per NTRIP (Network Transport of RTCM via Internet Protocol) über Mobilfunk oder Wi-Fi-Hotspot ist z. B. die Nutzung des SAPOS-Dienstes (SAPOS 2020) möglich, und somit eine präzise Positionsbestimmung in Echtzeit. Die Angaben des Herstellers (DJI 2020) zur Positionierungsgenauigkeit betragen:

- Vertikal: 1,5 cm + 1 ppm (RMS)
- Horizontal: 1 cm + 1 ppm (RMS)

Da das System auch Rohdaten im RINEX-Format (Receiver Independent Exchange Format) zur Verfügung stellt, ist grundsätzlich auch eine Positionsbestimmung mittels PPK (Post-Processed Kinematic) möglich. Die GNSS-Rohdaten eines Zweifrequenzempfängers (Code- und Trägerphasenbeobachtungen sowie die Ephemeridendaten) sind die Grundlage für eine nachträgliche Auswertung, die in der Regel zur Bestimmung von Positionslösungen höherer Genauigkeit führt. Dem Anwender bieten sich somit erweiterte Möglichkeiten zur Georeferenzierung der UAV-Bildflüge. Neben der (indirekten) Lagerung auf Bodenpasspunkte (Ground Control Points – GCP) ist sowohl eine direkte Lagerung auf die gemessenen Bildpositionen (Äußere Orientierung – ÄORI) als auch eine Kombination der beiden Ansätze (integrierte Orientierung) möglich.

Im Rahmen des Beitrags erfolgen Untersuchungen zur geometrischen Genauigkeit von UAV-Bildverbänden. Hierzu wurden 4 verschiedene DJI Phantom 4 RTK Systeme auf dem Areal des Präzisions-UAV-Testfelds des Industriemuseums „Zeche Zollern“ in Dortmund geflogen (Lage- und Höhengenaugigkeit der Passpunkte ca. 2 mm; PRZYBILLA et al. 2018). Alle Bildflüge erfolgten mit einer fest definierten Aufnahmekonfiguration (Kreuzverband mit 20% Höhendifferenz, Längsüberdeckung 80%, Querüberdeckung 60%, GSD 14 mm, manuelle Fokussierung (MF) der Kamera auf unendlich) an drei unterschiedlichen Tagen.

## 2 RTK-Bildflug

Planung und Durchführung des Bildflugs mit einem DJI RTK-Kopter-Systemen werden mittels einer in sich geschlossenen Software des Herstellers (DJI GS PRO) durchgeführt, die auch die Festlegungen für den Bezug der GNSS-Korrekturdaten enthält (Abb. 2).



Abb. 2: Definition der RTK-Einstellungen in der GS PRO Applikation (Fotos: Hersteller)

Während des Bildfluges werden für jedes Bild eine Bildnummer, ein Zeitstempel, die Komponenten des Hebelarms zwischen Antennenzentrum und der Bildmitte des CMOS-Sensors, die vollständigen Positionsdaten (im WGS84 bzw. ETRS89 bei Nutzung des SAPOS-Dienstes HEPS), zugehörige Genauigkeitsangaben sowie der RTK-Status protokolliert (Abb. 3).

Anzumerken ist hier, dass die Definition des Hebelarms – als Vektor zwischen Antennenzentrum und Projektionszentrum – von der in der Photogrammetrie üblichen Festlegung abweicht!

1	137007.745028	[2074]	26,N	2,E	191,V	51.51629799,Lat	7.33605200,Lon	219.190,Ellh	0.011534,	0.014892,	0.023045	50,Q
2	137011.540180	[2074]	39,N	-11,E	190,V	51.51634436,Lat	7.33592497,Lon	219.164,Ellh	0.011257,	0.014774,	0.022825	50,Q
3	137014.828655	[2074]	37,N	-13,E	190,V	51.51639034,Lat	7.33580400,Lon	219.238,Ellh	0.010862,	0.014808,	0.022815	50,Q
4	137018.129651	[2074]	42,N	-17,E	189,V	51.51643734,Lat	7.33568114,Lon	219.209,Ellh	0.010905,	0.014635,	0.022773	50,Q
5	137021.436674	[2074]	43,N	-13,E	189,V	51.51648456,Lat	7.33555970,Lon	219.129,Ellh	0.011128,	0.014976,	0.022773	50,Q
6	137024.768855	[2074]	44,N	-13,E	189,V	51.51653165,Lat	7.33543664,Lon	219.082,Ellh	0.011046,	0.014868,	0.022922	50,Q
7	137028.065583	[2074]	44,N	-18,E	189,V	51.51657771,Lat	7.33531410,Lon	219.076,Ellh	0.011286,	0.015268,	0.023078	50,Q
8	137031.362426	[2074]	43,N	-19,E	189,V	51.51662397,Lat	7.33519211,Lon	219.114,Ellh	0.011445,	0.015365,	0.023139	50,Q
9	137034.646846	[2074]	42,N	-13,E	190,V	51.51667034,Lat	7.33507188,Lon	219.140,Ellh	0.011212,	0.015130,	0.022813	50,Q
10	137037.943424	[2074]	43,N	-16,E	189,V	51.51671679,Lat	7.33494892,Lon	219.147,Ellh	0.011580,	0.015509,	0.023134	50,Q

Abb. 3: Auszug aus Timestamp.MRK mit den protokollierten RTK-Informationen

Weiterhin werden die Originaldaten der Satellitenbeobachtung sowie die Ephemeriden-Daten erfasst und in einer PPKRAW.bin-Datei im Format RTCM 3.2 gespeichert. Zusätzlich konvertiert das System die Satellitendaten on the fly in das RINEX-Format (Receiver Independent Exchange Format) und schreibt diese Daten in eine RINEX.obs-Datei. Damit stehen ebenfalls alle relevanten Informationen für eine PPK-Auswertung zur Verfügung. Diese kann – bedarfsorientiert – z. B. auf Basis der kostenfreien RTKLIB-Software (BÄUMKER 2014, TAKASU 2020) ausgeführt werden.

Zur weiteren Prozessierung müssen die vorliegenden Positionsdaten in der Regel in ein Zielkoordinatensystem konvertiert werden, in Deutschland vielfach ins Landeskoordinatensystem ETRS89. Da die Höheninformation nach dem Flug als ellipsoidische Koordinaten vorliegt, ist zudem eine Geoidundulation (aktuell im GCG2016 – German Combined Quasigeoidmodell) als Korrekturterm anzubringen. Als Resultat ergeben sich Höhen im Deutschen Haupthöhennetz DHHN2016. „Die horizontalen Variationen“ des Quasigeoids können Beträge bis zu 10 mm pro km annehmen. Quasigeoidvariationen müssen deshalb auch bei lokalen Höhenbestimmungen, z. B. mit GNSS, berücksichtigt werden“ (BKG 2020a). Neben einer Webapplikation (BKG 2020b) vertreibt das BKG hierzu auch eine Software als Desktoplösung.

### 3 Untersuchungsergebnisse

Die nachfolgenden Auswertungen erfolgten für vier verschiedene Systeme Phantom 4 RTK, welche an drei unterschiedlichen Tagen eingesetzt wurden. Die Auswertungen wurden mit der Software Agisoft Metashape durchgeführt. Alle Berechnungen basieren auf identischer Parametrisierung um somit eine Vergleichbarkeit der Resultate zu gewährleisten.

#### 3.1 Qualität der RTK-Messungen

Soll die Lagerung eines Bündelblocks auf der Grundlage gemessener Äußerer Orientierungen ausgeführt werden, ist vorab die Frage zur Qualität der gemessenen Positionsdaten zu klären. Die in Kap. 1 aufgeführten Herstellerangaben zu den erreichbaren Genauigkeiten bewegen sich in einem Bereich der sowohl eine optimale Satellitenkonfiguration als auch einen ungestörten Empfang der Echtzeit-Korrekturdaten voraussetzt. Von diesen Rahmenbedingungen kann nicht

durchgängig ausgegangen werden. Der Hersteller nennt hierzu vier verschiedene Qualitätsstufen zum RTK-Status, der auch während des Bildflugs in der Steuerungs-App angezeigt werden (Abb. 2 rechts):

- None
- RTK-FIX (Mehrdeutigkeiten / Ambiguities gelöst)
- RTK-FLOAT (keine Lösung der Mehrdeutigkeiten)
- SINGLE-GNSS

Abb. 4 zeigt die während der Bildflüge (die sich aus jeweils zwei Teilflügen einer Kreuzbefliegung zusammensetzen) erreichten Standardabweichungen der RTK-Messungen. Die vorliegenden Ergebnisse machen die zum Teil starken Variationen in der Qualität der gemessenen RTK-Positionen deutlich. Lediglich die Messungen des Systems A zeigen eine homogene Datenqualität, die einer FIX-Lösung entspricht. Eine geringe Anzahl von sichtbaren Satelliten, eine schlechte Geometrie der Satellitenkonstellation sowie eine schlechte Funkverbindung zwischen Basisstation und Rover können eine FIX-Lösung verhindern.

Der Haupteinflussfaktor ist bei den hier vorliegenden Daten in der Qualität der Datenverbindung zum SAPOS-Dienst HEPS zu sehen. Derartige Probleme sind nicht untypisch und durch die in Teilen schlechte Qualität des Mobilfunknetzes bedingt. Allerdings können zusätzlich auch atmosphärische Einflüsse eine Ursache darstellen. So fand z. B. die Kreuzbefliegung mit System D unter ungünstigen Witterungsbedingungen (Wind, aufkommende Regenfront) statt. Für die Flüge mit den Systemen B und C ist die RTK-Qualität für jeweils einen Teilflug akzeptabel, während für den zweiten Teilflug deutliche Genauigkeitsschwankungen vorliegen. Die Flüge mit System A weisen eine durchgängig hohe RTK Qualität auf.

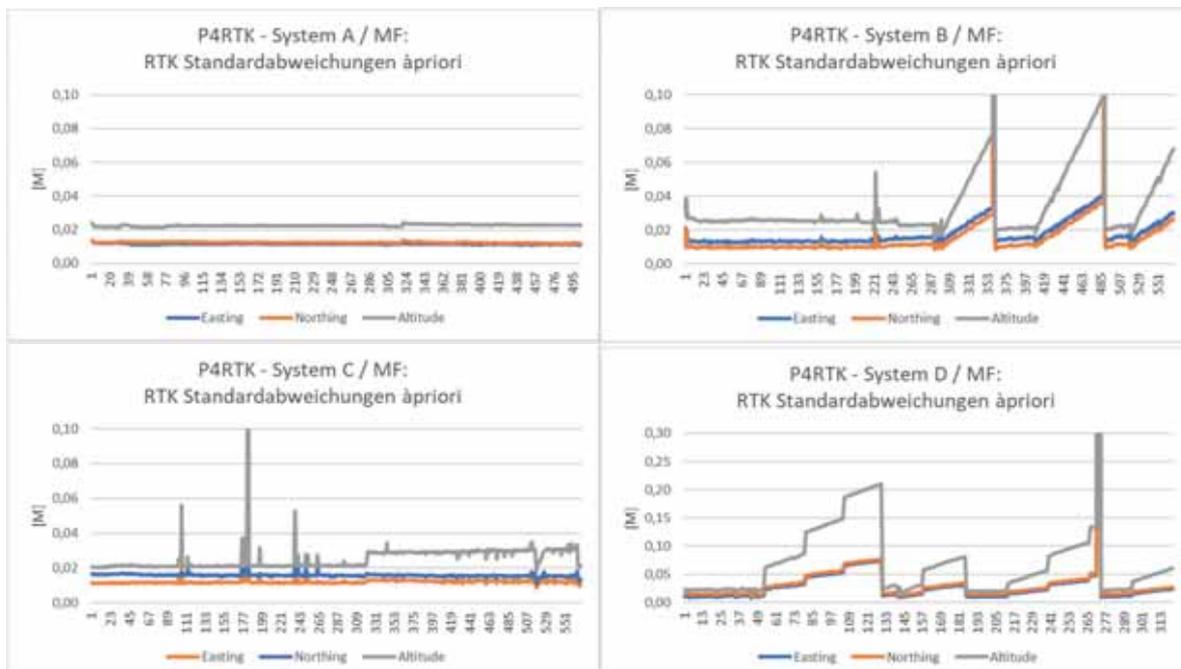


Abb. 4: Standardabweichungen (à priori) der mittels RTK bestimmten Bildpositionen (gemessene ÄORI) der Phantom 4 RTK Systeme A – D. Hinweis: die vertikale Skalierung für System D weicht von denen der Systeme A – C ab

### 3.2 Georeferenzierung der Bildverbände mittels RTK

Die in Kap. 3.1 gezeigten RTK-Messungen können als Basis für die Bildorientierung im Rahmen der Bündelblockausgleichung (BBA) genutzt werden, mit dem Ziel die Anzahl notwendiger Passpunkte zu reduzieren oder ggf. auch völlig darauf zu verzichten (PRZYBILLA et al. 2015, GERKE & PRZYBILLA 2016).

Die nachfolgenden Untersuchungen orientieren sich an typischen Lagerungskonfigurationen, die sich aus einer Kombination von Passpunkten und gemessener Äußerer Orientierung zusammensetzen (integrierte Orientierung). Im Gegensatz zu Bildverbänden der mantragenden Luftbildphotogrammetrie sind hierbei hinreichend genaue Messungen der Orientierungswinkel, die ein hochgenaues Inertiales Messsystem liefert, nicht verfügbar. Die in Tab. 1 gelisteten Konfigurationen wurden evaluiert. Abhängig vom individuellen Block variiert die maximale Anzahl der GCP zwischen 45 und 50. Für alle Konfigurationen wurde eine einheitliche Innere Orientierung (UNIFIED) für die zwei Teilflüge der Kreuzbefliegung eingeführt. Eine weitere Berechnung erfolgte mit zwei separaten Inneren Orientierungen (SEPARATE) für die Blocklagerung mit beobachteter ÄORI und 4 GCP in den Blockecken. Ziel dieser Variante ist einen eventuellen Einfluss veränderlicher Fokussierung der Kamera zu erkennen.

Tab. 1: Lagerungskonfigurationen aus gemessener ÄORI und Passpunkten

Innere Orientierung (IORI)	ÄORI	Vollständige GCP (45-50)	4 GCP in den Blockecken	ÄORI & 4 GCP	ÄORI & 1 GCP
UNIFIED	X	X	X	X	X
SEPARATE	-	-	-	X	-

Die während der Bildflüge mit den DJI-Systemen erfassten RTK-Daten wurden mit ihren à priori Genauigkeiten als Beobachtungen in die BBA eingeführt. In der Folge der Berechnungen wurde zudem der Einfluss unterschiedlicher Passpunktanordnungen evaluiert. Abb. 5 zeigt die Verbesserungen an den beobachteten Äußerer Orientierungen nach der Bündelblockausgleichung.

In Agisoft Metashape werden diese Größen als „ERROR“ bezeichnet. Die in der Abbildung dargestellten Resultate basieren auf einer Blocklagerung bestehend aus den RTK-Messungen sowie zusätzlichen 4 Passpunkten in den Blockecken. Die Verbesserungen liegen bei Bildflügen mit einem hohen Anteil von FIX-Lösungen in der Größenordnung der Beobachtungsgenauigkeiten (System A und C). Die Resultate für die Systeme B und D zeigen die (in Teilen) systematisch schlechten RTK-Messungen durch entsprechend hohe Verbesserungen an.

Die Überprüfung der Blockgeometrie, hier insbesondere auch von Blockdeformationen, erfolgt über die Restabweichungen (RMSE) an den Kontrollpunkten (Checkpoints – CP, Abb. 7). Im Gegensatz dazu sind die entsprechenden RMSE-Werte an den Passpunkten (Ground Control Points – GCP, Abb. 6) weniger aussagekräftig. Sie spiegeln lediglich wider, wie die Einpassung des Bildverbands auf die GCP erfolgt.

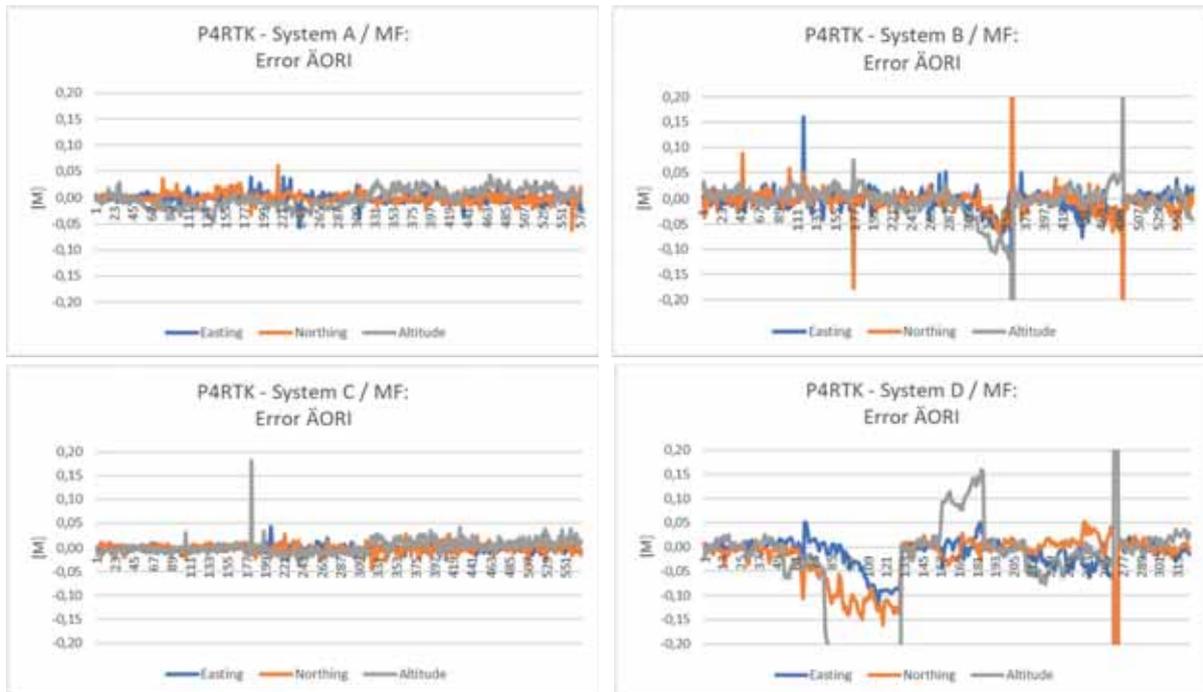


Abb. 5: Verbesserungen (Error) an den durch RTK bestimmten Bildpositionen (ÄORI) nach der BBA für die Systeme A – D



Abb. 6: RMSE-Werte an den Passpunkten (CP) in Abhängigkeit von der Lagerungsart (direkt: ÄORI – indirekt: GCP – integriert: ÄORI+GCP) (Systeme A – D)



Abb. 7: RMSE-Werte an den Kontrollpunkten (CP) in Abhängigkeit von der Lagerungsart (direkt: ÄORI – indirekt: GCP – integriert: ÄORI+GCP) (Systeme A – D). Hinweis: die vertikale Skalierung weicht von der der GCP (Abb. 6) um den Faktor 10 ab

In Abb. 7 werden die Einflüsse der Lagerungsart auf den jeweiligen Block sehr deutlich. Es können folgende Effekte abgeleitet werden:

- Direkte Lagerung auf die gemessene ÄORI  
Die Abweichungen an den CP liegen, bezogen auf die Lagekoordinaten, in der Größe der RTK-Genauigkeit (10-20 mm), jedoch zeigt sich eine signifikant große Abweichung in der Höhe. Diese liegt bei den Systemen A und D deutlich über 100 mm und bei System C knapp unter diesem Wert. Lediglich für System B liegt die Abweichung im Rahmen der Beobachtungsgenauigkeit. Eine zuverlässige Lagerung für diese Variante (ohne Bodenpasspunkte) ist nicht erkennbar.
- Indirekte Lagerung mit maximaler Passpunkt-Besetzung  
Da in dieser Variante alle GCP zur Lagerung genutzt werden ist eine Kontrolle mittels unabhängiger CP nicht möglich. In Abb. 6 ist jedoch die hohe Qualität der Einpassung auf die GCP erkennbar, die bei ca. 0,5 - 0,7 der GSD liegt. Die hier betrachtete Lagerungsvariante ist mit einem sehr hohen terrestrischen Aufwand verbunden.
- Indirekte Lagerung mit minimaler Passpunkt-Besetzung  
Aus der Lagerung über 4 GCP in den Blockecken ist deutlich erkennbar, dass keine hinreichende Stabilität in den Bildverbänden erreicht werden kann. Während die Lageabweichungen der CP noch im Bereich der GSD liegen, übersteigen die Höhenabweichungen diese um den Faktor 15 - 30. Eine Ursache für dieses Resultat ist in der Metrik der Kamera bzw. den offensichtlich unzureichenden Möglichkeiten zur simultanen Kalibrierung zu sehen (Kap. 3.3).
- Integrierte Lagerung auf die gemessene ÄORI und vier Passpunkte

Die vorliegenden Resultate zeigen die Wirksamkeit der integrierten Lagerung auf der Basis der RTK-Messungen in Verbindung mit einer Passpunktbesetzung in den Blockecken. Die Abweichungen an den CP liegen im Bereich der GSD, in Teilen auch darunter. Die erzielten Ergebnisse sind nur minimal schlechter als die Variante mit einer Vollpasspunkt-Besetzung.

- Integrierte Lagerung auf die gemessene ÄORI und einem Passpunkt  
Während die direkte Georeferenzierung mittels gemessener Äußerer Orientierung in den vorliegenden Datensätzen durch signifikante Höhenabweichungen gekennzeichnet ist, sind die positiven Auswirkungen eines zusätzlichen Passpunktes (in der Blockmitte) deutlich erkennbar. Die systematischen Höhenabweichungen an den CP reduzieren sich auf ca. 15 - 30 mm und liegen damit auf dem Genauigkeitsniveau der RTK-Messungen. Die Qualität der hier erzielten Genauigkeiten bei der Bildorientierung ist für topographische Anwendungen ausreichend.

### 3.3 Georeferenzierung der Bildverbände mittels Post-Processed Kinematic (PPK)

DJI RTK-Systeme bieten, auf Grund der Verfügbarkeit der Originaldaten der Satellitenbeobachtung sowie der Ephemeriden-Daten, die Möglichkeit einer verbesserten Positionsbestimmung im Post-processing (PPK). Die notwendigen Berechnungen können z. B. mit der freien Software RTKLIB (TAKASU 2020) ausgeführt werden. Da DJI über keine eigene Auswertesoftware verfügt, wird RTKLIB in den Workflows verschiedener Drittanbieter eingesetzt (AEROTAS 2020, KLAUPPK 2020). Allgemeine Hinweise auf GNSS-Workflows finden sich z. B. bei EMLID (2020). Ein besonderes Augenmerk ist bei der Auswertung auf die korrekte Anpassung/Interpolation der Positionsdaten auf den jeweiligen Zeitstempel der Bildaufnahme sowie die Hebelarm-Korrektur zu legen. Die hierzu notwendige Software wurde von BÄUMKER (2020) erstellt und für das Post-Prozessierung eingesetzt.

In Abb. 8 sind Vergleichsdaten (System B) aus RTK und PPK Prozessierungen dargestellt. Es wird deutlich, dass die beim RTK wiederholt auftretenden Beobachtungen mit großen Standardabweichungen à priori bei der PPK Prozessierung im Wesentlichen nicht mehr vorhanden sind. Es bleibt aber erkennbar, dass im zweiten Teilflug (Bildnummer ab ca. 245, Flughöhe über Grund: 50 m) – im Vergleich zum ersten (Flughöhe über Grund: 60 m) – eine offensichtlich schlechtere Satellitenkonfiguration vorlag, die zu einer schlechteren Beobachtungsgenauigkeit führt. Die Differenzen zwischen RTK- und PPK-Auswertung zeigen in weiten Bereichen eine gute Übereinstimmung, allerdings liegen auch hier partielle Probleme bei der Positionsbestimmung vor.

Die Verbesserungen (ERROR) an den Bildpositionen nach der BBA sind in weiten Bereichen sehr ähnlich, was für die insgesamt gute RTK-Lösung spricht. Dieses Resultat wird auch durch die Restabweichungen an den Kontrollpunkten bestätigt. Für alle Lagerungsarten ergeben sich nahezu identische Ergebnisse. Lediglich die direkte Blocklagerung auf die gemessene ÄORI weist Unterschiede auf, wobei bei der PPK-Lösung erstaunlicherweise höheren RMSE-Werte vorliegen.



Abb. 8: Vergleich der Auswertungen auf der Basis von RTK-Messungen (links) und PPK-Auswertungen (rechts) für das System B:  
 Oben: Standardabweichungen (à priori) der ermittelten Bildpositionen (ÄORI)  
 Mitte-oben: Differenzen zwischen RTK und PPK (links: Flug 1, rechts: Flug 2)  
 Mitte- unten: Verbesserungen (Error) an den Bildpositionen (ÄORI) nach der BBA  
 Unten: RMSE-Werte an den Kontrollpunkten (CP) in Abhängigkeit von der Lagerungsart

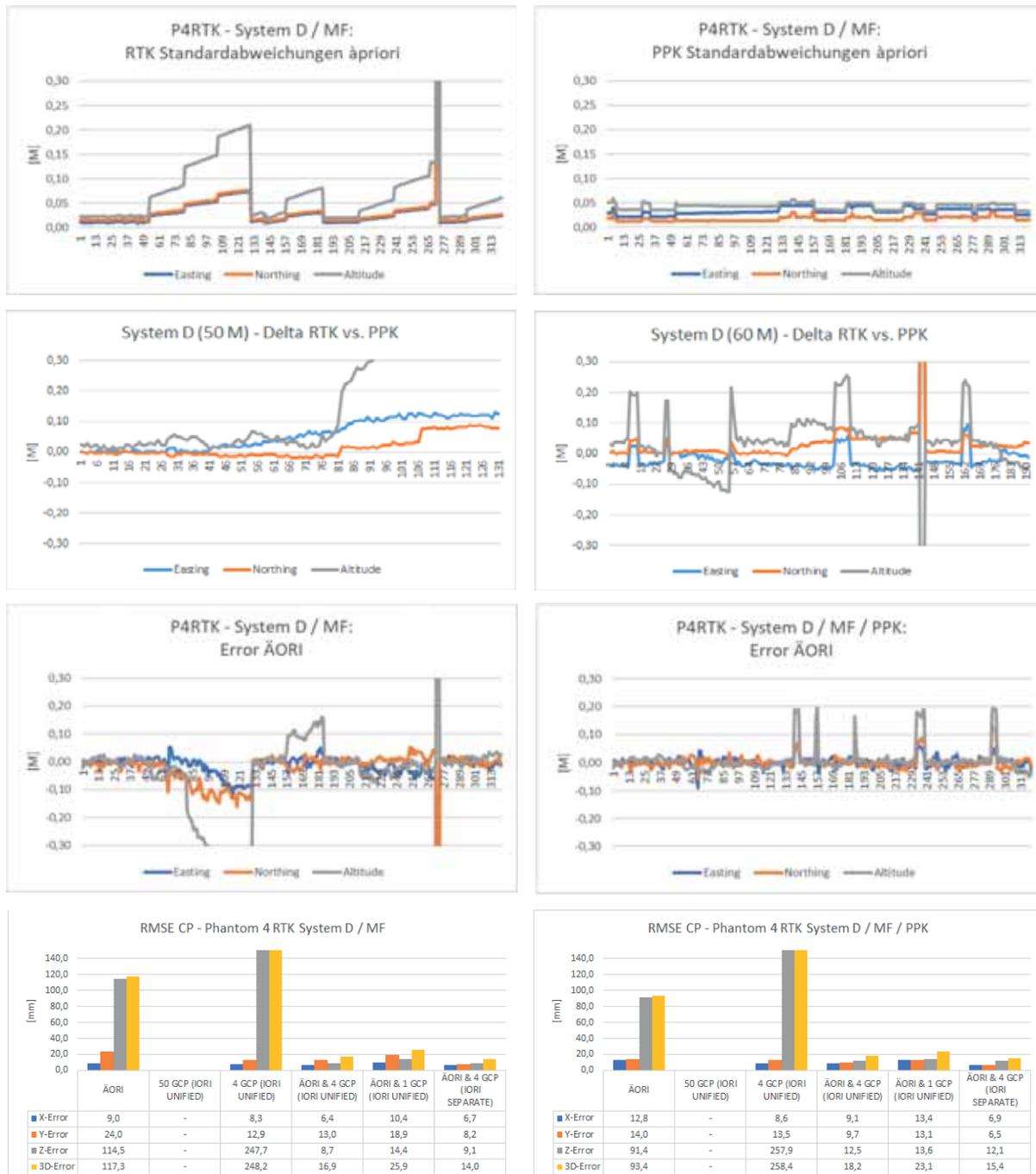


Abb. 9: Vergleich der Auswertungen auf der Basis von RTK-Messungen (links) und PPK-Auswertungen (rechts) für das System D:  
 Oben: Standardabweichungen (à priori) ermittelten Bildpositionen (ÄORI)  
 Mitte-oben: Differenzen zwischen RTK und PPK (links: Flug 1, rechts: Flug 2)  
 Mitte-unten: Verbesserungen (Error) an den Bildpositionen (ÄORI) nach der BBA  
 Unten: RMSE-Werte an den Kontrollpunkten (CP) in Abhängigkeit von der Lagerungsart

Die Vergleichsdaten RTK vs. PPK für das System D zeigt Abb. 9. Die vorliegenden RTK-Daten weisen für ca. 20% der insgesamt 321 Bilder des Fluges à priori Genauigkeiten  $> 10$  cm auf, so dass davon ausgegangen werden kann, dass hier keine FIX-Lösung vorliegt (Abb. 10). Der Flug weist im Vergleich zu denen der Systeme A – C die schlechteste RTK-Qualität auf. Die PPK-Lösung liefert gegenüber der des Systems B geringere Genauigkeiten. Diese sind zwar bei getrennter Betrachtung der zwei Teilflüge nahezu homogen in Lage und Höhe, jedoch im zweiten Teilflug (60 m) erkennbar schlechter. Häufigere, größere Differenzen zwischen RTK- und PPK-Auswertung sind die Folge (Abb. 9, Mitte-oben).

Der generelle Qualitätsgewinn der PPK-Lösung spiegelt sich nach Durchführung der BBA im Ergebnis nicht wider. Trotz homogener Qualität der Positionsbestimmungen liegen für verschiedene Bereiche des Bildverbandes Zuschläge (Error) an den gemessenen ÄORI von bis zu 20 cm vor. Die Ursachen sind möglicherweise in einer schlechten Satellitenkonfiguration in Verbindung mit sehr ungünstigen Witterungsbedingungen zum Zeitpunkt des Bildflugs zu suchen. Abb. 11 weist hier deutliche Systematiken für die an der ÄORI im Rahmen der BBA angebrachten Zuschläge auf.

Der abschließende Vergleich der RMSE-Werte an den Kontrollpunkten (Abb. 9 unten) liefert nahezu identische Resultate. Der vermutete Genauigkeitsgewinn durch die Post-Prozessierung ist im vorliegenden Datensatz nicht nachweisbar.

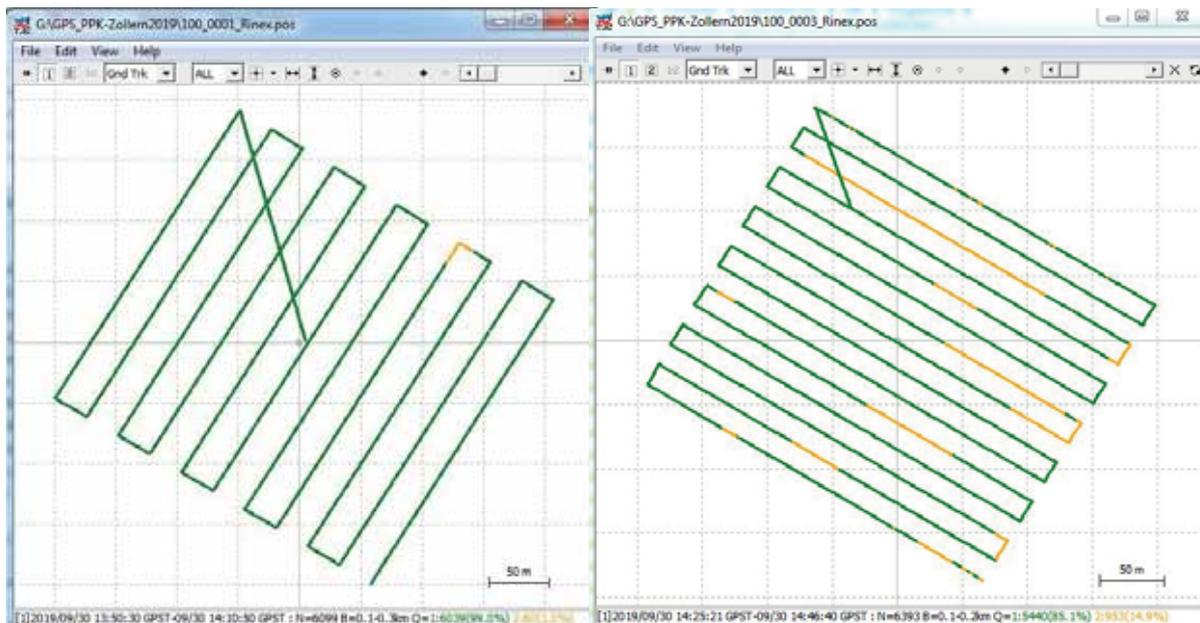
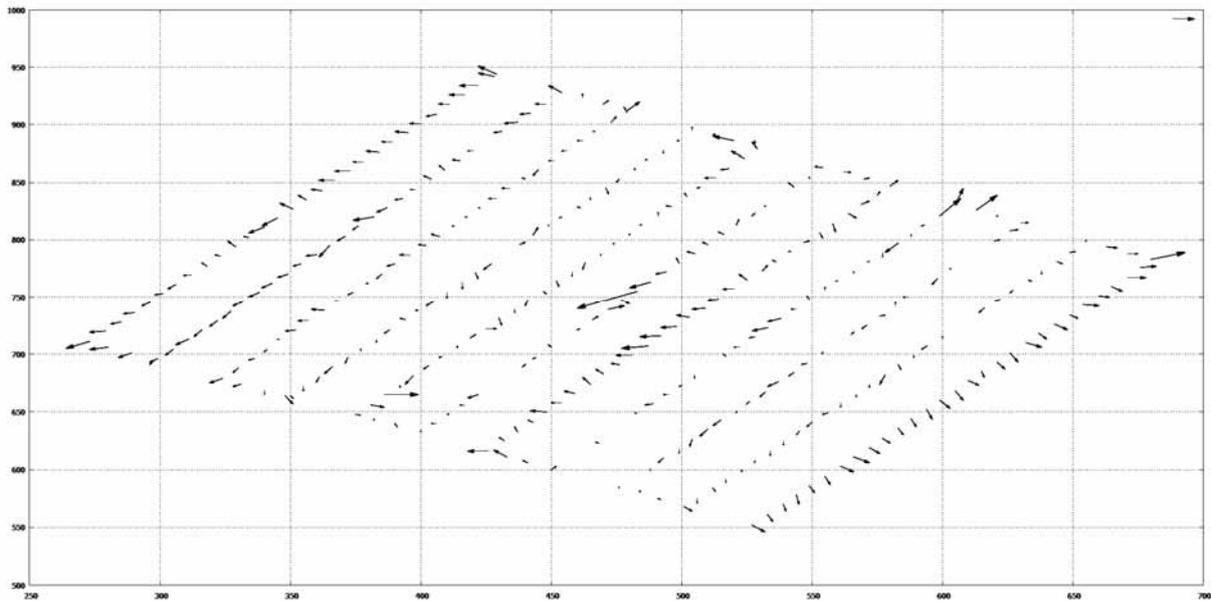
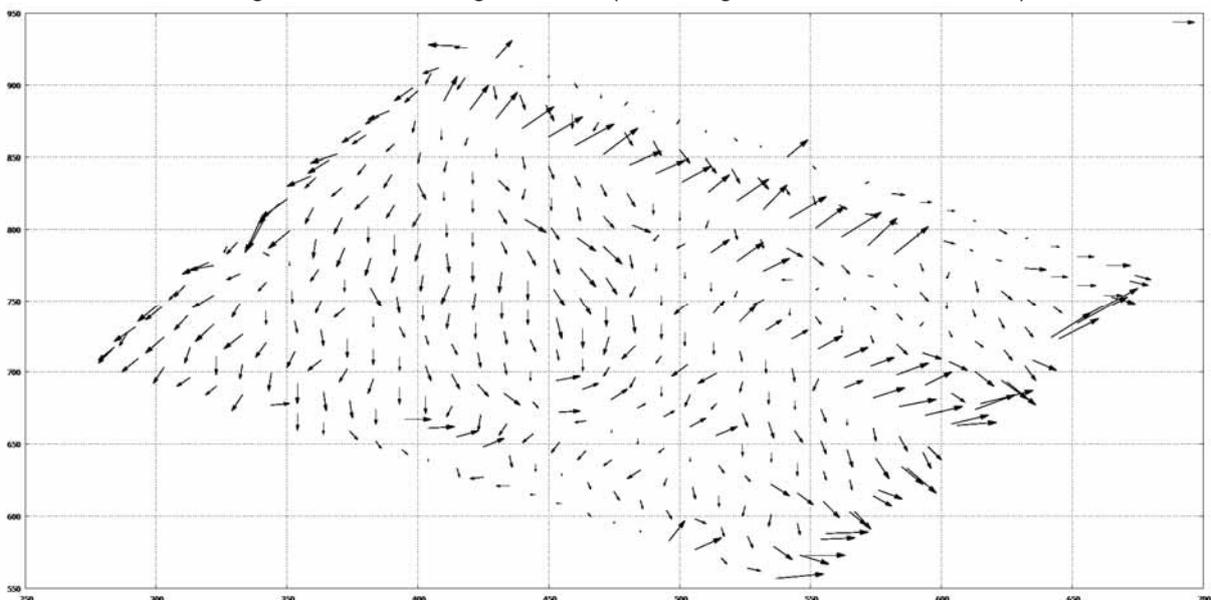


Abb. 10: Anteil der FIX (Grün) und FLOAT-Lösungen (Gelb) aus dem Post-Processing mit RTKLIB für Flüge mit System D (Flughöhe ü. G., links 50 m, rechts 60 m)



RMSE-Werte: Easting: 19,3 mm, Northing: 20,1 mm (nicht dargestellt: Altitude: 26,4 mm)



RMSE-Werte: Easting: 29,9 mm, Northing: 45,2 mm (nicht dargestellt: Altitude: 74,5 mm)

Abb. 11: RMSE-Werte der ÄORI nach BBA, mit deutlich erkennbaren Systematiken in den PPK-Messungen - System D – Flughöhe ü. G.: oben 50 m, unten 60 m – Lagerung auf 50 GCP

### 3.4 Innere Orientierung der Kamera

Voraussetzung für die in-situ Kalibrierung der Kamera in der BBA ist neben einer geeigneten Aufnahmekonfiguration (hier: Kreuzverband) das Vorliegen entsprechender Lagerungsinformationen (Passpunkte, gemessene Elemente der ÄORI). Abb. 12 zeigt Änderungen der Parameter Kamerakonstante ( $\Delta c$ ) und Hauptpunktlage ( $XH, YH$ ) in Abhängigkeit von der Blocklagerung.

Für alle Kameras liegen geringe Änderungen in der Hauptpunktlage vor ( $\leq 1$  Pixel), ein Resultat, das für die hohe Stabilität der jeweiligen Systeme spricht.



Abb. 12: Änderungen der Parameter der Inneren Orientierung ( $\Delta c$ , XH, YH) in Abhängigkeit von der Blocklagerung sowie gemeinsamer bzw. separater Parameter für die Teilflüge (Systeme A – D). Hinweis: die Abweichungen  $\Delta c$  beziehen sich auf einen einheitlichen Startwert

Der Einfluss der Blocklagerung auf den Parameter „Kamerakonstante“ wird bei allen Systemen sehr deutlich. Bei der Nutzung von ausschließlich vier Passpunkten in den Blockecken ist eine zuverlässige Bestimmung dieses Parameters nicht möglich. Die Abweichungen gegenüber allen anderen Varianten betragen zwischen 20 - 80 Pixel. Die entsprechenden (negativen) Folgen zeigt Abb. 7, hier insbesondere bei den dargestellten Höhenabweichungen an den Kontrollpunkten. Die verwendeten Ansätze einer gemeinsamen (UNIFIED) bzw. separaten (SEPARATE) Parametrisierung der Inneren Orientierung für die jeweiligen Teilflüge zeigt nahezu identische Resultate, sowohl für die RMSE-Werte an den Kontrollpunkten (Abb. 7) als auch für die Variation der Kamerakonstante (Abb. 8). Hieraus ist ableitbar, dass von einer Stabilität der untersuchten Parameter im Verlauf der zwei Teilflüge des Kreuzverbands ausgegangen werden kann.

## 4 Fazit & Ausblick

Mit der Verfügbarkeit von präzisen RTK-Lösungen für den UAV-Bildflug und der damit verbundenen direkten Georeferenzierung von photogrammetrischen Bildverbänden, ist ein wichtiger Schritt mit Blick auf den wirtschaftlichen Einsatz von UAV's in Geodäsie und Photogrammetrie getan. Die im Beitrag untersuchten UAV-Systeme Phantom 4 RTK liefern für Anwendungen im mittleren Genauigkeitsbereich ( $> 2 - 3$  cm) hinreichende Möglichkeiten den terrestrischen Aufwand für die Georeferenzierung deutlich zu reduzieren, indem auf umfangreiche Passpunktmes-

sungen weitestgehend verzichtet werden kann. Die vorliegenden Ergebnisse zeigen jedoch auch die zum Teil starken Variationen in der Qualität der gemessenen RTK-Positionen sowie deren Auswirkungen auf die Bildorientierung und die beteiligten Parameter. Eine Lagerung ausschließlich auf die gemessenen Elemente der Äußeren Orientierung scheint nach den vorliegenden Ergebnissen nicht sinnvoll, da hier vielfach systematische Höhenversätze erkennbar sind. Der integrierte Ansatz zur Georeferenzierung ist hier das Mittel der Wahl. Tab. 2 enthält einen Vorschlag für den Einsatz von RTK / PPK, in Abhängigkeit vom Anwendungsbereich.

Tab. 2: Nutzung von RTK und PPK in Abhängigkeit vom Anwendungsbereich

Parameter	Topographie	Kataster	Ingenieurvermessung
GSD	2,5 cm	1,5 cm	< 1cm
Bildmassstab	1:10.000	1:5.000	1:1.000
Genauigkeitsbereich	10 - 20 cm	1 - 3 cm	0,2 - 1 cm
Genauigkeit erreichbar	+++	++	+
Georeferenzierung	RTK / PPK	RTK / PPK & GCP*	GCP**
Punktwolke	x	x	x
Volumina	x	-	-
Höhenlinien	x	-	-
Profile	x	-	x
Orthophoto	x	x	-
Karte	x	x	-
3D Punkte	-	x	x

GCP\* - GNSS-Messung / GCP\*\* - Netzmessung

Die mit zusätzlichem Aufwand verbundene Post-Prozessierung der GNSS-Daten lieferte für die hier untersuchten zwei Datensätze der P4RTK keinen zusätzlichen Genauigkeitsgewinn, obwohl für ca. 20% der RTK-Daten des Systems D keine FIX-Lösungen erreicht werden konnten. Durch eine angepasste Gewichtung (auf der Basis der à priori Genauigkeiten) in der BBA ist der Einfluss dieser Beobachtungen gering, negative Auswirkungen auf die Blockgeometrie nicht erkennbar. Der PPK-Lösung kommt jedoch überall dort eine signifikante Bedeutung zu, wo eine schlechte Mobilfunk-Infrastruktur die Nutzung von Korrekturdiensten verhindert. Probleme eingeschränkter Satellitenempfangs können aber auch durch das PPK nicht aufgefangen werden. Positiv zu bewerten ist die Qualität der Kamera. Das in der Phantom 4 RTK genutzte System ist mit seinem Konzept baugleich mit der Zennuse X4S. Die Kamera verfügt über eine hohe Stabilität, die deutlich über der sonstiger Systeme des Herstellers liegt. Die Ursache hierfür ist in der Fix-Fokus Optik zu sehen, die mechanische Instabilitäten als Folge von Wechseloptiken ausschließt. Aus Sicht der Photogrammetrie ist sie trotz dieser positiven Eigenschaften keine Messkamera, eine in-situ Kalibrierung daher dringend notwendig.

## 5 Literaturverzeichnis

- AEROTAS, 2020: Phantom 4 RTK – PPK Processing Workflow. <https://www.aerotas.com/phantom-4-rtk-ppk-processing-workflow>. Letzter Zugriff: 19.01.2020.
- BÄUMKER, M., 2014: Zeitreihenanalyse der Daten der GNSS Referenzstation der Hochschule Bochum. In: Zeitabhängige Messgrößen – Ihre Daten haben (Mehr-)Wert. Beiträge zum 129. DVW-Seminar am 26. und 27. Februar 2014 in Hannover, Schriftenreihe des DVW, Band **74** (Hrsg. DVW e.V. – Gesellschaft für Geodäsie, Geoinformation und Landmanagement).
- BÄUMKER, M., 2020: Labor für Physikalische Messtechnik der HS Bochum. <https://www.hochschule-bochum.de/fbg/einrichtungen-im-fachbereich/labor-fuer-physikalische-messtechnik/>. Letzter Zugriff: 19.01.2020.
- BKG, 2020a: Die Höhenbezugsfläche von Deutschland. <https://www.bkg.bund.de/DE/Ueberdas-BKG/Geodaesie/Integrierter-Raumbezug/Hoehenbezugsflaeche/hoehenbezug.html>. Letzter Zugriff: 15.01.2020.
- BKG, 2020b: Onlineberechnung von Quasigeoidhöhen mit dem GCG2016, <http://gibs.bkg.bund.de/geoid/gscmp.php?p=g>. Letzter Zugriff: 15.01.2020.
- DJI, 2020: Phantom 4 RTK Technische Daten, <https://www.dji.com/de/phantom-4-rtk/info>. Letzter Zugriff: 15.01.2020.
- EMLID, 2020: GPS post-processing. <https://docs.emlid.com/reachm-plus/common/tutorials/gps-post-processing/>. Letzter Zugriff: 19.01.2020.
- GERKE, M. & PRZYBILLA, H.-J., 2016: Accuracy analysis of photogrammetric UAV image blocks: influence of onboard RTK-GNSS and cross flight patterns. PFG - Photogrammetrie, Fernerkundung, Geoinformation **2016**(1), 17-30.
- HANDELSBLATT, 2020: Diese chinesische Firma hat mit Drohnen den Massenmarkt erobert. <https://www.handelsblatt.com/technik/digitale-revolution/digitale-revolution-diese-chinesische-firma-hat-mit-drohnen-den-massenmarkt-erobert/25395412.html?ticket=ST-1134012-TrivDX4ReHLfHg6nRijj-ap2>. Letzter Zugriff: 19.01.2020.
- KLAUPPK, 2020: <https://klauppk.com/klauppk-software-for-the-phantom4-rtk-2/>. Letzter Zugriff: 19.01.2020.
- PRZYBILLA, H.-J., REUBER, C., BÄUMKER, M. & GERKE, M., 2015: Untersuchungen zur Genauigkeitssteigerung von UAV-Bildflügen. Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V., Band **24**, T. Kersten (Hrsg.), 35. Wissenschaftlich-Technische Jahrestagung der DGPF, 16.-18. März 2015 an der Universität zu Köln, Tagungsband auf CD-ROM, 45-54.
- PRZYBILLA, H.-J., BÄUMKER, M. & VIETEN, J., 2018: Das UAV-Testfeld Zeche Zollern in Dortmund. Schriftenreihe des DVW, Band **89**, 61-80, Wißner-Verlag, Augsburg.
- SAPOS, 2020: SAPOS-Dienste im Überblick. <https://www.sapos.de/dienste-im-ueberblick.html>. Letzter Zugriff: 15.01.2020.
- TAKASU, T., (2020): RTKLIB, Open Source Program Package for RTK-GPS. <http://gpspp.sakura.ne.jp/rtklib/rtklib.htm>. Letzter Zugriff: 15.01.2020.

# Erprobung des Leica SPL100 Laserscanners zur Erfassung der Wasseroberfläche an deutschen Binnengewässern

ROBERT WEIß<sup>1</sup> & GOTTFRIED MANDLBURGER<sup>2</sup>

*Zusammenfassung: Ein Baustein zur Sicherstellung der Sicherheit und Leichtigkeit des Verkehrs auf Bundeswasserstraßen ist die korrekte Erfassung der Wasseroberfläche. In der heutigen Zeit werden die entsprechenden Anforderungen immer höher und linienhafte Erfassungen in Form von Längsprofilen reichen für bestimmte Anwendungen nicht mehr aus. Infolgedessen besteht ein großes Interesse an einer Methode zur flächenhaften Erfassung von Wasseroberflächen. Die BfG wurde beauftragt zu untersuchen, inwieweit der Leica SPL100 Single Photon LiDAR Scanner für diese Aufgabe geeignet ist. Gleichzeitig soll untersucht werden, welche Performance der Scanner für topografische Aufnahmen hat. Im Rahmen der Untersuchung erfolgte eine Befliegung bestimmter Testgebiete entlang des Mittelrheins mit unterschiedlichen Eigenschaften in unterschiedlichen Flughöhen. In diesem Betrag werden erste Ergebnisse vorgestellt.*

## 1 Einleitung und Motivation

In Deutschland ist die Wasserstraßen- und Schifffahrtsverwaltung (WSV) für die Sicherstellung der Sicherheit und Leichtigkeit der Schifffahrt auf Bundeswasserstraßen zuständig. Dies beinhaltet auch die Sicherstellung der Hindernisfreiheit definierter Fahrrinnen und Mindesttiefen. Hierfür sind neben vielen weiteren Informationen auch die Erfassung der Geometrie der Wasseroberfläche nötig. Konventionell wird diese punktuell über Pegel bzw. linienhaft über Wasserspiegelfixierungen erfasst. In der heutigen Zeit reichen eindimensionale Informationen zur Wasseroberfläche nicht mehr für alle Anwendungen aus. Insbesondere für hydronumerische Modellierungen sind flächenhafte Informationen zur Wasseroberfläche von großem Interesse. Die Bundesanstalt für Gewässerkunde (BfG) ist eine dem Deutschen Bundesministerium für Verkehr und digitale Infrastruktur (BMVI) nachgeordnete wissenschaftliche Oberbehörde, deren Aufgabe auch darin besteht, die Wasserstraßen- und Schifffahrtsverwaltung fachtechnisch zu beraten. Im Rahmen dieser Aufgaben werden von der BfG unter anderem auch neuartige technische Systeme erprobt, sowie wissenschaftlich/technische Fragestellungen bearbeitet. Von Seiten des BMVI wurde die BfG beauftragt, die Möglichkeiten zur Erfassung der Wasseroberfläche mithilfe luftgestützter Airborne Laserscanning-Technik zu evaluieren. Bedingt durch die variierende Wasserspiegelhöhe treten dabei gegenüber einer konventionellen ALS-Befliegung zusätzliche Anforderungen auf. Durch seine technischen Möglichkeiten ist der Leica SPL100 eine vielversprechende Lösung (DEGNAN et al. 2016; DEGNAN et al. 2018; LEICA 2020). Erste Untersuchungen zur Erfassung der Wasseroberfläche mit dem Leica SPL100 wurden von MANDLBURGER & JUTZI (2018) und MANDLBURGER & JUTZI (2019) veröffentlicht.

---

<sup>1</sup> Bundesanstalt für Gewässerkunde, Am Mainzer Tor 1, D-56068 Koblenz, E-Mail: weiss@bafg.de

<sup>2</sup> Technische Universität Wien, Wiedner Hauptstraße 8-10, A-1040 Wien, Österreich, E-Mail: gottfried.mandlbuerger@geo.tuwien.ac.at

Im Rahmen einer Beauftragung durch die BfG erfolgte eine ALS-Befliegung bestimmter Testgebiete an Rhein und Mosel in unterschiedlichen Flughöhen (Abb. 1, links). Von Seiten des Auftragnehmers erfolgte die Datenlieferung der ersten Bearbeitungsstufe Mitte Januar 2020.

## 2 Problemstellung

Eine Unterhaltung von Wasserstraßen bedingt immer neben einer Erfassung der Gewässersohle auch die Definition bestimmter Bezugswasserstände, welche die Grundlage für nautische Karten darstellen. Praktisch bedeutet dies, dass an definierten Positionen (z.B. alle 100 m) den Bezugswasserständen Höhenwerte im amtlichen Höhenbezugssystem zugeordnet werden. Am Rhein ist ein Gleichwertiger Wasserstand (GIW) eingeführt, der seit vielen Jahrzehnten als maßgeblicher Wasserstand für den Ausbau und Unterhaltung der Fahrrinne am frei fließenden Rhein verwendet wird. Dieser wird in regelmäßigen Abständen von 10 Jahren überprüft und ggf. aktualisiert (WSV, 2014). Der Verlauf des Wasserspiegelgefälles am Rhein ist nicht linear und insbesondere von Strömungen und der Gewässerbodentopografie abhängig (Abb. 1, rechts).

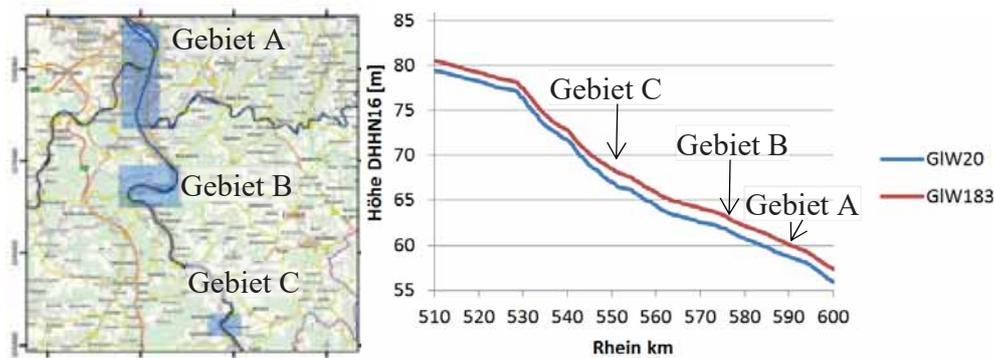


Abb. 1: Befliegungsgebiete (links) und Höhenverlauf des Bezugswasserstandes am Rhein (rechts)

Bedingt durch variierende Abflusszustände ist die Wasseroberfläche als dynamisch anzusehen und folglich kommt die zusätzliche Forderung nach einer sehr kurzen Erfassungszeit hinzu. Der Leica SPL100 ist für diese Anwendung aufgrund der potenziell großen Flughöhe eine naheliegende Lösung. Bedingt durch die Single Photon Lidar Technik ist dieser Scanner sehr sensitiv und in der Lage, auch schwache Wasseroberflächenechos zu empfangen. Weiterhin weist er mit 100 Einzelstrahlen i.V. mit einer sehr geringen Beamdivergenz eine extrem hohe Messrate bzw. hohe Punktdichte auf und kann auch aus einer sehr großen Höhe eingesetzt werden. Diese ist wiederum für eine große Streifenbreite am Boden und damit verbunden für eine Erfassung der Wasseroberfläche innerhalb einer kurzen Zeitspanne nötig.

MANDLBURGER & JUTZI (2019) haben allerdings anhand von Modellrechnungen auf Basis der Laser-Radar Gleichung gezeigt, dass für Standardflughöhen (ca. 4000 m) und -scanwinkel ( $15^\circ$  gegenüber der Nadirrichtung), selbst bei diesem Einzelphoton-sensitiven Sensor oftmals nicht genügend Strahlung von der Wasseroberfläche reflektiert wird, um gesichert Messpunkte bestimmen zu können. Im Zuge eines Pilotprojekts wurde daher eine Datenerfassung aus unterschiedlichen Flughöhen und mit einem Scanwinkel von  $10^\circ$  bezogen auf die Nadirrichtung

durchgeführt. Primäres Ziel war es dabei, eine flächenhafte Erfassung und geometrische Modellierung der Wasseroberfläche der Bundeswasserstraßen in den Befliegungsgebieten zu realisieren und dabei festzustellen, ab welchen Flughöhen dies zuverlässig möglich ist. Die Genauigkeit sollte dabei so hoch sein, dass auch geringere Variationen der Wasseroberflächenhöhen, wie beispielsweise Quergefälle in engen Kurven nachweisbar sind. Sekundäres Ziel sind belastbare Aussagen zur Performance des Leica SPL100 für topografische Aufnahmen trockener Oberflächen. Dies beinhaltet insbesondere Aussagen zur erreichbaren Punktdichte, Genauigkeit und Zuverlässigkeit bei bestimmten Flugparametern.

## 2.1 Datenerfassung

Zur Erfüllung des Auftrages wurde von Seiten der BfG eine Befliegung der drei genannten Gebiete in Flughöhen 3000 m, 2500 m, 1600 m und 800 m über Grund beauftragt. Jedes der drei Gebiete zeichnet sich durch besondere Eigenschaften aus. Innerhalb des Gebietes A sind neben dem Rhein als freifließendem Fluss auch staugeregelte Flüsse (Mosel und Lahn) vorhanden, deren Oberflächenbeschaffenheit sich von einem freifließendem Fluss unterscheidet.



Abb. 2: Befliegungsgebiet A (links) und Moselwehr mit Stausee (rechts)

Neben einer Gesamtübersicht sind in Abb. 2 und 3 einzelne Gebiete detaillierter dargestellt. Ein wesentliches Untersuchungsgebiet stellt die Mosel in Gebiet A dar. In diesem Gebiet befindet sich ein Stauwehr, der damit verbundene Moselstausee, Schleusenanlagen und ein freifließender Teil (Abb. 2, links). Die Mündung in den Rhein ist in Abb. 3 links dargestellt. Darüber hinaus steht das Gebiet A stellvertretend für ein urbanes Gebiet. Die Insel Niederwerth (Abb. 3, rechts) zeichnet sich durch eine Vielzahl unterschiedlicher Oberflächenarten, wie etwa Kiesstrand, Buschwerk, Wälder, landwirtschaftliche Flächen, Wiesen etc. aus.

Die Befliegungsgebiete B und C (Abb. 4) zeichnen sich vor allen durch mehrere enge Flusskurven (Gebiet B) und hydraulisch schwierige Verhältnisse (Gebiet C) aus. Die Hydraulik im Bereich des Gebietes C ist hydronumerisch äußerst komplex zu modellieren. Dies wird insbesondere durch enge Kurven, einer sehr variablen Gewässerbetttopografie sowie Felsen und Sandbänke verursacht. Infolgedessen wird in diesem Gebiet eine relativ stark variierende Wasseroberfläche erwartet.

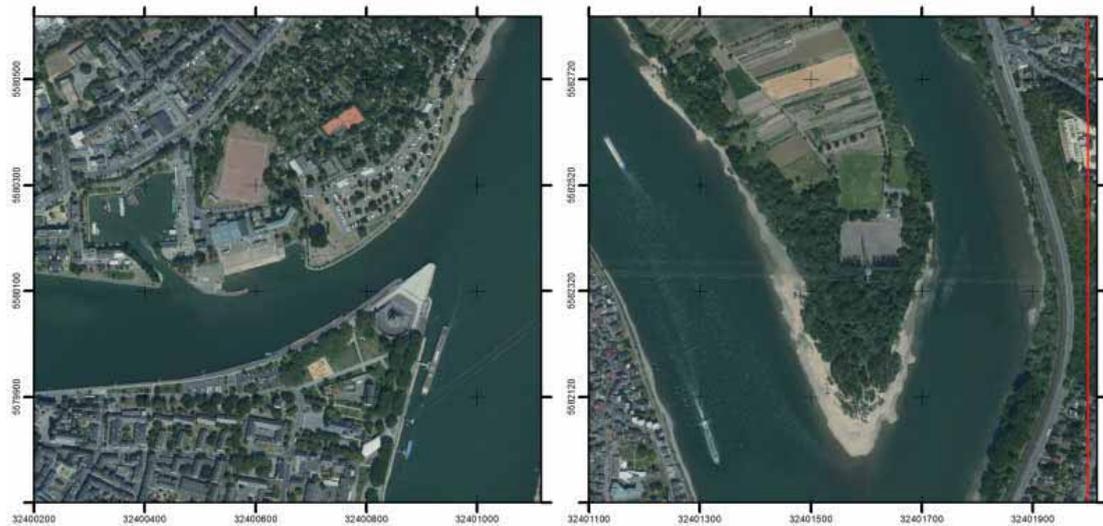


Abb. 3: Moselmündung (links) und Niederwerth (rechts)

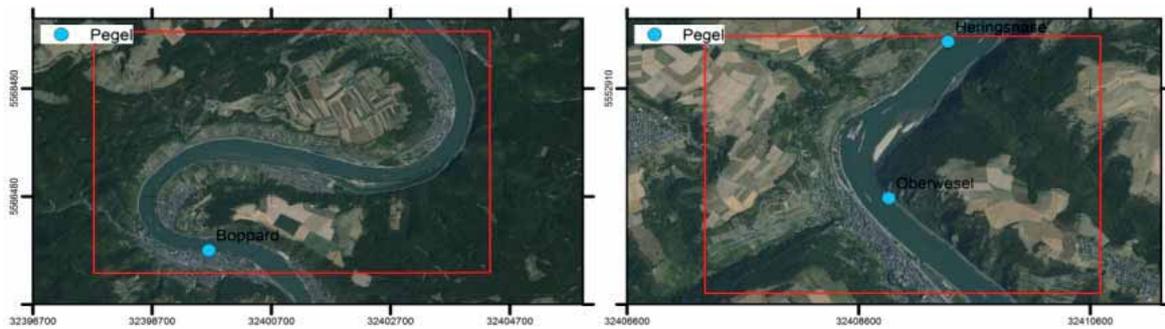


Abb. 4: Befliegungsgebiete B und C

Die Befliegung erfolgte am 30./31.10.2019. Zu diesem Zeitpunkt lagen die Wasserstände etwa 70 cm oberhalb des GIW20 und es wurden am Rhein Sichttiefen von 1,3-1,4 m ermittelt. Erfasst wurden die Daten aus Flughöhen von 3000 m (Gebiet A+B), 2500 m (Gebiet C), 1600 m (Gebiet A+B) und 800 m (Gebiet B+C) über Grund. Alle Scannerparameter sind in Tabelle 1 dargestellt. Die Auslenkung des Scanners betrug  $10^\circ$ , so dass ein Field of View von  $20^\circ$  realisiert wurde.



Abb. 5: Wetterbedingungen und Messflugzeug während der 3000m Befliegung

Tab. 1: Genutzte Flug- und Scanparameter

Gebiet	Flughöhe [m]	Geschw. [m/s]	Punkt- dichte je Streifen [Punkte/m <sup>2</sup> ]	Punktabstand (Durchschnitt)		Lücke zwischen Scans [m]	Scan- rate [Hz]	Puls- rate [kHz]
				radial [m]	tangential [m]			
A/B	3000	77	~25	0,15	0,18	-	20	45
C	2500	77	~32	0,15	0,15	-	20	45
A/B/C	1600	77	~55	0,15	0,07	0,97	20	50
A/B/C	800	77	~110	0,15	0,04	2,00	20	50

## 2.2 Referenzdaten

Als Referenzdaten liegen in den jeweiligen Untersuchungsgebieten mehrere Pegel sowie Flächen an Land vor. Bei den Pegeln handelt es sich um offizielle Pegel der Wasserstraßen- und Schifffahrtsverwaltung des Bundes. Diese setzen sich immer aus mindestens einer Pegellatte als maßgebenden Sensor, sowie mindestens einem kontinuierlich arbeitenden Wasserstandssensor zusammen (WEIß & SUDAU, 2012). Die durch den kontinuierlich arbeitenden Wasserstandssensor erfassten Daten werden innerhalb der Pegelmessstelle zu 15-Minuten Werten zusammengefasst, welche danach öffentlich zugänglich gemacht werden. Eine Nutzung der Rohmesswerte ist nicht möglich. Bei den Pegeln der Wasserstraßen- und Schifffahrtsverwaltung wird im Rahmen von Pegelprüfungen sichergestellt, dass die an den Pegelsensoren erfassten Werte auch den manuellen Ablesungen der Pegellatte entsprechen. Hierbei ist zu beachten, dass die Ablesegenauigkeit einer Pegellatte auf etwa 1 cm beschränkt ist. Darüber hinaus findet jährlich oder in zweijährigen Abständen eine Kontrolle der Höhenlage statt. Durch diese Kontrollen wird sichergestellt, dass die Pegellatten höhenmäßig korrekt montiert sind bzw. die Lattennullpunkte die zuvor definierten Höhen aufweisen. Gemäß angewandeter Pegelvorschrift sind an dieser Stelle Abweichungen von  $\pm 1$  cm zulässig (LAWA, 2018). Von Seiten der WSV erfolgten zeitnah zur Befliegung sowohl Kontrollen der Höhenlage der Pegelnullpunkte wie auch Pegelprüfungen. Die Ergebnisse zeigten, dass alle Pegel korrekt gearbeitet haben.

Neben den Pegeln existieren noch eine Vielzahl von Referenzflächen an Land, welche durch die WSV bereitgestellt bzw. von der BfG erfasst wurden. Zum Zeitpunkt der Verfassung dieses Artikels lagen noch keine finalen (streifenabgeglichenen) Ergebnisse vor. Aussagen zur Genauigkeit der erfassten Beobachtungen an Land können somit zu diesem Zeitpunkt noch nicht gemacht werden.

## 2.3 Datenprozessierung und Auswertung

Die komplette Datenprozessierung erfolgt beim Auftragnehmer. Die Ableitung der für die Punktkoordinaten notwendigen Trajektorie des Flugzeugs erfolgte im Postprocessing unter Nutzung der SAPOS<sup>®</sup>-Referenzstation Koblenz. Alle erfassten Rohbeobachtungen des SPL100 wurden mittels Leica HxMap Software in ETRS89/UTM Punktkoordinaten umgerechnet und streifenweise als LAS-Dateien geliefert. Im Rahmen einer ersten Klassifizierung wurden Streupunkte der Luft bzw. solare Fehlechos, sowie weitere Fehlmessungen eliminiert. Die Ergebnisse der weiterführenden Auswertung (u.a. Streifenabgleich, Plausibilisierung und Klassifizierung) liegen zum jetzigen Zeitpunkt noch nicht vor.

## 2.4 Erste Ergebnisse

Einen ersten Eindruck vermitteln die in Abb. 6 dargestellte Rohpunktcloud im Bereich des Deutschen Ecks (Mündungsbereich der Mosel (rechts) in den Rhein (links)). Die Flughöhe dieser Daten betrug 3000 m. Numerisch ergibt sich bei einer Strahldivergenz von 0,08 mrad aus einer Flughöhe von 3000 m ein Footprintdurchmesser im Nadir von ~0,24 m. Bei einem relativ geringen Auslenkwinkel von 10° werden die daraus resultierenden Vergrößerungen des Footprints bzw. die Variationen aufgrund der Roll- und Pitchbewegung des Flugzeuges vernachlässigt. Anhand der Abb. 6 zeigt sich, dass auch aus großen Höhen relativ dünne Strukturen, wie z.B. Geländer erfasst werden. Es zeigt sich weiterhin, dass im Bereich der beiden Flüsse viele Punkte erfasst wurden. Inwieweit diese die Oberfläche repräsentieren oder Volumenstreuer im Wasser darstellen, kann derzeit noch nicht beurteilt werden.

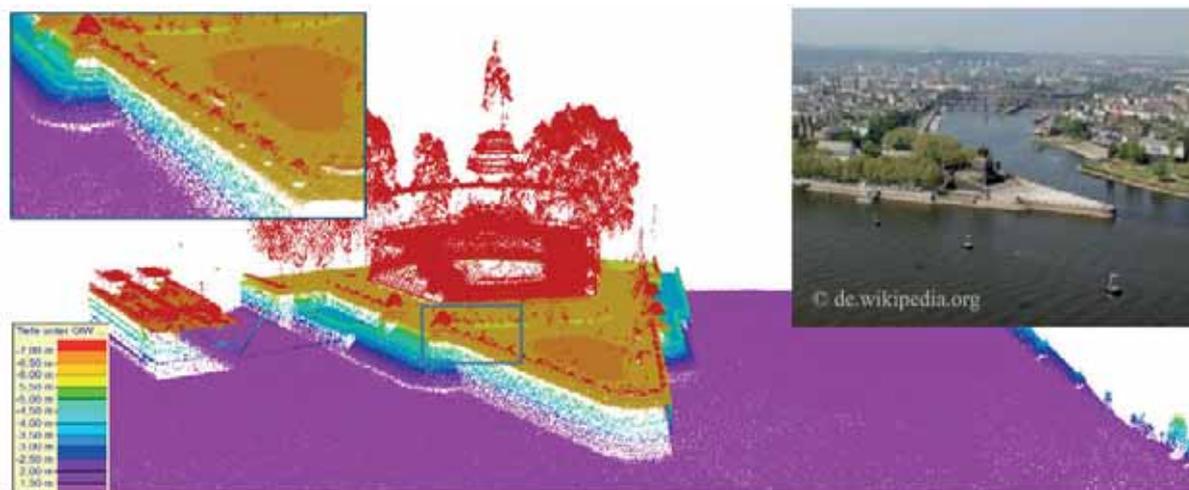


Abb. 6: Rohpunktcloud im Mündungsbereich der Mosel

### 2.4.1 Punktdichte

Bei der Beurteilung der Punktdichte ist zu beachten, dass z.T. Mehrfachechos vorhanden waren. Infolgedessen werden zur ersten Beurteilung der Punktdichte nur Gebiete herangezogen, die vegetationsfrei an Land liegen bzw. es werden nur die „last-pulse“ Daten näher betrachtet. Auch ist bei der Beurteilung der Punktdichte zu beachten, dass aufgrund des kreisförmigen Scannumsters (Palmer Scanner) eine inhomogene Punktverteilung innerhalb des Messstreifens vorliegt. In den Randbereichen ist diese um ein Vielfaches höher, als dies im Nadir des Scanners. In den folgenden Abbildungen ist jeweils ein Segment eines Flugstreifens dargestellt.

Abb. 7 zeigt die Messpunktdichte bei einer Flughöhe von 2500 m. Im linken Bereich werden zusätzlich die Minimaltiefen von 1 m Rasterzellen dargestellt. Aus diesen Werten lässt sich der Verlauf des Rheins ableiten. Aus der Abbildung geht hervor, dass auch im Zentralbereich auf trockenen Flächen Punktdichten von mehr als 30 Punkten/m<sup>2</sup> erreicht wurden (gelb). Im Bereich des Rheins sind weniger Echos vorhanden, wobei die Punktdichte aber dennoch deutlich höher als 10 Punkte/m<sup>2</sup> ist. Inwieweit es sich dabei aber um Wasseroberflächenreflexionen oder Volumenstreuer im Wasser handelt, ist aus den unklassifizierten Daten nicht abzuleiten.

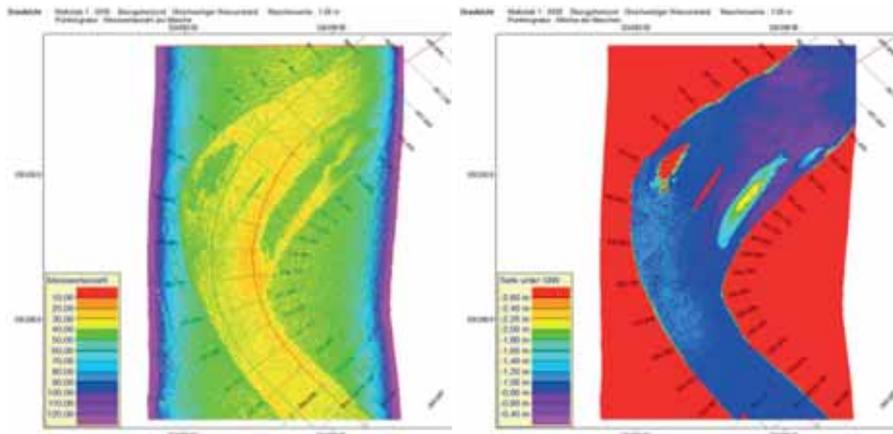


Abb. 7: Messpunktdichte (links) und Geländehöhe (rel. zum GIW) aus 2500m Höhe

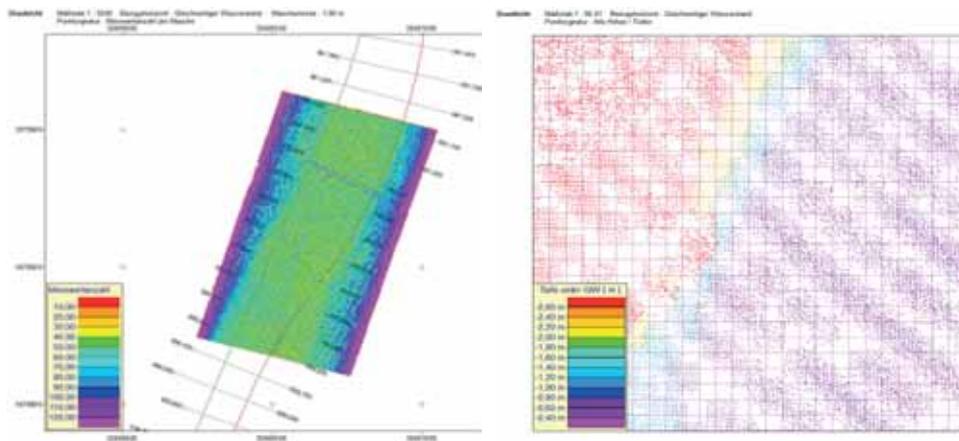


Abb. 8: Messpunktdichte und Bodenmuster/Geländehöhe aus einer Flughöhe von 1600m

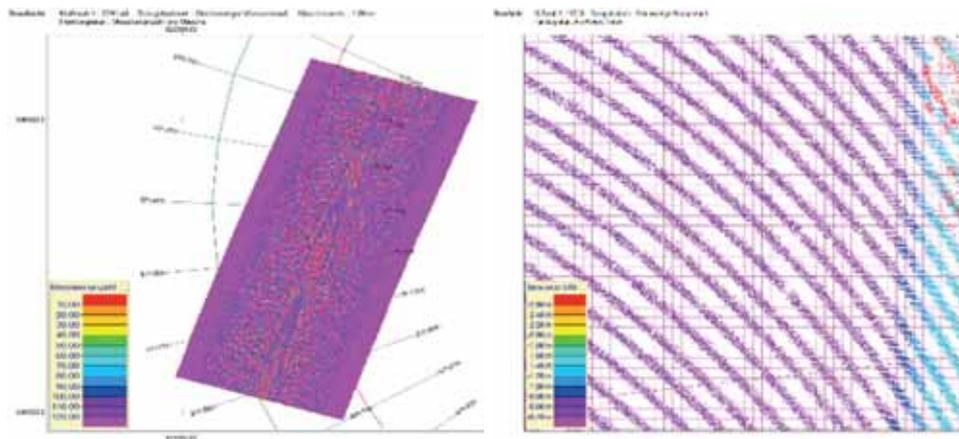


Abb. 9: Messpunktdichte und Bodenmuster/Geländehöhe aus einer Flughöhe von 800m

Aus den Flughöhen von 1600 m bzw. 800 m ergeben sich deutlich höhere Punktdichten (Abb. 8 und 9). Es ist dabei aber zu beachten, dass zwischen den einzelnen Scanzeilen Lücken entstehen,

welche auf die bewusste Unterschreitung der für den Scanner konzipierten Mindestflughöhe von 2000 m (LEICA 2020) zurückzuführen sind. Für eine bessere Darstellung des Scannmusters werden in den rechten Bereichen der Abbildungen Detailaufnahmen der Punktverteilung am Boden dargestellt. Zur Verdeutlichung des Scannmusters werden nur die vor dem Flugzeug erfassten Lasermesspunkte dargestellt. Zusätzlich ist ein 1 m Raster eingeblendet.

### 2.4.2 Relative Streuung der Bodenpunkte zueinander

Bedingt durch den bisher nicht erfolgten Streifenabgleich sind absolute Genauigkeitsaussagen noch nicht möglich. Zur Abschätzung der relativen Genauigkeiten untereinander werden im Folgenden die Standardabweichungen der Bodenpunkte innerhalb eines 1,0 m × 1,0 m Gitters näher beleuchtet. In Abb. 10 sind für das Deutsche Eck im Gebiet A (links) sowie für einen Bereich im Gebiete C (rechts) die Standardabweichungen der gerasterten Beobachtungen innerhalb eines Messstreifens dargestellt. In den grau dargestellten Flächen liegt diese unterhalb von 5 cm. In der Realität handelt es sich bei den grau dargestellten Flächen um Sandbänke bzw. Sandflächen.

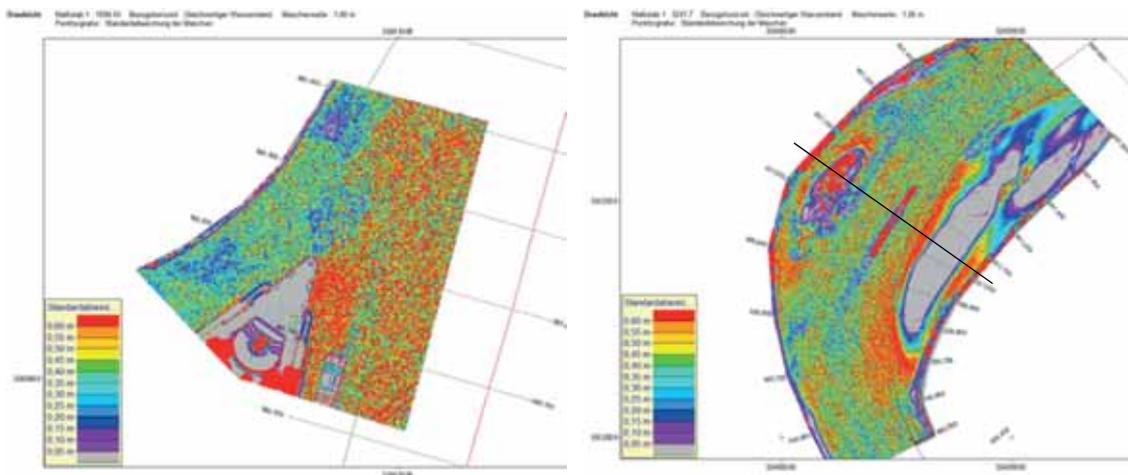


Abb. 10: Standardabweichungen der Beobachtungen innerhalb eines 1m Rasters

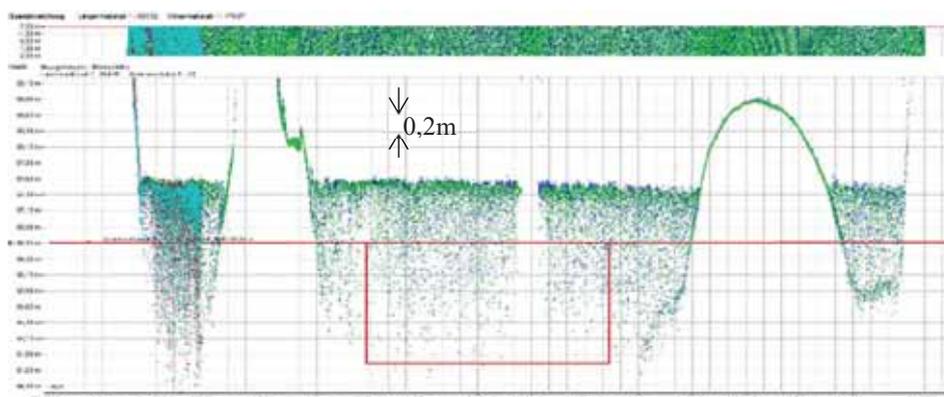


Abb. 11: Querprofil über die Rohbeobachtungen im Gebiet C

Zusätzlich sind in Abb. 11 ein Querprofil über die im Gebiet C liegende Sandbank dargestellt, dessen Lage in Abb. 10 rechts markiert ist. Das Querprofil stellt alle Beobachtungen innerhalb eines 4 m breiten Streifens dar. Alle der Abbildung zugrunde liegenden Daten wurden aus einer Flughöhe von 2500 m erfasst. Neben der sehr geringen Streuung der Beobachtungen über unbewachsenen trockenen Flächen zeigt sich außerdem, dass der Scanner über ein bestimmtes bathymetrisches Potenzial verfügt. Da Leica den Scanner als topografischen Scanner vermarktet, findet an dieser Stelle keine Brechungskorrektur statt.

### 2.4.3 Vergleich mit Pegeldata

Bisher liegen noch keine klassifizierte Daten vor. Folglich kann keine Unterscheidung zwischen Volumenstreuern im Wasser und Wasseroberflächenechos erfolgen. Für erste Abschätzungen wurden die streifenweise die Beobachtungen gerastert und aus jeder Gitterzelle die höchsten 5% aller Beobachtungen extrahiert. Dieses Verfahren wurde auf ein 50 m × 30 m großes Rechteck der Wasseroberfläche unmittelbar neben den Pegeln angewendet (Abb. 12, links). Die Mittelwerte aller Beobachtungen innerhalb des Rechtecks wurden den amtlich gemessenen Wasserständen gegenübergestellt. Es zeigt sich, dass im Mittel die Rohbeobachtungen wenige cm unterhalb der durch den Pegel erfassten Wasseroberfläche liegen. Die Beobachtungen selbst sind im rechten Bereich der Abb. 12 jeweils als Aufriss zusammen mit der am Pegel erfassten Wasserspiegelhöhe dargestellt.

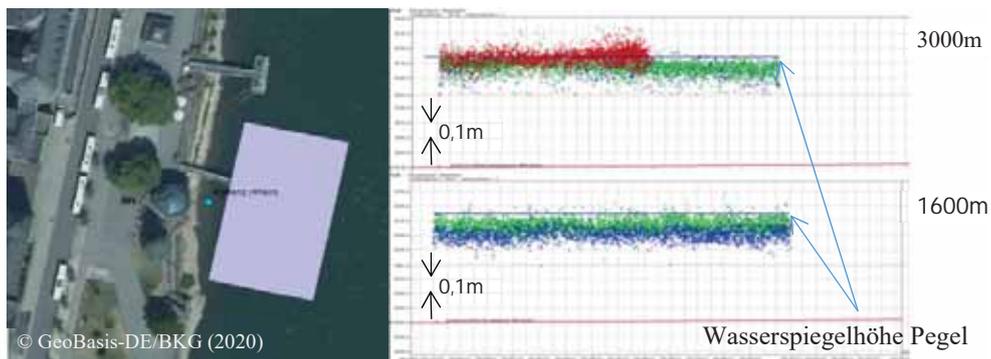


Abb. 12: Erfassung der Wasserspiegelhöhe am Pegel Koblenz

Die Ergebnisse sind exemplarisch für drei Pegel in Tabelle 2 dargestellt. Weiterführende Auswertungen sind erst dann möglich, wenn die finalen Daten durch den Auftragnehmer bereitgestellt werden.

Tab. 2: Abweichung zwischen den mittleren SPL100 Beobachtungen und Pegeldata

Pegel	Gebiet	Flughöhe [m ü.G.]	Wasserspiegelhöhe [m]		Abweichung (Soll-Ist)	
			Pegel	SPL100	Mittel	Stabw.
Heringsnase	C	800	67,27	67,18	0,09	0,03
Oberwesel	C	800	67,81	67,75	0,06	0,03
Koblenz	A	1600	59,20	59,13	0,07	0,04
Heringsnase	C	2500	67,29	67,25	0,04	0,05
Oberwesel	C	2500	67,85	67,81	0,04	0,04
Koblenz	A	3000	59,20	59,13	0,07	0,04

### 3 Fazit & Ausblick

Aufgrund der bisher vorliegenden Beobachtungen zeigt sich, dass der Leica SPL100 über ein großes Potenzial verfügt. Die ersten vorliegenden Rohdaten zeigen, dass die Erfassung trockener Flächen auch aus großen Höhen möglich ist. Die Messpunktdichte zeigt, dass auch auf Wasserflächen genügend Echos vorhanden sind. Inwieweit es sich dabei um Volumenrückstreuer im Wasser bzw. um Wasseroberflächenreflexionen handelt, kann zum bisherigen Stand nicht näher betrachtet werden. Es zeigt sich aber, dass die oberen 5% der Beobachtungen sich bis auf wenige cm den zeitgleich erfassten Pegelbeobachtungen annähern. Weiterführende Auswertungen diesbezüglich mit den finalen Daten sind Gegenstand künftiger Arbeiten. Erst dann ist eine abschließende Bewertung zur Vollständigkeit und Genauigkeit der Erfassung der Wasseroberfläche möglich.

Die Möglichkeiten des Scanners in Bezug auf die topografische Erfassung trockener Flächen scheinen sehr groß zu sein. Aus den größten Flughöhen streuen benachbarte Beobachtungen auf Sand- und Kiesflächen nur innerhalb weniger Zentimeter, wobei bei den zugrundeliegenden Daten keine vertikale Anpassung und kein Streifenabgleich durchgeführt wurden. Es ist daher zu erwarten, dass die später erfolgenden Genauigkeitsanalysen zu positiven Ergebnissen führen. MANDLBURGER et al (2019) haben für denselben Sensor im Vergleich zu konventionellem, wellenform-basierten Laserscanning eine Abhängigkeit des Genauigkeitsverhaltens von der Oberflächenneigung und -beschaffenheit festgestellt. Während die Single Photon LiDAR Technologie im Allgemeinen und der SPL100 im Speziellen bei horizontalen Flächen eine hohe Präzision aufweist, nimmt im Vergleich zu wellenform-basiertem Laserscanning größerer Neigung (z.B. Hausdächer) stärker ab. Die bisherigen Eindrücke bestätigen das geringe Messrauschen auf horizontalen Flächen.

Obwohl es sich bei dem Scanner um einen topografischen Laserscanner handelt, weist dieser dennoch eine gewisse bathymetrische Leistungsfähigkeit auf. Zum Zeitpunkt der Befliegung war der Rhein relativ klar bzw. die Sichttiefen mit ~1,4 m relativ groß. Augenscheinlich ergeben sich aus den Querprofilen Eindringtiefen von etwa der einfachen Sichttiefe. Mit den finalen Aussagen ist im Verlauf des Jahres zu rechnen.

### 4 Literaturverzeichnis

- WSV 2014: Der frei fließende Rhein - Relevante Niedrig- und Mittelwasserstände, Report der Wasserstraßen- und Schifffahrtsverwaltung des Bundes
- DEGNAN, J., 2016: Scanning, Multibeam, Single Photon Lidars for Rapid, Large Scale, High Resolution, Topographic and Bathymetric Mapping. *Remote Sensing* **8**(11), 923-958.
- DEGNAN, J., 2018: Arborne And Spaceborne Single Photon 3D Imaging Lidars. In *Proceedings of 21<sup>st</sup> International Workshop on Laser Ranging*, Canberra, Australia.
- BUND/LÄNDER ARBEITSGEMEINSCHAFT WASSER (LAWA), MINISTERIUM FÜR UMWELT, KLIMA UND ENERGIEWIRTSCHAFT BADEN-WÜRTTEMBERG, 2018: Leitfaden zur Hydrometrie des Bundes und der Länder – Pegelhandbuch, Kulturbuch-Verlag GmbH
- LEICA, 2020: Datenblatt des Seica SPL100 Single Photon LiDAR Sensors. <https://leica-geosystems.com/products/airborne-systems/topographic-lidar-sensors/leica-spl100>.

- MANDLBURGER, G. & JUTZI B., 2018: Feasibility investigation on Single Photon Lidar based water surface mapping. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* **IV-1**, 109-116.
- MANDLBURGER, G. & JUTZI, B., 2019: On the Feasibility of Water Surface Mapping with Single Photon LiDAR. *ISPRS International Journal of Geo-Information* **8**(188), 1-23.
- MANDLBURGER, G., LEHNER, H. & PFEIFER, N., 2019: A Comparison of Single Photon and Full Waveform LiDAR. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **IV-2/W5**, 397-404.
- PFEIFER, N., MANDLBURGER, G., OTEPKA, J. & KAREL, W., 2014: OPALS – A framework for airborne laser scanning data analysis. *Computers, Environment and Urban Systems* **45**, 125-126.
- WEIß, R. & SUDAU A., 2012: Geodätische Aspekte von Höhen und Höhenänderungen der Pegel und Pegelfestpunkte im Küstenbereich, *Hydrologie und Wasserbewirtschaftung (HyWA)* HW 56, Heft 5.

# Automatic Generation of Training Data for Land Use and Land Cover Classification by Fusing Heterogeneous Data Sets

SYLVIA SCHMITZ<sup>1,2</sup>, MARTIN WEINMANN<sup>1</sup>, UWE WEIDNER<sup>1</sup>,  
HORST HAMMER<sup>2</sup> & ANTJE THIELE<sup>1,2</sup>

*Abstract: Nowadays, automatic classification of remote sensing data can efficiently produce maps of land use and land cover, which provide an essential source of information in the field of environmental sciences. Most state-of-the-art algorithms use supervised learning methods that require a large amount of annotated training data. In order to avoid time-consuming manual labelling, we propose a method for the automatic annotation of remote sensing data that relies on available land use and land cover information. Using the example of automatic labelling of SAR data, we show how the Dempster-Shafer evidence theory can be used to fuse information from different land use and land cover products into one training data set. Our results confirm that the combination of information from OpenStreetMap, CORINE Land Cover 2018, Global Surface Water and the SAR data itself leads to reliable class assignments, and that this combination outperforms each considered single land use and land cover product.*

## 1 Introduction

A wide range of research in the field of environmental science relies on information on land use and land cover (LULC). For instance, such knowledge is used to develop climate models, to plan cities and agricultural land use efficiently, or to support forest management (PIELKE et al. 2011; HANSEN et al. 2013). In order to respond to the demand for up-to-date LULC products, a variety of methods has been developed to derive appropriate class assignments from remote sensing data. The use of remote sensing data, in particular satellite data, is justified by the fact that large areas can be covered, in most cases even with short revisit time. In recent years, the accessibility of remote sensing data has increased significantly due to the use of Earth observation satellites such as Sentinel, Landsat or TerraSAR-X. Because of the large amount of data available, machine learning methods offer high potential for the automatic analysis and interpretation of these data. For the task of LULC classification, particularly methods related to supervised learning are used. These methods include, among others, Random Forest (VAN BEIJMA 2014), Support Vector Machine (SVM) (HUANG et al. 2002) and Convolutional Neural Networks (CNNs) (MOHAMMADIMANESH 2019). In particular, algorithms with a high number of trainable parameters, such as those found in the area of deep learning, require many annotated training data for successful learning processes. The high volume and diversity of the training data significantly contributes to the convergence of the training process, as well as the ability of the classifier to generalize to unseen samples. Manual

---

<sup>1</sup> Karlsruhe Institute of Technology, Institute of Photogrammetry and Remote Sensing, Englerstraße 7, D-76131 Karlsruhe, E-Mail: [Sylvia.Schmitz, Antje.Thiele, Martin.Weinmann, Uwe.Weidner]@kit.edu

<sup>2</sup> Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Gutleuthausstraße 1, D-76275 Ettlingen, E-Mail: [Sylvia.Schmitz, Antje.Thiele, Horst.Hammer]@iosb.fraunhofer.de

annotation of data usually requires a considerable amount of expert knowledge and can only be realized with an enormous amount of time.

One apparent approach to bypass manual annotation is to use freely available LULC products e.g. crowdsourced data such as OpenStreetMap (OSM) (HAKLAY & WEBER 2008) or products developed and provided as part of Earth observation programs e.g. Copernicus. Such products contain a high amount of valuable information suitable for an initial labelling of remote sensing data. Especially the tagged data layers from OSM often serve as class labels within the training of classification and segmentation methods. For example, BRIAN & KOTARO (2016) extract OSM polygons of the categories *natural* and *landuse* to train a Random Forest in order to realise LULC mapping using time-series of satellite imagery. Further, to generate training data for forest/non-forest classification, PEKKARINEN et al. (2009) use class labels extracted from a map of the CORINE Land Cover (CLC) programme.

However, available LULC products also comprises some disadvantages that interfere with the training of robust and reliable classifiers. For example, one LULC product may have an insufficient minimum mapping unit (MMU), while others do not cover the entire training area, do not represent all target classes, or contain misclassifications. The validity of the training labels derived from a LULC product is therefore not ensured. However, the reliability of training data in supervised learning algorithms is a fundamental factor for the performance of the resulting classifier. Another issue concerning this strategy is that LULC products build on the analysis of data from the past. Thus, they may not represent the state at the acquisition time of the data to be annotated. In this paper, we present an approach to generate reliable annotations of remote sensing data with minimal manual labelling effort by fusing various sources of information. Instead of using only one LULC product as annotation, several complementary products are combined. The fusion of information eliminates the weaknesses of individual products and enhances their combined advantages at the same time. Besides various LULC products, information about the LULC state at the time of data acquisition are included. This information is derived from characteristic features of the data to be annotated. Using the Dempster-Shafer evidence theory, all information are combined to create a reliable training data set. We demonstrate the performance of the approach applying the automatic annotation on SAR data recorded over the German Wadden Sea.

This paper is organized as follows: in Section 2, we introduce the study area and the data collected in that region. In Section 3, all sources of information are described that are included in the automatic annotation and the approach to fuse these information is presented. Subsequently, in Section 4, the results that can be achieved with the approach are shown. Finally, in Section 5, we summarize the results, draw conclusions accordingly and give including suggestions for future work.

## 2 Study area – German Wadden Sea

The presented approach is to be used on the automatic annotation of SAR data captured from tide-influenced environments. With the superior goal of generating high-resolution geodata for coastal monitoring, airborne SAR data were acquired over the German Wadden Sea. As representative test areas, data were recorded over Otzumer Balje, as well as from the Elbe estuary Medem channel



Fig. 1: Polarimetric S-band SAR image mosaic, captured over the Medem channel (left) and Otzumer Balje (right). In the background, an image of the corresponding region from GoogleEarth is displayed

(see Fig. 1). The data acquisition was performed by the F-SAR system, developed at the German Aerospace Center (Deutsches Zentrum für Luft- und Raumfahrt; DLR) (HORN 2009). F-SAR is an airborne SAR system that is capable of simultaneously capturing fully polarimetric SAR data at different wavelengths. In the measurement campaigns, the X- and S-band antennas were used, which are able to realize single-pass polarimetric interferometry. In addition to the single-pass interferometry, repeat-pass measurement mode was used to assure baseline flexibility. The flights took place in February and July 2019 on days with low tide. We have chosen times with low tides to observe large areas of dry fallen mudflats, as the structures found there are of particular interest for coastal protection. By observing variations in the morphology of tide ways, tidal flats, beaches and dunes, it is possible, for instance, to predict the formation of sandbanks and the erosion of dunes.

The presented approach automatically generates reference data for a part of the collected data in order to enable the training of various classifiers by means of supervised learning processes. These classifiers shall serve to distinguish between water, mudflats, different vegetation classes and man-made objects.

### 3 Automatic data annotation

With the aim of generating training data, a representative section of available airborne data is annotated automatically with respect to five classes: *water*, *mudflats*, *high vegetation* (e.g. trees, high shrubs), *ground* (including fields, meadow and arable land) and *man-made objects*. An overview of the approach for the generation of training data is illustrated in Fig. 2. The input data comprises classified polygons from OSM data, the CORINE Land Cover (CLC2018) map, information from the Global Surface Water (GSW) data set as well as interpreted SAR data (coherence mask and entropy mask). In order to enable a pixel-wise fusion, all information are first mapped to a uniform grid. To simplify the fusion we introduce a homogenisation of the class nomenclature. Consequently, there are up to five class assignments for each pixel to be annotated. These are regarded as evidences and fused using the Dempster-Shafer rule of combination. The resulting combined basic probability assignment function is used to determine a final class label whose certainty is

quantified with a confidence value. Finally, the resulting training data set is composed of all class labels whose corresponding confidence value exceeds a fixed threshold value.

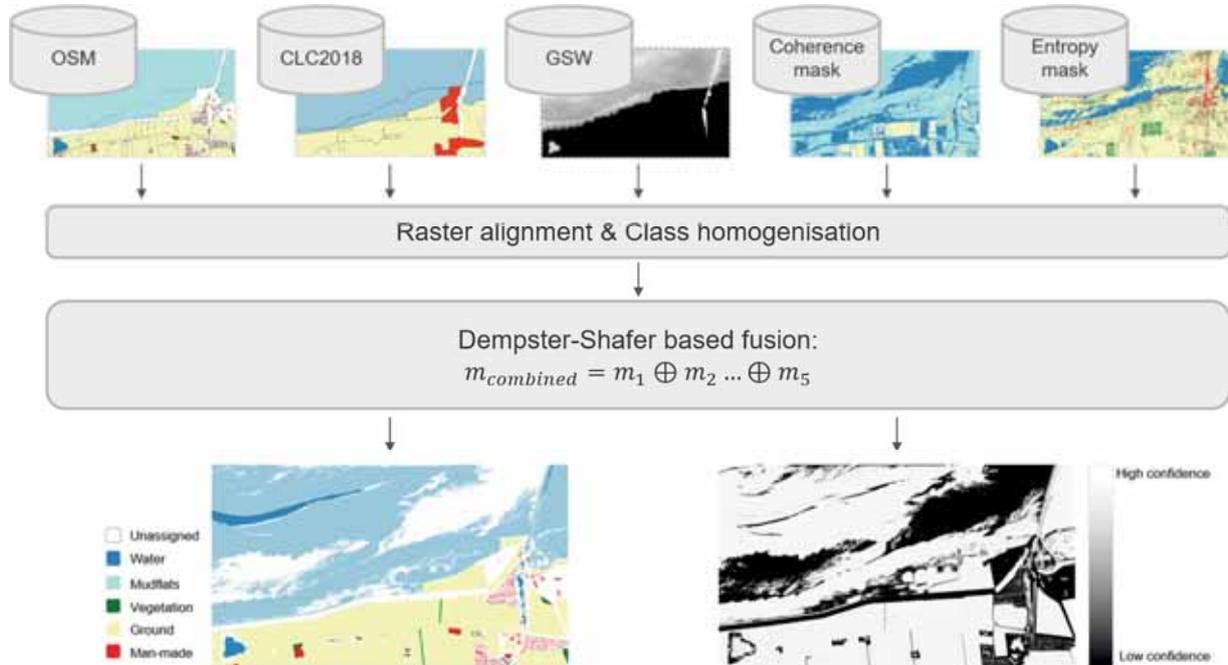


Fig. 2: Overview of the approach to generate annotated training data by fusing several LULC products (OSM, CLC2018) and interpreted SAR data (Coherence mask, Entropy mask)

### 3.1 Information base

The following three LULC products serve as the information basis for automatic annotation:

1. OpenStreetMap (OSM) data (HAKLAY & WEBER 2008), a crowdsourced data set to which everyone can voluntarily make his own contribution.
2. CORINE Land Cover 2018 (CLC2018) (BÜTTNER et al. 2004), a data product developed within the European Earth observation programme Copernicus by the analysis and interpretation of satellite images and in-situ data.
3. Global Surface Water (GSW) (PEKEL et al. 2016), a data set, which was also developed as part of Copernicus to reflect the temporal distribution of water surfaces over the last 3.5 decades.

The main characteristics of the data sets are listed in the first three columns of Tab. 1 and described in detail in the following sections.

#### 3.1.1 OpenStreetMap data

The collaborative project OpenStreetMap aims at creating and maintaining freely available map material, which captures mainly information related to land use, transport infrastructure and buildings as well as the localisation of national and coastal borders. Since 2004 information has been collected on a voluntary basis by more than two million OSM users. The generation of map sections is a two-step process. First, raw data is recorded, for example by GPS measurements collected while following roads, paths or rivers. In a second step, users transfer the data in form of polygons,

lines and points to the map and assign them with attributes, so-called *tags*. These tags consist of key – value pairs. The key indicates the general topic or category of an object (e.g. *landuse*) and the value assigns the explicit class (e.g. *crop*).

Tab. 1: Properties of each included source of information.

	OSM	CLC2018	GSW	COHERENCE MASK	ENTROPY MASK
<b># CLASSES</b>	61	44	-	3	5
<b>SPATIAL RES- OLUTION</b>	-	MMU: 25ha	30m × 30m	1 m × 1 m	1 m × 1 m
<b>TEMPORAL COVERAGE</b>	2004 to to- day	2017 to 2018	1984 to 2018	July 2019	July 2019
<b>DATA BASIS</b>	Voluntarily contributed data	Sentinel-2 Landsat-8	Landsat-5 Landsat-7 Landsat-8	F-SAR data	F-SAR data

Concerning the goal of annotating acquired SAR data with LULC classes, the publicly available OSM data can make a significant informational contribution. The information we require within our approach predominantly lies in polygons that are tagged with the category (key) *natural*, *waterways*, *buildings*, *man-made* and *landuse*. Thus, all polygons located within the test area and annotated with one of those keys are extracted from the OSM data along with their attributes. The information contained therein forms an important building block for the automatic data annotation approach.

### 3.1.2 CORINE Land Cover 2018

The EU's CORINE Land Cover (CLC) programme was initiated in 1985 with the aim of generating and providing of standardised localised geographical information on the land cover of all EU member states. The latest data set produced within the project refers to the status of LULC in 2018 (CLC2018). Research groups from 39 member countries are involved in the development of the LULC map. On a national basis, satellite data are evaluated by visual interpretation and semi-automated processes. In addition, information from in-situ data and GIS data are partially integrated. The resulting product maps LULC with a minimum mapping unit (MMU) of 25 hectares to 44 classes, following the hierarchical 3-level CORINE nomenclature. The coarsest level (Level 1) consisting of *artificial surfaces*, *forest and semi natural areas*, *wetlands* and *water bodies* corresponds directly to the five class assignments that we want to map in our training data.

As shown in Fig. 2, the CLC2018 data set provides less detailed spatial information compared to polygons extracted from OSM data. However, it offers complete coverage of the test area in contrast to the sparse coverage of OSM polygons. Thus, it complements the OSM data on the one hand and serves for the validation of the OSM polygons on the other hand.

### 3.1.3 Global Surface Water

The GSW-Explorer provides water data sets developed by the European Commission's Joint Research Centre in order to support scientists and decision-makers, especially in the field of water management. The data sets were developed based on the analysis of more than 3 million satellite images acquired by Landsat 5, 7, and 8 between March 1984 and October 2015. The resulting maps represent information for the entire globe on the local and temporal occurrence of water surfaces, as well as their changes. Natural and artificial water surfaces are mapped with a spatial resolution of  $30\text{ m} \times 30\text{ m}$ .

Since the test area consists mainly of water and mudflats, which need to be distinguished, information regarding water occurrences are of particular importance. For this reason, we include the Surface Water Occurrence map of the GSW-Explorer in the annotation process. This map displays the spatial and temporal variations of surface water in a single product, giving the frequencies with which water surfaces occur. With the background knowledge that the data to be annotated were captured at low tide, this information supports the distinction between water, mudflats and land.

## 3.2 Information from SAR data

The three selected products do not necessarily reflect the state of LULC at the desired time, but represent results that may be outdated or show a mixture of past occurrences. However, for the annotation of currently available remote sensing data, it is of great importance to map LULC at the time of acquisition. This is particularly important in the tidal areas we are investigating, as water and mudflats are constantly changing. In order to be able to map the corresponding water levels correctly, it is necessary to include information derived directly from present SAR data. In order to obtain relevant information, meaningful SAR features are extracted. To differentiate between water, mudflats and land areas, we make use of the multi-path constellation that was applied during data acquisition. Based on two co-registered S-band SAR images with VV polarisation, a coherence image is calculated. Due to the reflective properties of water, low coherence values result in the corresponding areas. With the addition of X-band amplitude images, a threshold value method is used to assign the label *water*, *mudflat* or *land* to each pixel. For example, pixels with low coherence and low amplitude belongs to class *water*. The resulting coherence mask can reliably distinguish between these three classes. However, the differentiation of different land cover classes (*ground*, *vegetation*, *man-made*) is not performed. Therefore, another classification mask based on polarimetric features is generated. Following the approach proposed by CLOUDE & POTTIER (1997), the complex-valued polarimetric scattering matrix is decomposed to extract entropy (H), alpha ( $\alpha$ ) and anisotropy values. Based on the resulting H/ $\alpha$ -plane, each pixel receives one of the class labels: *water*, *mudflat*, *ground*, *vegetation* or *man-made*. The corresponding interpretation of the H/ $\alpha$ -plane for class assignments is shown in Fig. 3.

The maps resulting from the two methods (hereinafter referred to as entropy mask and coherence mask) are used in the annotation process to incorporate knowledge about the LULC at the appropriate time.

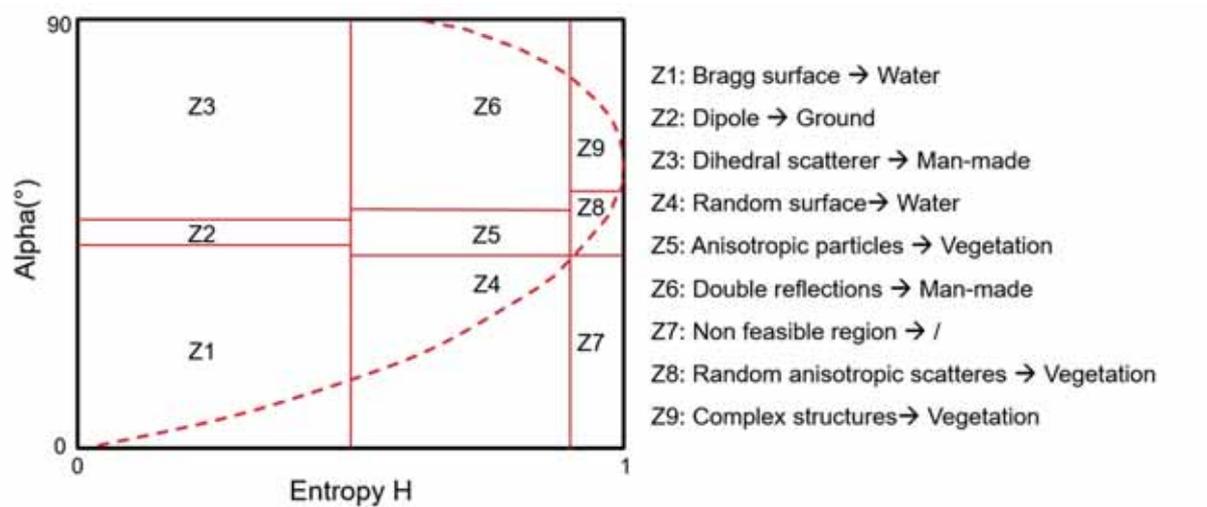


Fig. 3: The H-alpha plot (Cloude & Pottier, 1997) shows how Entropy and alpha values are used to distinguish between five LULC classes

### 3.3 Homogenisation of class nomenclatures

To simplify the final fusion, information stored in the presented LULC products are mapped to up to five output classes: *water*, *mudflats*, *high vegetation*, *ground* and *man-made objects*. To this

Tab. 2: Mapping of OSM tags, CLC nomenclature and water occurrences to five output classes.

	OSM	CLC2018	GSW
<b>WATER</b>	Natural: <i>water</i> Waterway: <i>all values</i>	Water bodies	Occurrence > 0.95
<b>MUDFLATS</b>	Natural: <i>wetland, mud</i> Landuse: <i>salt_pond</i>	Wetlands	0.95 > Occurrence > 0.5
<b>VEGETATION</b>	Landuse: <i>vineyard, orchard, greenfield</i> Natural: <i>grassland, greenfield, scrub, heath, forest</i>	Forest and semi-natural areas	Occurrence < 0.5
<b>GROUND</b>	Landuse: <i>greenhouse_horticulture, farmland, meadow</i> Natural: <i>fell, bare_rock, sand, rock, cliff</i>	Agricultural areas	
<b>MAN-MADE</b>	Landuse: <i>industrial, commercial, retail, quarry, construction, allotments, farmyard, garages</i> Man_made: <i>all values</i> Office: <i>all values</i> Buildings: <i>all values</i> Shop: <i>all values</i>	Artificial surfaces	

end, a custom assignment scheme, presented in Tab. 2 is applied. For the mapping of information from OSM data, several key – value pairs are grouped and one of the output classes per group is assigned. In the case of CLC2018 data, the class mapping is straightforward adopting the Level-1 labels. The GSW data provide values between 0% and 100%. Two thresholds were chosen to separate *water*, *mudflat* and *land* (see Tab. 2). In this case we do not differentiate between *vegetation*, *ground* and *man-made*, because the GSW data do not contain any information about different land types.

### 3.4 Fusion methodology

As a result of the processing steps described above, for each pixel, we obtain up to five statements regarding the class label. In order to come to an overall decision about the class label, we fuse the partially conflicting statements using the Dempster-Shafer theory (DST). The DST, which has been developed by DEMPSTER and SHAFER (1968, 1992), provides a tool for knowledge representation and reasoning with uncertainty in expert systems. It can be applied to combine information (evidences) from different sources into an overall statement. In contrast to traditional probability theory principles, the Dempster Shafer approach allows for the representation of ignorance. Formally, the theory is described by the following definitions:

- (1) *Frame of discernment*: In DST, all mutually exclusive hypotheses form a finite non-empty set  $\theta$ , commonly known as frame of discernment.  $\Omega(\theta)$  denotes the power set comprising all  $2^\theta$  subsets of  $\theta$ , including itself and the null hypotheses  $\emptyset$ . For the case we investigate, the distinction between five LULC classes, this results in the frame of discernment:

$$\theta = \{\text{'water'}, \text{'mudflats'}, \text{'vegetation'}, \text{'ground'}, \text{'man-made'}\}$$

The resulting power set  $\Omega(\theta)$  contains all 32 combinations of these hypotheses, but only some subsets are of interest. One relevant example is the subset

$$A = \{\text{'vegetation'}, \text{'ground'}, \text{'man-made'}\},$$

which implies that the LULC class is one of the land classes *vegetation*, *ground* or *man-made*.

- (2) *Basic probability assignment function*: To quantify the certainty with which each subset of  $\Omega(\theta)$  is supported by a distinct evidence, the basic probability assignment (bpa) function  $m$  is introduced. It is defined as a mapping that assigns a value between 0 (no belief) and 1 (total belief) to every element of the power set. The formal representation is given by:

$$m : \Omega(\theta) \rightarrow [0, 1]. \quad \text{Eq. 1}$$

with

$$m(\emptyset) = 0, \quad \sum_{A \subset \Omega(\theta)} m(A) = 1. \quad \text{Eq. 2}$$

The bpa of an individual hypothesis  $m(A)$  quantifies the portion of the total belief committed exactly to the subset  $A$ . However, it does not provide any information about the amount of support across the different subsets of  $A$ . This shall be demonstrated by an example related to our application. Based on the information extracted from GSW, which indicate that the frequency of water occurrences was 0, we can assign a high portion of the total belief to the subset  $A = \{\text{'vegetation'}, \text{'ground'}, \text{'man-made'}\}$ . Note that by this the amount of support of the individual classes, e.g. *vegetation*, is not given.

To represent ignorance, the total amount of belief 1 is given to the full set  $\theta$  and consequently no belief (0) to any other subset of  $\theta$ . Applied to our example, this representation is reasonable if a dataset does not provide any information for one considered pixel. This occurs especially in connection with the sparse OSM data.

- (3) *Dempster's rule of combination*: In order to combine several bpa functions, derived from evidence from independent sources, Dempster's rule of combination can be applied. Given two bpa functions  $m_1$  and  $m_2$  in the same frame of discernment  $\theta$ , a joint bpa is calculated using the following equation:

$$m_{1,2}(A) = (m_1 \oplus m_2)(A) = \frac{1}{K-1} \sum_{B \cap C = A} m_1(B) m_2(C), \quad \text{Eq. 6}$$

$$\text{with } K = \sum_{B \cap C = \emptyset} m_1(B) m_2(C) > 0. \quad \text{Eq. 7}$$

Thus,  $K$  indicates the amount of conflict among the evidences, obtained from two different sources of information. The larger the conflict, the less informative is the resulting combination of sources. The following generalisation results for the combination of more than two sources of evidence to an aggregated hypothesis:

$$m(A) = (m_1 \oplus m_2 \oplus \dots \oplus m_n)(A), \quad \text{Eq. 8}$$

$$\frac{1}{1-K} \sum_{\cap_{i=1}^n E_i = A} m_1(E_1) \cdot m_2(E_2) \cdot \dots \cdot m_n(E_n). \quad \text{Eq. 9}$$

We use the previously explained evidence theory to fuse the information contained in OSM, CLC2018, GSW, entropy mask and coherence mask for each pixel. As already indicated, the frame of discernment is determined by:

$$\theta = \{\text{'water'}, \text{'mudflats'}, \text{'vegetation'}, \text{'ground'}, \text{'man-made'}\}.$$

The first substantial step of our approach includes the assignment of bpa functions to the respective evidence, derived from the different information sources. In other words, we need to quantify the reliability of the evidence that we extract from the selected LULC products. This is accomplished by including prior knowledge about the class significance of the different data sets. This prior knowledge is obtained by comparing class predictions of the data sets to a manually labelled training area. Following the approach proposed by DENG et al. (2006), we construct bpa functions  $m_i^{\varphi_k}$  for each class  $c_i$  ( $i = 1 \dots N$ ) in every data set  $\varphi_k$  ( $k = OSM, CORINE, GSW, entropy\ mask, coherence\ mask$ ) based on recall and precision rates. Using the labelled training area, that contains the true classes  $c_i$ , we compute the confusion matrices  $C_{\varphi_k}$  to quantify the performance of a data set  $\varphi_k$ :

$$C_{\varphi_k} = \begin{bmatrix} n_{11} & n_{12} & \dots & n_{1N} \\ n_{21} & n_{22} & \dots & n_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ n_{N1} & n_{N2} & \dots & n_{NN} \end{bmatrix}. \quad \text{Eq. 10}$$

Here,  $n_{ij}$  represents the number of pixels belonging to the true class  $c_i$  and classified by the data set  $\varphi_k$  as  $c_j$ . Note that we exclude all unclassified pixels. With the precision rate  $r_{ij}^r$  and recall rate  $r_{ij}^p$  given by:

$$r_{ij}^r = \frac{n_{ij}}{\sum_{j=1}^{N+1} n_{ij}} \quad \text{and} \quad r_{ij}^p = \frac{n_{ij}}{\sum_{i=1}^{N+1} n_{ij}}, \quad \text{Eq. 11}$$

we derive the corresponding matrices:

$$C_{\varphi_k}^r = \begin{bmatrix} r_{11}^r & r_{12}^r & \dots & r_{1N}^r \\ r_{21}^r & r_{22}^r & \dots & r_{2N}^r \\ \vdots & \vdots & \ddots & \vdots \\ r_{N1}^r & r_{N2}^r & \dots & r_{NN}^r \end{bmatrix} \quad \text{and} \quad C_{\varphi_k}^p = \begin{bmatrix} r_{11}^p & r_{12}^p & \dots & r_{1N}^p \\ r_{21}^p & r_{22}^p & \dots & r_{2N}^p \\ \vdots & \vdots & \ddots & \vdots \\ r_{N1}^p & r_{N2}^p & \dots & r_{NN}^p \end{bmatrix}. \quad \text{Eq. 12}$$

Based on  $C_{\varphi_k}^r$  and  $C_{\varphi_k}^p$ , two bpa functions  $m_{\varphi_k,i}^r$  and  $m_{\varphi_k,i}^p$  are derived for every class  $c_i$  with:

$$m_{\varphi_k,i}^r(\{c_i\}) = \frac{r_{ii}^r}{\sum_{j=1}^N r_{ji}^r} \quad \text{and} \quad m_{\varphi_k,i}^p(\{c_i\}) = \frac{r_{ii}^p}{\sum_{j=1}^N r_{ji}^p}, \quad c_i \in 2^\theta, \quad \text{Eq. 13}$$

$$m_{\varphi_k,i}^r(\{\theta\}) = 1 - m_{\varphi_k,i}^r(\{c_i\}) \quad \text{and} \quad m_{\varphi_k,i}^p(\{\theta\}) = 1 - m_{\varphi_k,i}^p(\{c_i\}), \quad c_i \in 2^\theta, \quad \text{Eq. 14}$$

$$m_{\varphi_k,i}^r(\{A\}) = 0 \quad \text{and} \quad m_{\varphi_k,i}^p(\{A\}) = 0, \quad \forall A \in 2^\theta \setminus \{c_i, \theta\}. \quad \text{Eq. 15}$$

Note that the considered classes  $c_i$  can represent both singletons as well as subsets such as  $\{\text{'vegetation'}, \text{'ground'}, \text{'man-made'}\}$  of  $\theta$ .

To generate the final bpa functions  $m_i^{\varphi_k}$ , that reflect the ability of a dataset  $\varphi_k$  to recognise the class  $c_i$  the two bpa functions  $m_{\varphi_k,i}^r$  and  $m_{\varphi_k,i}^p$  are combined:

$$m_i^{\varphi_k} = m_{\varphi_k,i}^r \oplus m_{\varphi_k,i}^p. \quad \text{Eq. 16}$$

After the generation of bpa functions based on a training area, we can produce a fused map for an independent test area. For each pixel  $p$  within the test area, every data set assigns a predicted class  $c_p^{\varphi_k}$ . According to the predictions, the corresponding bpa functions  $m_i^{\varphi_k}$  with  $i = c_p^{\varphi_k}$  are selected. For the case of unclassified samples,  $m^{\varphi_k}(\theta) = 1$  is used. For the fusion of the five bpa functions, we apply Dempster's rule of combination and obtain the combined bpa function:

$$m_p^{\text{combined}} = f(\varphi_1, c_p^{\varphi_1}) \oplus f(\varphi_2, c_p^{\varphi_2}) \oplus \dots \oplus f(\varphi_5, c_p^{\varphi_5}). \quad \text{Eq. 17}$$

Here  $f(\varphi_k, c_p^{\varphi_k})$  denotes the associated bpa  $m_i^{\varphi_k}$ , belonging to the predicted class  $c_p^{\varphi_k}$  for pixel  $p$ . The last step contains the decision-making regarding the final class, depending on  $m^{\text{combined}}$ . For this, we apply the pignistic transformation (SMETS et al., 1994):

$$c_p^{combined} = \arg \max \left\{ \sum_{A \in \theta, x \in A} \frac{1}{|A|} \frac{m_p^{combined}(A)}{1 - m_p^{combined}(\emptyset)} \right\}, x \in \theta. \quad \text{Eq. 18}$$

As a result, we obtain the final class label  $c_p^{combined}$  for each pixel  $p$ . Furthermore, information about the reliability of the class assignments is stored for each label. To quantify this information, we use the maximum value calculated in the argument of Eq. 16, denoted as confidence value. For our goal of generating annotated training data, the complete annotation of each individual pixel is not necessary. More important is the correctness of the annotation of pixels that build up the training data set. Therefore, the intended training data set consists only of pixels whose class assignments are marked with a high confidence value.

## 4 Results

We use the presented approach in order to annotate a part of SAR images, captured over the German Wadden Sea. For the evaluation of the proposed fusion method, a test area was selected that covers a coastal strip south of Spiekeroog. Fig. 4 shows the SAR image of the test area overlaid with manually annotated polygons. The annotated area was divided into training and validation areas, whereby in both areas all output classes are included. The training area comprises 2,098,003 annotated pixels with a pixel size of  $1 \text{ m} \times 1 \text{ m}$ . The annotated pixels of this area are used to determine the bpa functions, which define the reliability of the data to be fused. For this purpose, we calculate the confusion matrices and derive the corresponding recall and precision matrices for each included LULC product.

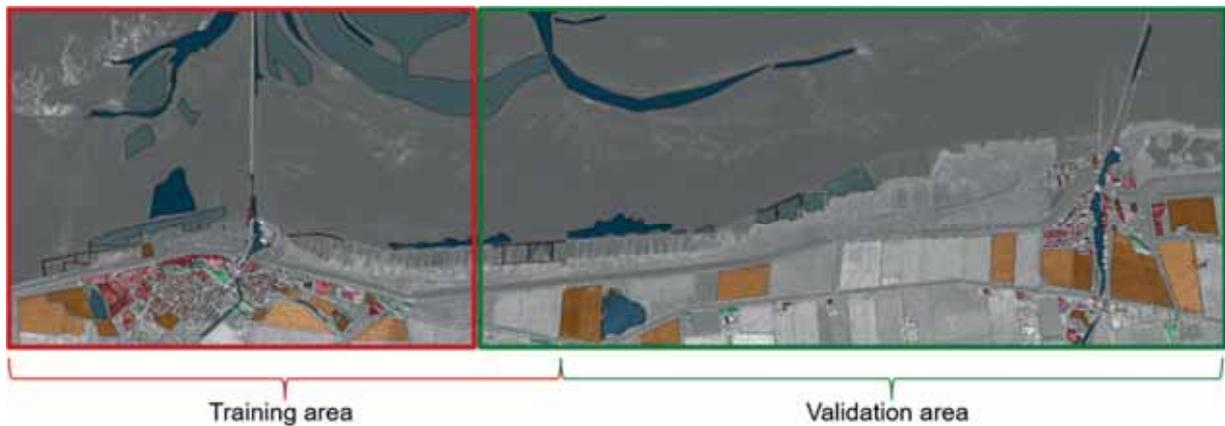


Fig. 4: Test area, divided into training and validation area. The S-band amplitude image is overlaid with manually annotated polygons

The validation area contains 1,851,813 annotated pixels. Within this area, class labels are determined by fusing five LULC products, using the presented approach. For the automatic selection of reliable annotations, the fused map is filtered using a threshold value of 0.9 on the provided confidence values. In this way, 69.10% of the pixels within the validation area are assigned a class

label. 1,456,739 of these pixels have a manual annotation and are used for evaluation. In the following, we present the obtained results and compare them with the performance of the individual LULC products OSM and CLC2018. For the evaluation, the confusion matrices including precision and recall rates per class are shown. For the validation area, we obtain performance scores that are summarized in Tab. 3 to 5. With precision values between 0.87 and 0.99, the annotation resulting from the fusion provides highly reliable class information. The separation of the five classes is successful and only few misclassifications remain after the fusion and the pixel selection. The only poor performance value is the recall rate of 0.62 for *man-made objects*. This class is often mistakenly classified as *mudflats*, since tiny coastal buildings are not captured in any of the used LULC products. Compared to the performance of OSM and CLC2018, the strength of the fused result is particularly evident in the distinction between *water* and *mudflats*. Due to the constant change in the tidal influenced area, the static LULC products OSM and CLC2018 do not reflect the water level at the time of data acquisition. However, the fused result takes into account time developments and current information, so that significantly higher precision and recall rates can be achieved. As Tab. 4 shows, the performance of the OSM data is high regarding the distinction of the three land classes. This strength is also exploited in the fusion, so the strong performance is reflected in the fused result.

Tab. 3 Confusion matrix for the fused result (FUSED).

	Man-Made	Ground	Vegetation	Mudflats	Water	Recall
Man-Made	<b>46,433</b>	3,837	156	24,104	204	0.62
Ground	2,319	<b>807,730</b>	0	0	0	0.997
Vegetation	864	1,567	<b>25,158</b>	0	0	0.91
Mudflats	0	0	0	<b>269,435</b>	0	1.0
Water	0	677	0	16,056	<b>282,903</b>	0.94
Precision	0.94	0.99	0.99	0.87	0.99	

Tab. 4 Confusion matrix for OSM data.

	Man-Made	Ground	Vegetation	Mudflats	Water	Recall
Man-Made	<b>46,697</b>	3,837	156	21,583	9,180	0.57
Ground	2,319	<b>807,730</b>	0	0	4,010	0.99
Vegetation	864	1,567	<b>25,158</b>	0	0	0.91
Mudflats	0	0	0	<b>269,473</b>	0	1.0
Water	0	732	0	107,611	<b>133,971</b>	0.55
Precision	0.94	0.99	0.99	0.67	0.91	

Tab. 5 Confusion matrix for CLC2018 data.

	Man-Made	Ground	Vegetation	Mudflats	Water	Recall
Man-Made	<b>63,582</b>	49,544	0	31,936	11,984	0.40
Ground	10,809	<b>904,358</b>	0	0	45	0.99
Vegetation	80	33,583	<b>0</b>	0	0	0
Mudflats	0	0	0	<b>269,473</b>	0	1.0
Water	4,290	87,624	0	168,787	<b>215,718</b>	0.45
Precision	0.81	0.84	0	0.57	0.95	

A qualitative assessment can be deduced from Fig. 5 that shows the annotation of the different LULC products (OSM, CLC2018; FUSED) over the validation area. It is noticeable that, in many cases, reliable class labels from the OSM data will prevail in the context of the fusion for the land area. In the tidal influenced area, the dominance of the OSM classes is significantly lower. Due to stronger conflicts between the fused LULC products in the tidal area, the occurrence of unclassified areas increases. By excluding pixels with high uncertainty, wrong class assignments are prevented. Thus, only those pixels remain in the tidal area for which a reliable distinction can be made between *water* and *mudflats*.



Fig. 5: Assigned classes by CLC2018, OSM and FUSED (our result)

## 5 Conclusion

In order to apply supervised learning to train LULC classifiers that rely solely on remote sensing data, a high amount of annotated data is necessary. In this paper, an approach is proposed to annotate remote sensing data automatically. Using the example of SAR data from the German Wadden Sea, we show how to use Dempster-Shafer theory to fuse information from LULC products in order to generate reliable training data. The presented results show that reliable class assignments are achieved by fusing data derived from OSM, CLC2018, GSW and from polarimetric and coherent features of the SAR data itself. The results testify that the fused map represents the LULC classes at the acquisition time of the SAR data more precisely than a single data set. It receives details in contrast to the roughly resolved CLC2018 map, and represents all required classes as opposed to the GSW map, coherence mask and entropy mask. By including information based on SAR features, the fusion result also maps the current water occurrence and distinguishes reliably between *water* and *mudflats*. This makes it more suitable as a training set than for example only OSM data. Additionally, the automatic selection of certain class assignments also greatly reduces misclassifications, which is especially relevant for the use as training data. In the future, we intend to use the proposed method for the annotation of spatially and temporally distributed SAR images to generate a diversified training data set. Such a data set will enable the robust training of CNN-based classifiers with a high number of parameters. In particular, a training from scratch is facilitated, which is essential for the analysis of SAR data to learn SAR-specific low-level and high-level features.

## 6 Acknowledgement

This study is part of the GeoWAM project that is funded by the German Federal Ministry of Transport and Digital Infrastructure within the framework of the Modernity Fund ("mFUND").

## 7 Literature

- BÜTTNER, G., FERANEC, J., FAFFRAIN, G., MARI, L., MAUCHA, G. & SOUKUP, T., 2004: The CORINE Land Cover 2000 Project. *EARSeL eProceedings* **3**(3), 331-346.
- DEMPSTER, A. P., 1968: A generalization of Bayesian inference. *Journal of the Royal Statistical Society* **30**, 205-247.
- DENG, X., LIU, Q., DENG, Y. & MAHADEVAN, S., 2016: An improved method to construct basic probability assignment based on the confusion matrix for classification problem. *Information Sciences* **340**, 250-261.
- HAKLAY, M. & WEBER, P., 2008: OpenStreetMap: User-generated street maps. *IEEE Pervasive Computing* **7**(4), 12-18.
- HANSEN, M. C., POTAPOY, V. P., MOORE, R., HANCHER, M., TURUBANOVA, S. A., TYUKAVINA A., THAU, D., STEHMAN, S. V., GOETZ, S. J., LOVELAND, T. R., KOMMAREDDY, A., EGOROV, A., CHINI, L., JUSTICE, C. O. & TOWNSHEND, J. R. G., 2013: High-resolution global maps of 21st-century forest cover change. *Science* **342**(6160), 850-853.
- HORN, R., JAEGER, M., KELLER, M., LIMBACH, M., NOTTENSTEINER, A., PARDINI, M., REIGBER, A. & SCHEIBER, R., 2017: F-SAR - Recent upgrades and campaign activities. 18<sup>th</sup> International Radar Symposium (IRS), 1-10.
- HUANG, C., DAVIS, L. S. & TOWNSHEND, J. R. G., 2002: An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing* **23**(4), 725-749.
- MOHAMMADIMANESH, F., SALEHI, B., MAHDIANPARI, M., GILL, E. & MOLINIER, M., 2019: A new fully convolutional neural network for semantic segmentation of polarimetric SAR imagery in complex land cover ecosystem. *ISPRS Journal of Photogrammetry and Remote Sensing* **151**, 223-236.
- PEKEL, J., COTTAM, A., GORELICK, N. & BELWARD, A. S., 2016: High-resolution mapping of global surface water and its long-term changes. *Nature* **540**(7633), 418-422.
- PEKKARINEN, A., REITHMAIER, L. & STROBL, P., 2009: Pan-European forest/non-forest mapping with Landsat ETM+ and CORINE Land Cover 2000 data. *ISPRS Journal of Photogrammetry and Remote Sensing* **64**(2), 171-183.
- PIELKE SR, R. A., PITMAN, A., NIYOGI, D., MAHMOOD, R., MCALPINE, C., FAISAL, H., GOLDEWIJK, K., NAIR, U., BETTS, R. & FALL, S., 2011: Land use/land cover changes and climate: modeling analysis and observational evidence. *Wiley Interdisciplinary Reviews: Climate Change*, **2**(6), 828-850.
- SHAFER, D., 1992: Dempster-Shafer theory. *Encyclopedia of Artificial Intelligence* **1**, 330-331.
- SMETS, P. & KENNES, R., 1994: The transferable belief model. *Art. Intelligence* **66**(2), 191-234.
- VAN BEIJMA, S., COMBER, A. & LAMB, A., 2014: Random forest classification of salt marsh vegetation habitats using quad-polarimetric airborne SAR, elevation and optical RS data. *Remote Sensing of Environment* **149**, 118-129.

# Neues Normungsverfahren „InSAR – Radarinterferometrie für die Bodenbewegungserfassung“

NICO ADAM<sup>1</sup>, RALF REULKE<sup>2</sup> & VOLKER SPRECKELS<sup>3</sup>

*Zusammenfassung: In letzten Jahren ist das Interesse an der Radarinterferometrie für die Bodenbewegungserfassung stark gewachsen und kommerzielle und behördliche Dienste stehen zur Verfügung. Das Ziel des neuen DIN-Normungsverfahrens ist es, vereinheitlichte Begriffe, validierte Verfahren und Standards bei den Vermessungsprodukten und der Qualitätssicherung zu etablieren und damit fachkundig und rechtssicher für die betriebliche und behördliche Praxis einsetzbar zu machen. Dieser Beitrag beschreibt den Kontext der Initiative, den Stand der bisherigen Validierungen und Standardisierungen und Beispiele von aktuellen Bodenbewegungsdiensten und Corner-Reflektoren.*

## 1 Einleitung

Die SAR Interferometrie ist ein Fernerkundungsverfahren, mit dem die Topographie und Bewegungen der Erdoberfläche mit Hilfe eines kohärenten Radars bildhaft gemessen werden. Die genaueste Technik für das Bodenbewegungsmonitoring ist die Persistent Scatterer Interferometry (PSI). Diese ermöglicht unter Nutzung von Radar-Satelliten die großflächige und kontinuierliche Vermessung mit Millimeter-Genauigkeit. Auf Grund dieser Messempfindlichkeit gibt es zahlreiche Anwendungen wie z.B. die Detektion und Vermessung von Auswirkungen durch Energiespeicherung, Geothermie, Bergbau, Gas- und Ölförderung sowie von vulkanischen und tektonischen Bewegungen. Daher ist in den letzten Jahren in der Forschung aber auch seitens behördlicher und kommerzieller Nutzer das Interesse an dieser neuen fernerkundlichen Vermessungstechnik stark gewachsen. Unterstützt wird dieser Trend durch die allgemeine Verfügbarkeit von Radar-Satelliten mit synthetischer Apertur (SAR). Die hochaufgelöst erfassenden TerraSAR-X und die großflächig beobachtenden Sentinel-1 Satelliten sind speziell für interferometrische Anwendungen entwickelt worden und bieten eine kontinuierliche und erschwingliche Datenbasis. In Anwendungsprojekten wurden interferometrische Auswertungen bisher für kleine Gebiete aufwändig einzeln beauftragt und oft mit Prototypensoftware realisiert. Die Nutzer solcher Daten haben jedoch beständig den Wunsch geäußert, über vergleichbare Auswertungen, Referenzvermessungen, vereinheitlichte Begriffe und validierte Verfahren zu verfügen. Aus Aspekten der Qualitätssicherung und auch der Akzeptanz wurde daher das Thema „Radar-Norm“ bzw. „InSAR-Norm“ bereits mehrfach von Anwendern aus Industrie sowie Bundes- und Länderbehörden und Kommunen an den Normungsausschuss herangetragen. Mit dem neuen Normungsverfahren „InSAR – Radarinterferometrie für die Bodenbewegungserfassung“ wird das nun end-

---

<sup>1</sup> Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Institut für Methodik der Fernerkundung, Münchner Str. 20, 82230 Weßling, E-Mail: Nico.Adam@dlr.de

<sup>2</sup> Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Institut für Optische Sensorsysteme, Rutherfordstrasse 2, 12489 Berlin, E-Mail: Ralf.Reulke@dlr.de

<sup>3</sup> RAG Aktiengesellschaft, SB BG Standort-/Geodienste - BG N Nachbergbau, Wilhelmstrasse 98, 44649 Herne, E-Mail: Volker.Spreckels@rag.de

lich möglich (REULKE 2015). Nutzer sind Behörden, Kommunen, die Wirtschaft und die Bevölkerung. Die Kenntnis über die Stabilität des Untergrundes bietet ein sehr breites Anwendungsspektrum für praktisch alle Entscheidungsträger, die auf Geodaten angewiesen sind, von der Raum- und Verkehrswegeplanung bis zum Küsten- und Hochwasserschutz.

## 2 Stand der Validierungen und Standardisierungen

Die satellitengestützte SAR Interferometrie hat sich in den 90-er Jahren mit den ERS-Satelliten schnell in der Wissenschaft etabliert (MASSONNET et al. 1994; GOLDSTEIN 1995; ZEBKER et al. 1997). Im Februar 2000 wurde mit der Shuttle Radar Topography Mission (SRTM) diese Technik zum ersten Mal zur operationellen globalen (Breitengradbereich 54°S - 60°N) 3D-Vermessung der Erde eingesetzt (FARR et al. 2007; WERNER 2001; RABUS et al. 2003). Das *Digital Elevation Model* (DEM) war das Hauptmissionsziel, daher ist es durch die in der Raumfahrt üblichen Spezifikations-, Design- und Reviewprozesse standardisiert. Seit 2015 ist das SRTM DEM mit einer Auflösung von einer Bogensekunde (ca. 25 m) in Länge und Breite frei verfügbar. Die dadurch erfolgte zahlreiche Nutzung hat diesen Datensatz zusätzlich validiert. Die deutsche TanDEM-X (*TerraSAR-X add-on for Digital Elevation Measurements*) Mission hatte zum Ziel, eine vollständige und noch genauere 3D-Vermessung der Erde zu liefern (KRIEGER et al. 2007; EINEDER et al. 2013; HUBER et al. 2007). Dieses DEM ist 2016 mit einer Pixelgröße von 0,4 Bogensekunden (ca. 12 m) fertig gestellt worden und steht seit 2018 mit einer reduzierten Pixelgröße von 3 Bogensekunden (ca. 90 m) für die wissenschaftliche Nutzung kostenfrei zur Verfügung (GEOSERVICE DLR 2018). Die hochspezialisierten interferometrischen DEM Missionen setzen und werden selbst durch die umfangreiche Datennutzung Standards.

Anders ist es mit den neu entwickelten Anwendungen der SAR Interferometrie. Hier ist besonders die interferometrische Bodenbewegungserfassung zu nennen, die in der Forschung aber auch auf der behördlichen und kommerziellen Nutzerseite starkes Interesse geweckt hat. Der Grund ist der Durchbruch in der Genauigkeit der relativen Bodenbewegungsbestimmung mit ca. einem Millimeter pro Jahr, der durch FERRETTI et al. (1999; 2000; 2001) mit der PSI Technik erzielt wurde. Die PSI gibt ein Konzept vor, um typische städtische Gebiete mit vorhandenen Daten, einem einheitlichen Vorgehen und bisher unerreichter Genauigkeit zu vermessen. Die ESA hatte das Potential dieser Technik erkannt und auf der FRINGE 2003 entschieden, eine Validierung der unabhängig entwickelten Verfahren d.h. der *Permanent Scatterer*<sup>TM</sup> Interferometrie von FERRETTI et al. (1999; 2000; 2001), dem *Small Baseline Subset* (SBAS) *algorithm* von BERNARDINO et al. (2002), der *Interferometric Point Target Analysis* (IPTA) von WERNER et al. (2003) und andere PSI Entwicklungen von z.B. HANSEN et al. (2005), ADAM et al. (2003) bzw. KAMPES & ADAM (2005) zu fördern. Die Abbildung (Abb. 1) visualisiert einen Ausschnitt des PSIC4 Testgebiets mit dem Gardanne Bergbauggebiet. Die unabhängige Validierung sollte Bedenken bei den Endnutzern, die keine Radarspezialisten sind, ausräumen (RAUCOULES et al. 2007). Eine weitere SBAS-Entwicklung ist das „Intermittent Small Baseline Subset“ (ISBAS) Verfahren (LANARI et al. 2007).

Das Nachfolgeprojekt der ESA hieß Terrafirma und war ein *service element* im *Global Monitoring for Environment and Security* (GMES) Programm. Hauptziel des Projektes war es, einen Europa-weiten *ground motion hazard information service* (Gefahreninformationsdienst bezüg-

lich Bodenbewegungen) in jedem der damaligen 27 EU Staaten anzubieten. Die generierten Bodenbewegungsprodukte der vier *operational service provider* (OSPs) sind online kostenfrei verfügbar (TERRAFIRMA 2005 - 2015). Das Hauptziel wurde durch eine Validierung unterstützt, um die Zuverlässigkeit und Präzision der PSI Bewegungsmessungen zu demonstrieren und Standards bei den PSI Produkten und der Qualitätssicherung zu etablieren (ADAM et al. 2010). Dazu wurde eine Product Validation Workgroup (PVW) gebildet, die zwei sich gegenseitig ergänzende Validierungsstrategien umsetzte. Die eine Validierungsmethode war die Product Validation, welche die geocodierten Ergebnisse mit Referenzdaten (ground truth) verglich (HANSSEN et al. 2008; CROSETTO et al. 2009). Die andere war die Process Validation, welche für die PSI Technik bis dahin noch nicht angewendet wurde und zudem eine Zertifizierung der PSI Prozessierungssysteme als langfristig wirkende Qualitätssicherung beinhaltete (ADAM et al. 2009). Im Rahmen dieser Process Validation wurden Zwischenprodukte der PSI Verarbeitungen mit der DLR Referenzprozessierung verglichen und validiert. Die Abbildung (Abb. 2) visualisiert das Testgebiet mit dem DLR Referenzdatensatz. Im Rahmen dieses Projektes haben ADAM & PARIZZI (2008) auch eine kurzfristig wirkende Qualitätssicherung erarbeitet. Die Qualitätsparameter einer typischen Verarbeitung werden dazu in einem Quality Control Protocol zusammengeführt und dem Endnutzer mit dem PSI Datensatz als standardisierter Teil des Prozessierungsreports bereitgestellt. Die Abbildung (Abb. 3) veranschaulicht die Komponenten des ganzen Terrafirma Services und die jeweiligen Validierungsarten. Die Abbildung (Abb. 4) visualisiert, dass alle Prozessschritte im Ablauf des Terrafirma Bodenbewegungsservices durch die Kurzzeit- und Langzeit-Validierungen abgedeckt sind.

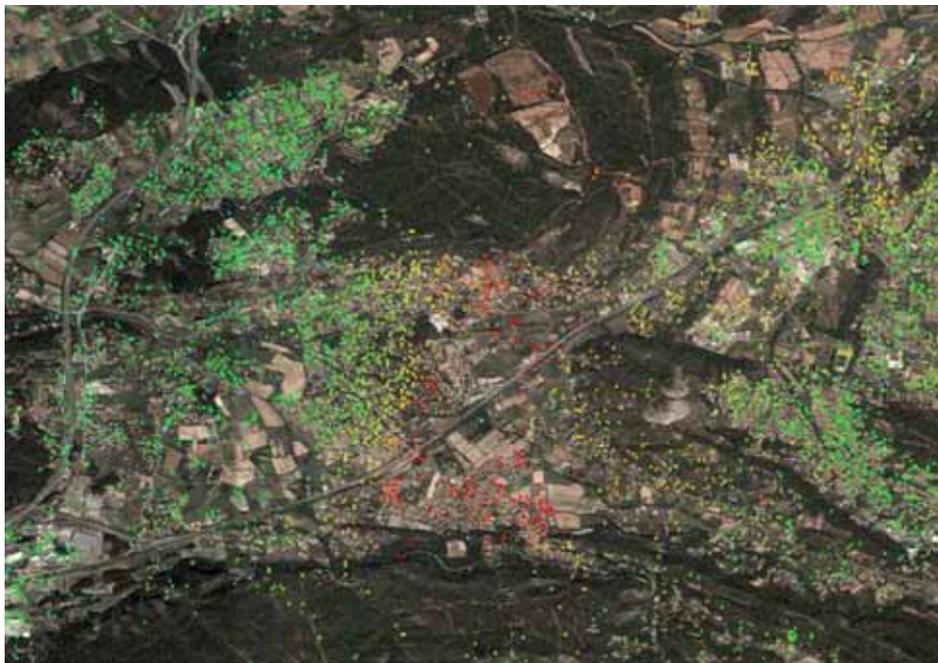


Abb. 1: Ausschnitt des PSIC4 Testgebiets mit dem Gardanne Bergbaugebiet, Quelle: Abbildung 9 in KAMPES & ADAM (2005)

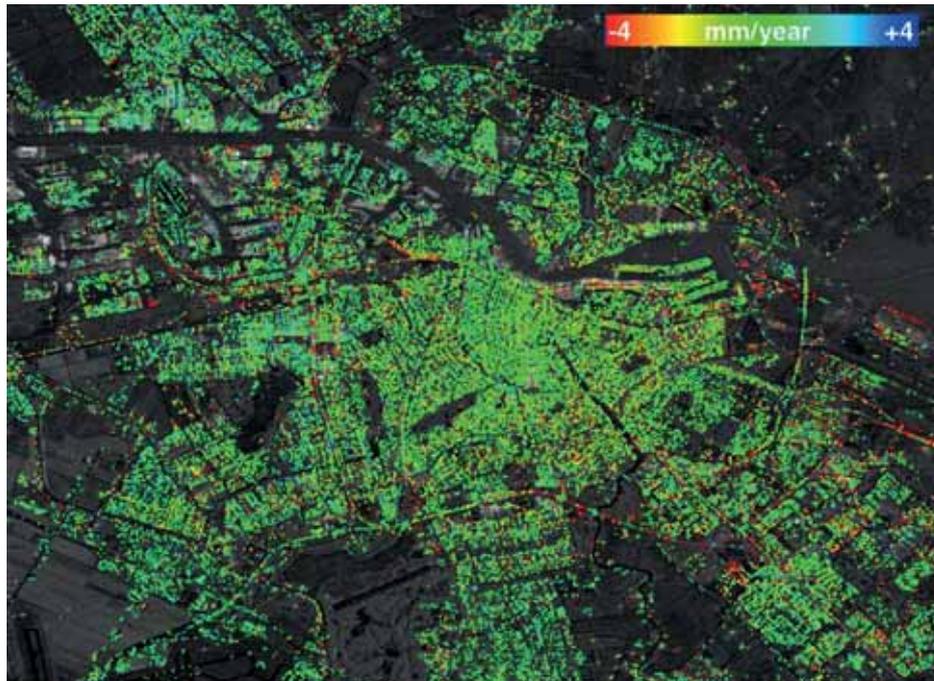


Abb. 2: DLR Referenz-Prozessierung für die Produkt- und Prozessvalidierung (TerraFirma Testgebiet Amsterdam)

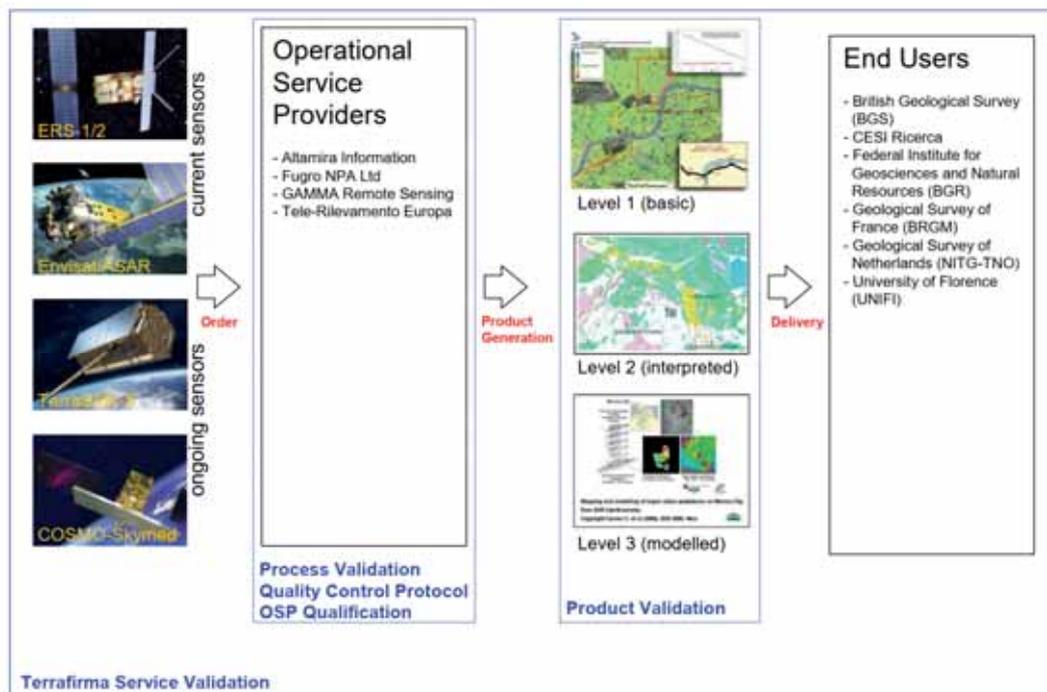


Abb. 3: Komponenten des ganzen TerraFirma Services und die jeweiligen Validierungsarten, Quelle: Abbildung 1 in ADAM et al. (2010)

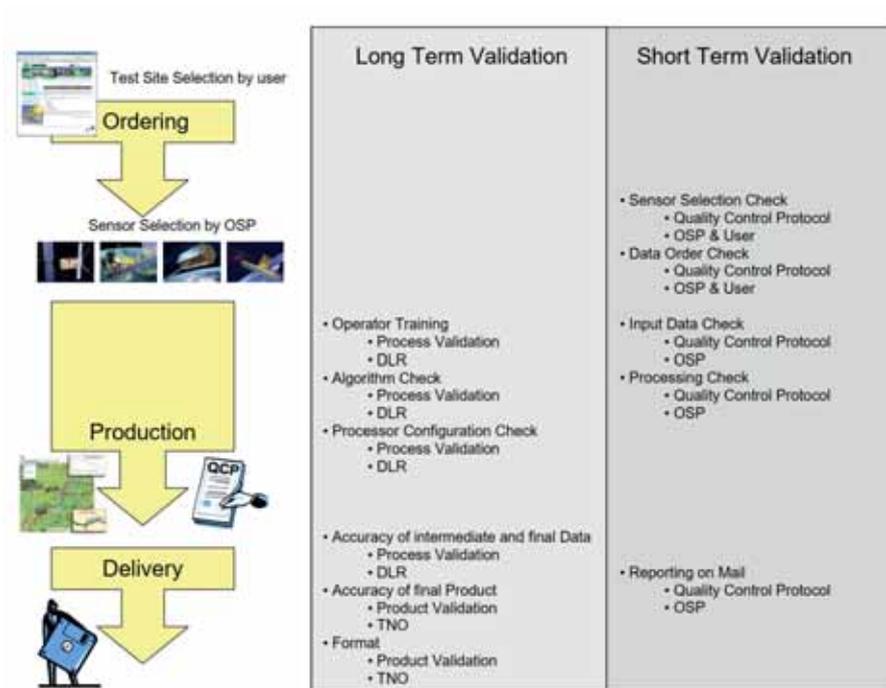


Abb. 4: Serviceelemente (gelb) für L1 Basisprodukte und die Beiträge der Langzeit- und Kurzzeit-Validierung, Quelle Abbildung 3 in ADAM et al. in (2010)

Unabhängig von den internationalen Aktivitäten, gab es auch nationale Initiativen z.B. mit dem Arbeitspapier des DEUTSCHER MARKSCHEIDER-VEREIN E.V. (2013). Die formulierten Grundsätze und Empfehlungen zum Einsatz der Radarinterferometrie sollen diese Vermessungstechnik fachkundig und rechtssicher für die betriebliche und behördliche Praxis einsetzbar machen. Die Grundsätze beinhalten eine Kurzbeschreibung des Messverfahrens, die Planung eines Radarinterferometrieprojektes, die Dokumentation der grundlegenden Projektinformationen, der genutzten Satellitendaten und der Prozessierung aber auch der Qualitätskontrolle. Diesen Grundsätzen hat sich der Bundesverband Erdgas, Erdöl und Geoenergie e.V. (BVEG) angeschlossen (BVEG 2018).

Beim Deutschen Institut für Normung (DIN) wird seit 2015 im Rahmen des Normungsausschuss NA 005-03-02 AA „Photogrammetrie und Fernerkundung“ die Norm 18740 Radarinterferometrie arrangiert (BALTRUSCH & REULKE 2017). Die Themen für die neue InSAR-Norm sind

- Sensoren, Prozessoren, Algorithmen z.B. Informationen zu den und eindeutige Festlegung der diversen Verfahren (z.B. InSAR, SBAS, ISBAS, PSI, SAR-Tomographie)
- Datenverfügbarkeit (z.B. ERS-1/2, Envisat/ASAR, TerraSAR-X, Sentinel-1)
- Anforderungen an die Prozessoren
- Präzise Geolokalisierung und Georeferenzierung von radarinterferometrischen Daten
- Genauigkeitsanforderungen an die Georeferenzierung z.B. über Corner-Reflektoren
- Zusätzliche notwendige Korrekturen (z.B. Atmosphäre)

### 3 Aktuelle Beispiele von Bodenbewegungsdiensten

Neue Normungsanforderungen ergeben sich aus der allgemeinen Verfügbarkeit von Bodenbewegungsdaten. Hier sind, behördlich beauftragte und damit geprüfte Vermessungen und unabhängig promotete Daten zu unterscheiden. Behördlich hat Italien (COSTANTINI et al. 2017) aus ERS, Envisat und COSMO-SkyMed Daten sein ganzes Territorium bezüglich Bodenbewegungen vermessen. Auch in den Niederlanden (NETHERLANDS GEODETIC COMMISSION 2018) und in Deutschland (BGR 2019) ist jeweils eine solche landesweite Vermessung realisiert worden. In Deutschland hat die BGR im Zeitraum von 2016 bis 2019 den Bodenbewegungsdienst Deutschland (BBD) umgesetzt (KALIA et al. 2017; KALIA & FREI 2019). In diesem Projekt werden Copernicus Daten der Satelliten Sentinel-1A und B mit dem vom DLR entwickelten PSI *Wide Area Processing* (WAP) (ADAM et al. 2011) durch die BGR in einem WebGIS kostenfrei bereitgestellt. Die Datennutzer werden mit Nutzungshinweisen (KALIA 2019) bei der Verwertung der Daten unterstützt. Mit dem DLR Prozessierungssystem ist im Rahmen von Pilotstudien und eigenständigen Entwicklungsarbeiten auch eine Bodenbewegungskarte der 90-er Jahre, basierend auf ERS Daten, generiert worden. Die beiden erstellten landesweiten Bodenbewegungskarten sind in der Abbildung (Abb. 5) veranschaulicht. Weitere unabhängige Bodenbewegungsprozessierungen stellt die Firma Geomatic Ventures Limited (GVL) von Großbritannien (GVL 2018a), den Niederlanden (GVL 2018b) und Deutschland (GVL 2018c) unter Verwendung des *Intermittent Small Baseline Subset* (ISBAS) Verfahrens (SOWTER et al. 2013) kostenfrei bereit.

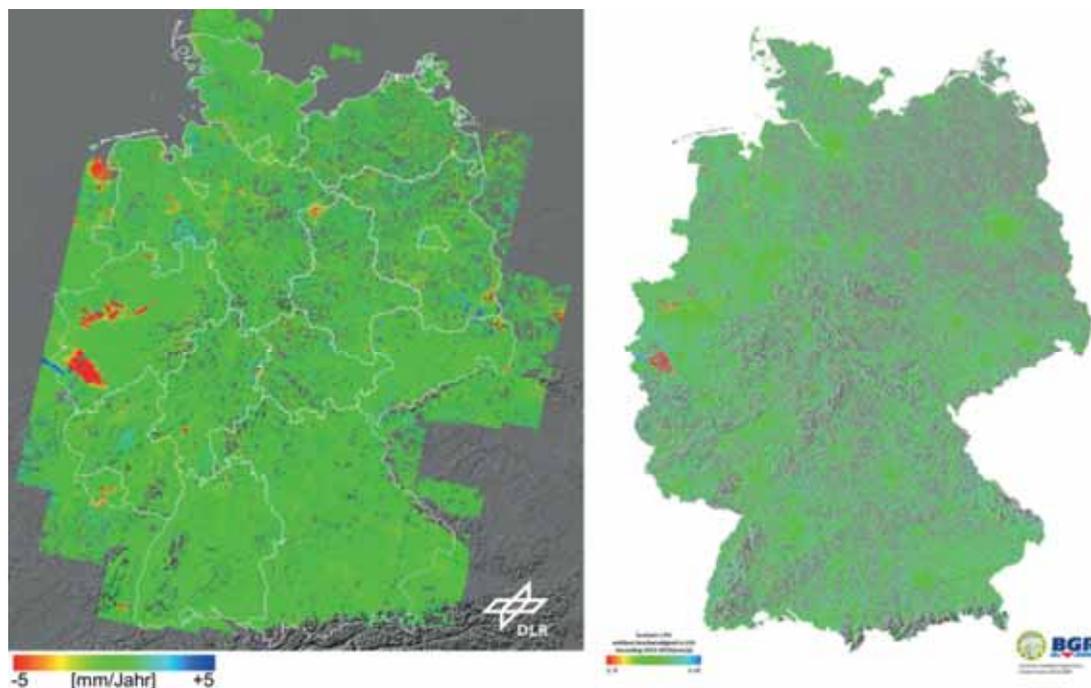


Abb. 5: links: PSI Prozessierung mit ERS Daten der 90-er Jahre. (ADAM et al. 2010); rechts: PSI Prozessierung mit Sentinel-1A/B Daten der Jahre 2014-2019 (KALIA, & FREI 2019)

Zukünftige Bodenbewegungsdienste sind national aber auch im europäischen Rahmen als Copernicus Service Element geplant (Frei 2017). Vertikale Bodenbewegungen sollen mit den Bodenbewegungskatastern der Vermessungsverwaltungen der Länder Nordrhein-Westfalen und Saarland zur Verfügung gestellt werden. In NRW wird das Bodenbewegungskataster NRW nach derzeitigem Stand 2021 online gehen (RIECKEN et al. 2019). Im Jahr 2020 werden vom saarländischen Landesamt für Geoinformation, Vermessung und Landentwicklung (LVGL) im Rahmen des „Saarländischen Bodenbewegungskataster“ (SaarBoBeKa) eigene PSI-Bodenbewegungsanalysen bereitgestellt. Hier ist vorgesehen, die Radarsatellitendaten aus den Ascending- und den entsprechenden Descending-Orbits zu kombinieren (YIN & BUSCH, 2019), was über Multisensor-Stationen (MSST) erfolgen wird (Abb. 6), die derzeit im Saarland errichtet werden (Abb. 7).

## 4 Corner Reflektoren

Corner Reflektoren (CR) erlauben die Validierung der Vermessung und den Abgleich der relativen interferometrischen Bewegungsänderungen mit zeitlich und örtlich ausreichend dichten Referenzdaten für Lage- und Höhenveränderungen (relative Lagerung). Letztendlich folgt die absolute Lagerung in einem Lage- und Höhenkoordinatensystem, was mit höchster Genauigkeit nur über Corner-Reflektoren mit einem Lage- und Höhenanschluss an das jeweilige Messnetz zu gewährleisten ist.

Das Bundesamt für Kartographie und Geodäsie (BKG) hat bereits einige GNSS-Standorte (Hof, Bautzen, Lindenberg, Bremgarten) mit Corner-Reflektoren erweitert, wobei je ein CR für den Ascending- und ein CR für den Descending-Orbit eingerichtet wurde (Abb. 8). Das BKG baut diese Stationen weiter aus, wobei je nach Standorteignung passive Corner-Reflektoren oder die Installation von aktiven elektronischen *Compact Active Transponder* (CAT) vorgesehen ist (FRIEDLÄNDER & LIEBSCH 2019).

Am Landesamt für Vermessung, Geoinformation und Landentwicklung (LVGL) des Saarlands und der RAG sind dafür im Rahmen des „Saarländischen Bodenbewegungskataster“ Doppel-Corner-Reflektoren (Doppel-CR) für die Multisensor-Stationen (MSST) entwickelt worden. Die MSST bestehen aus einem ca. 2 m × 2 m × 2 m mächtigen Stahlbetonfundament, welches einen massiven Doppel-CR von 2 m Durchmesser und 1 m Kantenlänge aus 10 mm Edelstahlplatten, eine permanent registrierende GNSS-Station mit kalibrierter 2-Frequenz Choking-Antenne zum Anschluss an das amtliche SAPOS-Netz und mehrere Nivellementpunkte zum Anschluss an die Höhenmesslinien der Länder birgt. Testmessungen konnten am RAG Standort Duhamel bei Saarlouis an einem Prototypen des Doppel-CR und einem benachbarten einfachen CR erfolgen. Der Doppel-CR wurde im April 2018 mobil auf einer Palette belassen, um Änderungen der Ausrichtung, Verkipnungen und eventuelle GNSS-Mehrwegeeffekte untersuchen zu können. Der einfache CR wurde im Dezember 2010 in descending Ausrichtung gesetzt. Die Untersuchungen zu Sentinel-1 und TerraSAR-X Satellitendaten und haben ergeben, dass der Doppel-CR sowohl in Sentinel-1 asc/desc und TerraSAR-X ascending gut erkennbar ist. Langzeit-statische GNSS-Messungen des LVGL über und neben dem Doppel-CR sowie an unterschiedlichen Standorten ergaben, dass es keine signifikanten Mehrwegeeffekte durch den Doppel-CR auf die GNSS-

Messung gibt. Somit sind die MSST zur Validierung und Referenzierung aller fernerkundlichen und terrestrischen Messdaten geeignet (SPRECKELS et al. 2020).

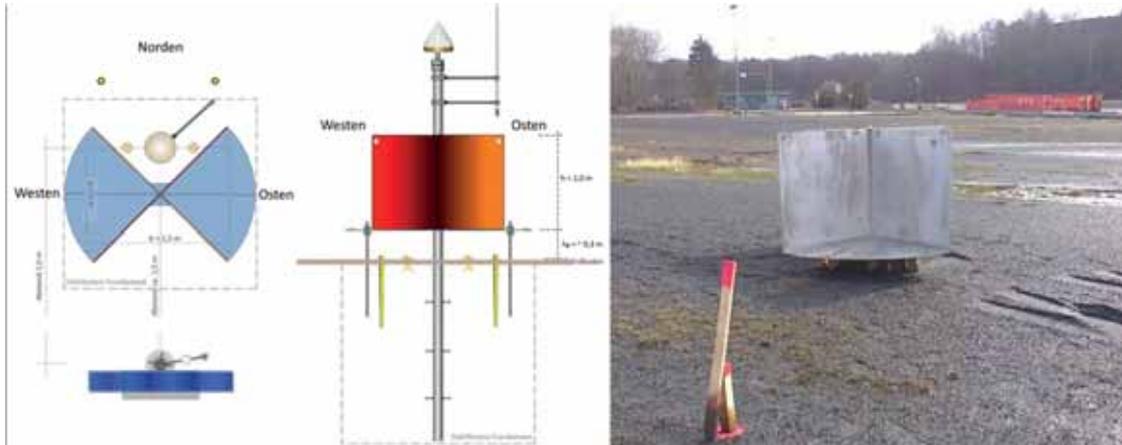


Abb. 6: Konstruktionsskizze einer RAG Multisensor-Referenzstation (MSST) aus Doppel-CR, GNSS-Monitoringstation, Nivellementanschlussspunkten, Datenübertragung, autarker Energieversorgung: Aufsicht (links), Seitenansicht von Süden (Mitte). RAG Doppel-CR für Testreihen am Standort Duhamel, Saarland (rechts). (Mit freundlicher Genehmigung der RAG Aktiengesellschaft)

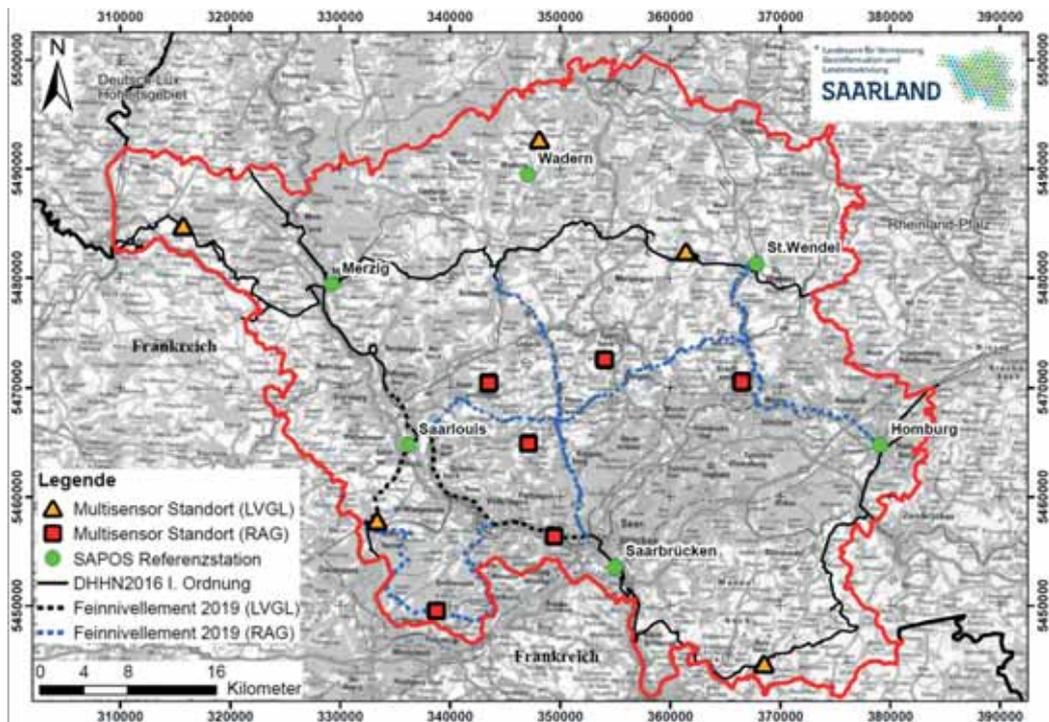


Abb. 7: Umsetzung des Monitoringkonzepts SaarBoBeKa im Saarland (Quelle: SPRECKELS et al. 2020)

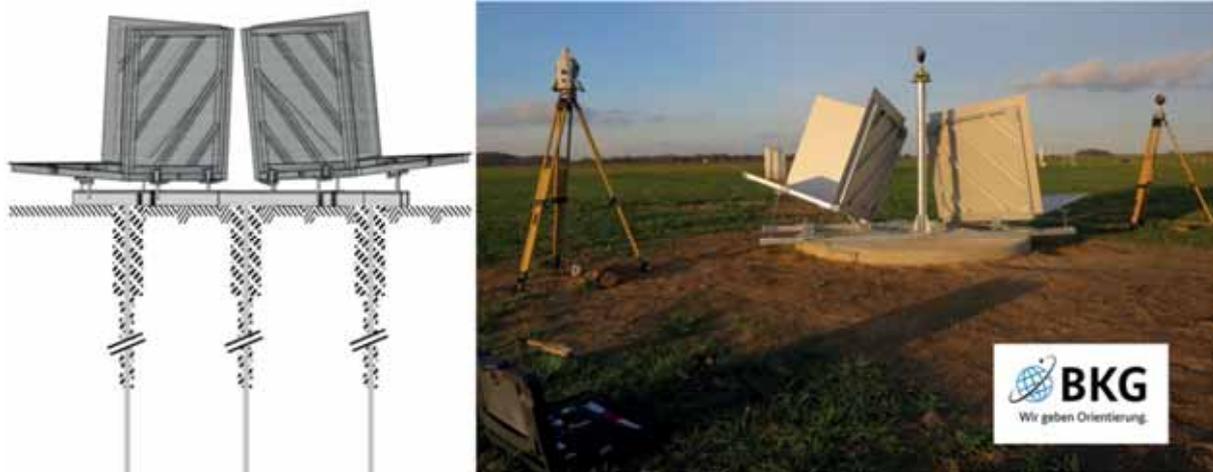


Abb. 8: Doppel-CR des BKG: Konstruktionsskizze (links), Einmessung (rechts) (Mit freundlicher Genehmigung vom BKG)

## 5 Fazit & Ausblick

Das Interesse an der Radarinterferometrie für die Bodenbewegungserfassung ist in den letzten Jahren stark gewachsen und erreicht inzwischen Nutzer ohne Radar- und Vermessungserfahrungen. Der Grund liegt in der kostengünstigen oder kostenfreien Verfügbarkeit kommerzieller und behördlicher Dienste. Das neue Normungsverfahren soll die Lücke zwischen den Experten d.h. den Service Providern und den Anwendern schließen. Daher ist es das Ziel, vereinheitlichte Begriffe, validierte Verfahren und Standards bei den Vermessungsprodukten aber auch bei der Qualitätssicherung zu etablieren. Diese DIN-Initiative stärkt und untermauert alle bisherigen Entwicklungen, stärkt das Vertrauen der Nutzer in diese neue Technik und macht die interferometrische Bodenbewegungserfassung rechtssicher und in der betrieblichen und behördlichen Praxis alltäglich.

Dieser Beitrag soll auch auf das neue Normungsverfahren „InSAR – Radarinterferometrie für die Bodenbewegungserfassung“ aufmerksam machen, um interessierte Kreise für Beiträge anzusprechen und neue Experten für diese Initiative einzuladen.

## 6 Literaturverzeichnis

- ADAM, N., KAMPES, B. M., EINEDER, M., WORAWATTANAMATEEKUL, J. & KIRCHER, M., 2003: The development of a scientific permanent scatterer system. ISPRS Workshop High Resolution Mapping from Space **6**, <https://www.ipi.uni-hannover.de/fileadmin/institut/pdf/adam.pdf>.
- ADAM, N. & PARIZZI, A., 2008: Terrafirma Quality Control Protocol for Level 1 Products”, Terrafirma Project Document, **1(5)**, 1-61.
- ADAM, N., PARIZZI, A., EINEDER, M. & CROSETTO, M., 2009: Practical persistent scatterer processing validation in the course of the Terrafirma project. *Journal of Applied Geophysics*, **69(1)**, 59-65.

- ADAM, N., PARIZZI, A. & KAMPES, B., 2010: TerraFirma Service Validation Protocol C5. ESA GMES TerraFirma Technical Note V.37, ESRIN/Contract no. 17059/03/I-IW, 1-139.
- ADAM, N., RODRIGUEZ GONZALEZ, F., PARIZZI, A. & LIEBHART, W., 2011: Wide Area Persistent Scatterer Interferometry: Algorithms and Examples. In Proceedings of Fringe, 1-5.
- BALTRUSCH, S. & REULKE, R., 2017: DIN-Normungsarbeit in der Photogrammetrie und Fernerkundung – Stand und Perspektiven. Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V., Band 26, T. Kersten (Hrsg.), 37. Wissenschaftlich-Technische Jahrestagung der DGPF, 7.-10. März 2017 in Würzburg, Tagungsband auf CD-ROM, 280-287.
- BERARDINO, P., FORNARO, G., LANARI, R. & SANSOSTI, E., 2002: A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. *IEEE Transactions on Geoscience and Remote Sensing* **40**(11), 2375-2383.
- BGR, 2019: WebGIS BodenBewegungsdienst Deutschland. online: <https://bodenbewegungsdienst.bgr.de/mapapps/resources/apps/bbd/index.html?lang=de>.
- BVEG, 2018: Höhenüberwachung durch Radarinterferometrie. online: <https://www.bveg.de/Erdgas/Erdgasspeicher/Radarinterferometrie>.
- COSTANTINI, M., FERRETTI, A., MINATI, F., FALCO, S., TRILLO, F., COLOMBO, D., NOVALI, F., MALVAROSA, F., MAMMONE, C., VECCHIOLI, F., RUCCI, A., FUMAGALLI, A., ALLIEVI, J., CIMINELLI, M. G. & COSTABILE, S., 2017: Analysis of surface deformations over the whole Italian territory by interferometric processing of ERS, Envisat and COSMO-SkyMed radar data. *Remote Sensing of Environment* **202**, 250-275, doi: 10.1016/j.rse.2017.07.017.
- CROSETTO, M., MONSERRAT, O., JUNGNER, A. & CRIPPA, B., 2009: Persistent Scatterer Interferometry: potential and limits. ISPRS Hannover Workshop 2009 High-Resolution Earth Imaging for Geospatial Information, ISPRS Archives – Volume XXXVIII-1-4-7/W5.
- EINEDER, M., BAMLER, R., CONG, X., GERNHARDT, S., FRITZ, T., ZHU, X., BALSS, U., BREIT, H., ADAM, N. & FLORICIOIU, D., 2013: Globale Kartierung und lokale Deformationsmessungen mit den Satelliten TerraSAR-X und TanDEM-X. *ZfV - Zeitschrift für Geodäsie, Geoinformation und Landmanagement* **1**, 75-84.
- FARR, T. G., ROSEN, P.A., CARO, E., CRIPPEN, R., DUREN, R., HENSLEY, S., KOBRIK, M., PALMER, M., RODRIGUEZ, E., ROTH, L., SEAL, D., SHAFFER, S., SHIMADA, J., UMLAND, J., WERNER, M., OSKIN, M., BURBANK, D. & ALSDORF, D., 2007: The Shuttle Radar Topography Mission. *Rev. Geophys.* **45**(RG2004), doi:10.1029/2005RG000183.
- FERRETTI, A., PRATI, C. & ROCCA, F., 1999: Permanent scatterers in SAR interferometry. *IEEE 1999 International Geoscience and Remote Sensing Symposium* **3**, 1528-1530.
- FERRETTI, A., PRATI, C. & ROCCA, F., 2000: Nonlinear subsidence rate estimation using permanent scatterers in differential SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing* **38**(5), 2202-2212, doi: 10.1109/36.868878.
- FERRETTI, A., PRATI, C. & ROCCA, F., 2001: Permanent scatterers in SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing* **39**(1), 8 - 20, doi: 10.1109/36.898661.
- FREI, M., 2017: European Ground Motion Service (EU-GMS) - A proposed Copernicus service element. online: <https://land.copernicus.eu/user-corner/technical-library/egms-white-paper>.

- FRIEDLÄNDER, S. & LIEBSCH, G., 2019: Monitoring geodätischer Infrastruktur – Das Potential der SAR-Interferometrie für das Vermessungswesen. Vortrag auf dem DIN-Normungsausschuss Photogrammetrie und Fernerkundung, Berlin, 08.10.2019.
- DEUTSCHER MARKSCHEIDER-VEREIN E.V., 2013: Grundsätze zum Einsatz von satellitengestützten Verfahren der Radarinterferometrie zur Erfassung von Höhenänderungen. online: [https://www.dmv-ev.de/images/stories/uploads/DMV\\_Radar\\_interferometrie\\_Grundsaeetze\\_2013\\_09\\_16.pdf](https://www.dmv-ev.de/images/stories/uploads/DMV_Radar_interferometrie_Grundsaeetze_2013_09_16.pdf).
- GEOSERVICE DLR, 2018: The TanDEM-X 90m Digital Elevation Model, online: <https://geoservice.dlr.de/web/dataguide/tdm90/>.
- GOLDSTEIN, R., 1995: Atmospheric limitations to repeat-track radar interferometry. *Geophysical Research Letters* **22**, 2517-2520.
- GVL, 2018a: United Kingdom Relative Deformation Map. online: <https://mangomap.com/geomatic-ventures-limited/maps/72883/united-kingdom-relative-deformation-map>.
- GVL, 2018b: Netherlands Relative Deformation Map. online: <https://mangomap.com/geomatic-ventures-limited/maps/80063/netherlands-relative-deformation-map-2016-2018#>.
- GVL, 2018c: Germany Relative Deformation Map. online: <https://mangomap.com/geomatic-ventures-limited/maps/76398/germany-relative-deformation-map>.
- HANSEN, R. F., VAN LEIJEN, F., KETELAAR, G., MARINKOVIC, P. S. & GEHLOT, S., 2005: PSIC4: PSI-processing over a validation test site. applicability report. TU Delft Technical Report, ESA Contract Nr. 18707/04/I-LG.
- HANSEN, R. F., VAN LEIJEN, F. J., VAN ZWIETEN, G. J., BREMMER, C., DORT-LAND, S. & KLEUSKENS, M., 2008: Validation of existing processing chains in TerraFirma stage 2; Product validation: validation in the Amsterdam and Alkmaar area, in GMES TerraFirma Report, ESRIN/contract No. 19366/05/I-E, 1-85.
- HUBER, M., WESSEL, B., HABERMEYER, M. & ROTH, A., 2007: Quality of Orthorectified TerraSAR-X Products. IEEE 2007 International Geoscience and Remote Sensing Symposium, 3937-3937 doi: 10.1109/IGARSS.2007.4423707.
- KALIA, A. C., FREI, M. & LEGE, T., 2017: A Copernicus downstream-service for the nationwide monitoring of surface displacements in Germany. *Remote Sensing of Environment* **202**, 234-249.
- KALIA, A. C. & FREI, M., 2019: BodenBewegungsdienst Deutschland – BBD. online: [https://www.bgr.bund.de/DE/Themen/GG\\_Fernerkundung/BodenBewegungsdienst\\_Deutschland/bodenbewegungsdienst\\_deutschland\\_node.html](https://www.bgr.bund.de/DE/Themen/GG_Fernerkundung/BodenBewegungsdienst_Deutschland/bodenbewegungsdienst_deutschland_node.html).
- KALIA, A., 2019: Nutzungshinweise BBD WebGIS. Version 01, online: [https://www.bgr.bund.de/DE/Themen/GG\\_Fernerkundung/Downloads/nutzungshinweise-bbd-webgis.html?nn=12782094](https://www.bgr.bund.de/DE/Themen/GG_Fernerkundung/Downloads/nutzungshinweise-bbd-webgis.html?nn=12782094)
- KAMPES, B. & ADAM, N., 2005: The STUN algorithm for persistent scatterer interferometry. Proceedings of FRINGE, **SP-610**, 1-14, [https://elib.dlr.de/110190/1/kampes\\_fringe05.pdf](https://elib.dlr.de/110190/1/kampes_fringe05.pdf)
- KRIEGER, G., MOREIRA, A., FIEDLER, H., HAJNSEK, I., WERNER, M., YOUNIS, M. & ZINK M., 2007: TanDEM-X: A satellite formation for high-resolution SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing* **45**(11), 3317-3341, doi: 10.1109/TGRS.2007.900693.

- LANARI, R., CASU, F., MANZO, M., ZENI, G., BERARDINO, P., MANUNTA, M. & PEPE, A., 2007: An Overview of the Small Baseline Subset Algorithm: A DInSAR Technique for Surface Deformation Analysis. *Pure and Applied Geophysics*, **164**, 637-661, doi: 10.1007/s00024-007-0192-9.
- MASSONNET, D., FEIGL, K., ROSSI, M. & ADRAGNA, F., 1994: Radar interferometric mapping of deformation in the year after the Landers earthquake. *Nature* **369**, 227-230.
- NETHERLANDS GEODETIC COMMISSION, 2018: The Netherlands Bodemdalingskaart. online: <https://bodemdalingskaart.nl/portal/index>.
- RABUS, B., EINEDER, M., ROTH, A. & BAMLER, R., 2003: The Shuttle Radar Topography Mission (SRTM) – A New Class of Digital Elevation Models Acquired by Spaceborne Radar. *ISPRS Journal of Photogrammetry & Remote Sensing* **57**(4), 241-262, doi: 10.1016/S0924-2716(02)00124-7.
- RAUCOULES, D., BOURGINE, B., DE MICHELE, M., LE COZANET, G., CLOSSET, L., BREMMER, C., VELDKAMP, H., TRAGHEIM D., BATESON, L., CROSETTO, M. & AGUDO, M., 2007: Persistent Scatterers Interferometry Independent Validation and Intercomparison of Results – Executive Summary. ESA Report BRGM/RP-55640-FR.
- REULKE, R., 2015: Standardisierung und Qualitätssicherung. Arbeitskreis der Deutsche Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V. (DGPF), online: <https://www.dgpf.de/aks/sqs.html>.
- RIECKEN, J., KRICKEL, B., GEFELLER, V. & REIFENRATH, P., 2019: Nutzung der Radarinterferometrie im geodätischen Raumbezug. In *Zeitschrift für Geodäsie, Geo-Information und Landmanagement (zfv)* **144**, 354-361, doi: 10.12902/zfv-0281-2019.
- SPRECKELS, V., SCHULZ, M., NIEMEIER, W., ENGEL, T., SCHÄFER, F., KEMKES, E., RÜFFER, J., TENGEN, D., BECHERT, S., SCHLIENKAMP, A., DROBNIOWSKI, M., MÜLLER, M. & SCHMITT P., 2020: GNSS, Nivellement und Radar – einheitliche Multisensor-Referenzstationen zur Überwachung von Bodenbewegungen in ehemaligen Bergbaubereichen. Tagungsband GeoMonitoring 2020.
- SOWTER, A., BATESON, L., STRANGE, P., AMBROSE, K. & SYAFIUDIN, M. F., 2013: DInSAR estimation of land motion using intermittent coherence with application to the South Derbyshire and Leicestershire coalfields. *Remote Sensing Letters*, **4**(10), 979-987.
- TERRAFIRMA, 2005-2015, Projektwebseite, <http://www.europe-geology.eu/terrafirma/#mapobj>.
- WERNER, M., 2001: Shuttle Radar Topography Mission (SRTM), Mission overview, *J. Telecomm.* **55**, 75-79.
- WERNER C., WEGMÜLLER, U., STROZZI, T. & WIESMANN, A., 2003: Interferometric point target analysis for deformation mapping. *Proc. IGARSS 2003*.
- YIN, X. & BUSCH, W., 2018: Nutzung der Sentinel-1 Aufnahmekonfigurationen zur Ableitung von Bodenbewegungskomponenten im Rahmen eines radarinterferometrischen Bodenbewegungsmonitorings. Tagungsband GeoMonitoring, 119-138.
- ZEBKER, H. A., ROSEN, P. A. & HENSLEY, S., 1997: Atmospheric effects in interferometric synthetic aperture radar surface deformation and topographic maps. *Journal of Geophysical Research: Solid Earth* **102**(B4), 7547-7563.

# Entwicklung einer interaktiven, immersiven Virtual-Reality-Applikation des historischen Stadtmodells Stade 1620

THOMAS P. KERSTEN<sup>1</sup> & ALEXANDER P. WALMSLEY<sup>1</sup>

*Zusammenfassung: Virtual Reality wird durch die immer leistungsfähigeren 3D-Dokumentations- und Modellierungstechnologien ein erschwingliches und interessantes Werkzeug für Visualisierungen in Architektur, Planung, Denkmalpflege, u.v.m. Daher wird es immer wichtiger, kostengünstige Methoden für die Erstellung von immersiven virtuellen 3D-Umgebungen zu entwickeln. Somit wird diese Technologie für Institutionen wie Museen und andere kulturelle Bildungseinrichtungen, die oft innerhalb eines engen Budgets arbeiten, eine attraktive Form der Wissensvermittlung. In diesem Beitrag beschreiben wir einen Workflow zum Aufbau einer interaktiven, immersiven Virtual-Reality-Applikation für ein virtuelles 3D-Stadtmodell der Stadt Stade im Jahr 1620. Dieses virtuelle Stadtmodell basiert auf einem physischen 3D-Modell der Stadt, das im Stader Rathaus ausgestellt ist. Der Workflow beginnt mit der Digitalisierung des Modells mittels digitaler Photogrammetrie, gefolgt von der anschließenden Low- und High-Polygon-Modellierung der einzelnen Architekturobjekte in Autodesk Maya, führt weiter zum Texture Mapping mit Substance Painter und endet schließlich mit der Visualisierung in der Unreal Engine 4. Das Ergebnis dieses Workflows ist eine detaillierte, historische 3D-Umgebung mit einem hohen Maß an geometrischer Realität, in der Interaktivität später einfach hinzugefügt werden kann. Darüber hinaus verfolgt der Workflow einen sehr iterativen Ansatz, der es ermöglicht, die Leistung der virtuellen Umgebung in der Game Engine in jeder Phase des Prozesses zu überwachen und schnell anzupassen. Um das Potenzial und die Attraktivität der virtuellen Umgebung als Werkzeug für Bildung und Kommunikation zu erhöhen, werden zukünftig interaktive Elemente, Animationen und einfache Lernspiele integriert.*

## 1 Einleitung

In den letzten Jahren hat sich Virtual Reality zu einer günstigen und leistungsfähigen Technologie entwickelt, wodurch sie in einer viel breiteren Öffentlichkeit für Visualisierungszwecke und Spiele eingesetzt wird. Diese Entwicklung eröffnet kulturellen Bildungseinrichtungen, wie beispielsweise Museen, spannende neue Möglichkeiten, um ein neues und vor allem junges Publikum für die Besichtigung ihrer Sammlungen und Archive zu gewinnen. Insbesondere in Verbindung mit digitalen 3D-Rekonstruktionsverfahren kann Virtual Reality dem Publikum ermöglichen, historische Orte im Maßstab 1:1 in einer räumlich immersiven, visuellen und auditiven Umgebung zu erleben. Darüber hinaus fördert die Nutzung von Interaktion innerhalb dieser virtuellen Welten ein gewisses Maß an sensorisch-motorischer Immersion (HOLOPAINEN & BJÖRK 2005) und ist damit ein sehr mächtiges Werkzeug zum spielerischen Lernen. Infolgedessen hat dies ein zunehmendes Interesse an der Anwendung spiel-typischer Elemente (Gamification) für den Bildungsauftrag im Bereich des kulturellen Erbes geweckt. Die neue Form der Bildung und Wissensvermittlung wird heute mit dem Begriff "Edutainment" bezeichnet.

---

<sup>1</sup> HafenCity Universität Hamburg, Labor für Photogrammetrie & Laserscanning, Überseeallee 16, D-20457 Hamburg, E-Mail: [Thomas.Kersten, Alexander.Walmsley]@hcu-hamburg.de

In diesem Beitrag zeigen wir, dass diese Technologien kostengünstig eingesetzt werden können, um eine interaktive, immersive Virtual-Reality-Umgebung zu erschaffen. Auf der Grundlage der digitalen 3D-Rekonstruktion des historischen Stadtmodells der Stadt Stade im Jahr 1620 wurde eine VR-Applikation entwickelt. Dieses Projekt wurde in Zusammenarbeit mit dem Museum Schwedenspeicher in Stade konzipiert, wo die VR-Applikation als pädagogisches Instrument zukünftig für Museumsbesucher und Schulklassen eingesetzt werden soll, um so spielerisch einen Einblick in die Geschichte Stades zu geben.

## 2 Verfügbare historische Stadtmodelle

Es gibt einen kleinen, aber wachsenden Bestand an digitalen historischen 3D-Stadtmodellen, der in den letzten Jahren durch die zunehmende Verfügbarkeit der notwendigen Soft- und Hardware entstanden ist. Physische und skalierte 3D-Gelände- und Stadtmodelle gewannen ab dem 16. Jahrhundert an Popularität, vor allem als Werkzeug für militärische Planungen (WARMOES 2018). Tatsächlich gibt es mehrere Beispiele für digitale 3D-Modelle, die von einem vorhandenen physischen 3D-Stadtmodell entwickelt wurden. Dazu gehören die Stadt Solothurn (Schweiz) von Nebiker et al. (2009), Duisburg von JEDRZEJAS & PRZYBILLA (2009), Prag (Tschechien) von SEDLACEK & ZARA (2009), Toul (Frankreich) von CHEVRIER et al. (2010), Hamburg in 4D von KELLER et al. (2011), Lüttich (Belgien) von PFEIFFER et al. (2013) und Genf (Schweiz) vom Auguste Magnin Relief Projekt, koordiniert von der Stadt Genf ([http://making-of.geneve1850.ch/3D/FR/#Geneve\\_1850-3D](http://making-of.geneve1850.ch/3D/FR/#Geneve_1850-3D)). Die in diesen Projekten entwickelten digitalen Modelle haben ein breites Anwendungsspektrum. Während einige zur Dokumentation und Präsentation von Museumssammlungen gebaut wurden (z.B. CHEVRIER et al. 2010), wurden bei anderen das Potenzial als Forschungsinstrument genutzt (z.B. PFEIFFER et al. 2013). Selbstverständlich schließen sich diese beiden Ziele nicht gegenseitig aus.

Die Nutzung von Virtual Reality zur Erforschung dieser Stadtmodelle in der First-Person-View (FPV) ist jedoch vergleichsweise seltener. Dies liegt vielleicht an der Herausforderung, Städte mit genügend Details zu modellieren, um innerhalb der virtuellen Umgebung überzeugend realistisch zu erscheinen, während gleichzeitig technische Kriterien streng eingehalten werden müssen, die die zu ladende Datenmenge für eine einwandfreie Visualisierung begrenzen. Das Potenzial von Virtual-Reality-Stadtmodellen, die im Maßstab 1:1 erlebt und erforscht werden können, ist dennoch groß, nicht nur für die Präsentation und Planung, sondern auch für die Bildung und die Wissensvermittlung (THOMPSON et al. 2006). Virtual Reality-Stadtmodelle bieten die Möglichkeit, extrem komplexe Informationen für Menschen mit unterschiedlichem Fachwissen durch Immersion und Interaktivität zu vereinfachen.

Insbesondere historische Architektur- und Kulturobjekte wie z.B. Monumente oder Städte sind ideal für eine interaktive Visualisierung in der virtuellen Realität, da sie heute oft als Wahrzeichen oder als touristische Attraktion in Ausstellungen thematisiert werden. Aus diesem Grund haben solche Websites bereits ein großes, interessiertes Publikum, das durch VR-Erfahrungen angesprochen werden kann. Mehrere Projektbeispiele von der HafenCity Universität Hamburg über VR-Visualisierungen von historischen Geodaten wurden bereits veröffentlicht. Das Alt-Segeberger Bürgerhaus wurde als eines der ersten virtuellen Museen mit dem VR-System HTC Vive vorgestellt (KERSTEN et al. 2017b). Darüber hinaus wurden zwei historische Städte (einschließlich der

umliegenden Landschaften) als VR-Anwendungen für Besuche in der Vergangenheit entwickelt: Duisburg im Jahr 1566 (Tschirschwitz et al. 2019) und Segeberg im Jahr 1600 (Kersten et al. 2018a). Zwei religiöse Kulturdenkmäler stehen auch als VR-Erlebnisse zur Verfügung: die Selimiye Moschee in Edirne, Türkei (Kersten et al. 2017a) und das Holzmodell des Salomonischen Tempels (Kersten et al. 2018b).

Ein weiteres historisches Beispiel in VR stellt der sanierte post-industrielle Standort der "Zeche Holland" in Bochum-Wattenscheid dar (Edler et al. 2019). Einen etwas anderen Workflow zeigt Edler et al. (2018) für den Aufbau einer interaktiven kartographischen VR-Umgebung zur Erforschung urbaner Landschaften.



Abb. 1: Das historische 3D-Stadtmodell von Stade im Jahr 1620 (links) und ein Ausschnitt des physikalischen Modells im Rathaus

### 3 Das historische Stadtmodell von Stade

Die Hansestadt Stade liegt im Bundesland Niedersachsen 45 km westlich von Hamburg. Die durch den Nebenfluss Schwinge mit der Elbe verbundene Stadt Stade war einst ein mächtiges Mitglied der Hanse, die vom 12. bis 16. Jahrhundert den Handel in Mittel- und Nordeuropa dominierte. Die Stadt, die 1601 aus dem Städtebund der Hanse ausgeschlossen wurde, befand sich ab 1645 unter schwedischer und ab 1712 unter dänischer Herrschaft. Das Museum Schwedenspeicher in Stade verfügt über zahlreiche Stadtpläne und Karten aus der Zeit, die einen Eindruck von der Ausdehnung der Stadt und ihrer Umgebung vermitteln. Im Foyer des Rathauses befindet sich ein physisches 3D-Modell der Stadt, basierend auf ihrem Grundriss aus dem Jahr 1620, das Auskunft über die zahlreichen Gebäude und Kirchen in Stade gibt. Viele davon existierten bis zum großen Brand von 1659 und zeugen vom reichen religiösen und kulturellen Leben der Stadt in der Frühen Neuzeit. In ihrer Blütezeit hatte die Stadt fünf Stadtteile mit jeweils einer eigenen Kirche: das Klosterviertel mit der Kirche St. Georg, das Kaufmannsviertel mit St. Cosmas et Damiani, das Bischofsviertel mit St. Wilhadi, das Fischerviertel mit St. Nicolai und das Schlossviertel mit St. Pankratii. Das physische 3D-Modell der Stadt (Abb. 1) hat Abmessungen von 2,1 m × 1,3 m und ist im Maßstab ca. 1:550 dargestellt.

Leider ist wenig über die Geschichte des physischen Modells und dessen Entstehung selbst bekannt. Es wurde 1977 für die Eröffnung des Schwedenspeichers in Auftrag gegeben und 2010, als das Museum im Umbau begriffen war, in das Foyer des Rathauses verlegt. Der Name des Künstlers

oder auch die genauen historischen Quellen, auf denen die Rekonstruktion basiert, sind nicht bekannt. Da es jedoch ursprünglich vom Museum in Auftrag gegeben wurde, kann man davon ausgehen, dass das Modell als interpretatives Hilfsmittel für Besucher gedacht war.

#### 4 Die Unreal Game Engine und das VR-System HTC Vive Pro

Eine Game Engine ist eine Software-Entwicklungsumgebung, die entwickelt wurde, um Videospiele für Konsolen, mobile Geräte und PCs zu erstellen. Seine wichtigste Funktionalität umfasst typischerweise eine Rendering-Engine für 2D- oder 3D-Grafiken zur Visualisierung texturierter 3D-Modelle. Zusätzlich enthält sie unter anderem eine Physik-Engine oder eine Kollisionserkennung zur Interaktion mit Objekten, ein Audiosystem zur Ausgabe von Sound, Skripting, Werkzeuge für die Erstellung von Animationen, künstliche Intelligenz, Streaming und Performance Management. Eine Game Engine ist eine leistungsstarke Plattform zur Darstellung von 3D-Landschaftsdaten in VR, da sie dem Benutzer mit den notwendigen Funktionalitäten das Eintauchen in eine virtuelle Welt ermöglicht.

Derzeit bieten mehrere Spiele-Engines wie die Unity-Engine von Unity Technologies, die CryEngine des deutschen Entwicklungsstudios Crytek und die Unreal Engine 4 (UE4) von Epic Games für Entwickler freien Zugang. Einen aktuellen Überblick und Vergleich verschiedener Game Engines findet sich in GAMESPARKS (2019). Die Auswahl einer geeigneten Engine für die Projektbearbeitung richtet sich nach den oben genannten Komponenten und der Anpassungsfähigkeit an einen bereits bestehenden Arbeitsprozess. Für dieses Projekt wurde UE4 gewählt, weil sie - im Gegensatz zu anderen Engines - eine visuelle Skriptsprache mit Blueprints implementiert hat. Diese visuelle Programmierung erfordert kein Schreiben von maschinenkonformem Quellcode und ermöglicht Anwendern mit geringen Programmierkenntnissen einen viel einfacheren Einstieg in die Entwicklungsumgebung. Darüber hinaus bietet Unreal ausreichenden Zugriff auf Dokumentationen, Tutorials und Asset Stores. Aufgrund der begrenzten Zeit für dieses Projekt wurde keine andere Engine getestet.

Als Virtual Reality System kommt die HTC Vive Pro zum Einsatz (Abb. 2 links). Die HTC Vive Pro ([www.vive.com](http://www.vive.com)) ist ein Virtual-Reality-Headset für raumbezogene VR. Es wurde von HTC und Valve Corporation entwickelt. Die erste Version der HTC Vive wurde am 5. April 2016 auf den Markt gebracht. Basiskomponenten sind das Headset für das immersive Erlebnis, zwei Controller für Benutzerinteraktionen und zwei Lighthouse-Basisstationen zur Verfolgung der Benutzerbewegung durch Infrarotlaser. Das komplette HTC Vive Pro (Full Kit) ist derzeit für EUR 1199 (Januar 2020) auf dem Markt erhältlich. Die Lighthouse-Technologie erlaubt eine hochpräzise und zeitlich hochaufgelöste Positionsbestimmung des Nutzers innerhalb der  $4,6 \text{ m} \times 4,6 \text{ m}$  großen Interaktionsfläche (Abb. 2 rechts). Das VR-System verwendet einen Gyroskop-, einen Beschleunigungs- und Laserinfrarotsensor, um die Bewegungen des Kopfes bis auf ein Zehntel Grad genau zu verfolgen (PAINTER, 2015). Die zwei im Raum diagonal positionierten Lighthouse-Stationen senden Laserlinien mit einer Wellenlänge im Infrarotbereich aus, die von mehreren einfachen Fotosensoren in der VR-Brille und in den Controllern für deren 3D-Positionsbestimmung genutzt werden. Die Vive-Controller sind speziell für VR mit intuitiver Steuerung und realistischem haptischem Feedback konzipiert, um virtuelle Umgebungen frei zu erkunden und mit ihnen zu interagieren. Die visuelle Darstellung erfolgt in der VR-Brille mit einer Auflösung von  $1440 \times 1600$

Pixel und einer Bildwiederholrate von 90 Hz. Der Datenaustausch zwischen Game Engine und der HTC Vive wird über SteamVR realisiert. Abb. 2 zeigt den Aufbau des VR-Systems HTC Vive Pro einschließlich des Interaktionsbereichs (blau).

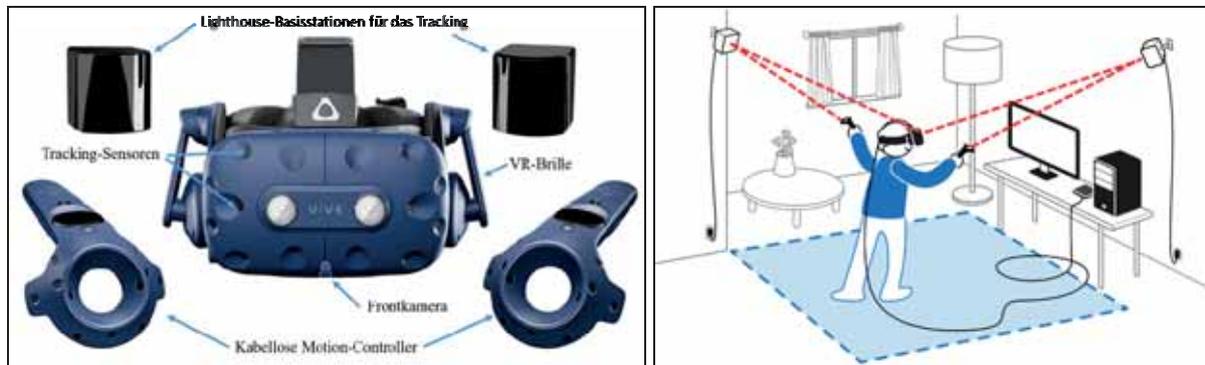


Abb. 2: HTC Vive Pro mit zwei Controllern und dem Lighthouse-Tracking-System ([www.vive.com](http://www.vive.com)) (links) und Aufbau der HTC Vive Pro einschließlich der Lighthouse-Basisstationen (rechts), die die beweglichen zwei Controller und das Headset im blauen Interaktionsbereich in Echtzeit verfolgen ([www.htc.com](http://www.htc.com))

## 5 Der Workflow

Der folgende Workflow wurde entwickelt, um das virtuelle Stadtmodell zu erstellen: (1) Datenerfassung durch digitale Photogrammetrie, (2) Bildorientierung, dichte Punktwolkengenerierung (Abb. 3 links) und Dreiecksvermaschung (Abb. 3 rechts) mit der Software Agisoft PhotoScan/Metashape, (3) Erstellung einer Höhenkarte (Height Map) aus der photogrammetrischen Dreiecksvermaschung in Autodesk Maya (Abb. 4 links oben), (4) Laden der Höhenkarte als Grauwertbild in UE4 und Erstellen von einfachen Klötzchenmodellen der Objekte mit wenigen Polygonen (Abb. 4), (5) detaillierte 3D-Modellierung der Gebäude und Objekte mit Maya unter Verwendung der photogrammetrischen Dreiecksvermaschung (Mesh) und Referenzmaterial, (6) Texturierung der 3D-Modelle mit Maya und Substance Painter, (7) Einsetzen der texturierten, detaillierten 3D-Modelle in die Szene in UE4, (8) Modellierung und Texturierung der Landschaften mit den UE4-Landscape-Editor, (9) Integration von Bewegungssteuerungen und Interaktionen in UE4, und (10) immersive und interaktive Visualisierung des Stadtmodells im VR-System HTC Vive Pro mit SteamVR - als Schnittstelle zwischen der Game Engine und dem Head Mounted Display.

## 6 Datenerfassung

Im Jahr 2011 digitalisierte das Labor für Photogrammetrie & Laserscanning der HafenCity Universität Hamburg das physikalische 3D-Stadtmodell von Stade mittels Photogrammetrie. 106 Fotos wurden mit einer Spiegelreflexkamera Nikon D90 mit Nikkor-Objektiv 20 mm und Zoom-Objektiv (Brennweite 85 mm) aufgenommen. Alle Fotos wurden in Agisoft PhotoScan verarbeitet, um eine dichte Punktwolke von 20,7 Millionen Punkten mit hoher Qualität zu erzeugen (Abb. 3). Für die Skalierung der Punktwolke wurden die Distanzen von drei Maßstäben gemessen und in

PhotoScan mit einer Genauigkeit von 0,3 mm ermittelt. Schließlich wurde ein texturiertes 3D-Modell mit 4,1 Millionen Dreiecksflächen als Grundlage für die weitere 3D-Modellierung erstellt.



Abb. 3: Dichte 3D-Punktwolke des physikalischen Stadtmodells von Stade im Jahr 1620 (links) und perspektivische Ansicht des texturierten, vermaschten 3D-Modells (rechts)

## 7 3D-Modellierung, Texturierung und VR-Integration

Im Rahmen dieses Projektes wurde der Workflow für die Modellierung und Texturierung optimiert, um die Bearbeitungszeit zwischen verschiedenen Versionen des virtuellen Modells zu minimieren und um die genutzte Computerleistung bis zur endgültigen VR-Applikation laufend zu überprüfen.

Die Datenmenge der Dreiecksvermaschung (Abb. 3 rechts) war zu groß, um es direkt in die Game Engine zu integrieren. Aus diesem Grund wurde mit Autodesk Maya eine Höhenkarte aus der Dreiecksvermaschung erstellt, die anschließend als Verschiebungskarte in UE4 eingebracht wurde, um ein hochauflösendes (4K) Landschaftsnetz zu erzeugen. Mit den Modellierungswerkzeugen von UE4 wurden von den Gebäuden einfache Klötzchenmodelle (mit jeweils wenigen Polygonen bzw. Flächen) erstellt und an den entsprechenden Stellen in der Landschaft platziert (Abb. 4). Um die Zeit für die Modellierung zu verkürzen und gleichzeitig eine ausreichende Variabilität in den Gebäudetypen zu erreichen, wurden insgesamt 13 verschiedene Gebäudevarianten mit geringer Anzahl von Polygonen geschaffen, um alle 974 Stadthäuser von Stade darzustellen. Diese Varianten mit wenigen Polygonen wurden entwickelt, um die Variation der Haustypen im physischen Modell so genau wie möglich darzustellen, die sich je nach Variablen wie Größe und Form der Grundfläche, Gebäudehöhe, Anzahl der Stockwerke und Dachtyp unterscheiden. Neben den Häusern und anderen Wohnbauten wurden weitere Objekte mit wenigen Polygonen für die Verteidigungs- und Hafeninfrastrukturen der Stadt wie Stadtmauern, Brücken, Verteidigungstürme und Hafenkräne modelliert. Darüber hinaus wurde eine Reihe von Sonderbauten in der Stadt für eine individuelle Modellierung ausgewählt. Dazu gehören die Kirchen St. Pankratii, St. Nicolai, Cosmae et Damiani und St. Wilhadi, die Kloster St. Marien, das Kloster St. Johannis, das Rathaus Stade, ein historisches Kaufmannshaus und die Ruinen des Klosters St. Georg, die sich 1620 noch in der Stadtlandschaft befanden.



Abb. 4: Ansichten vom ersten Bearbeitungsschritt in der Unreal Engine: Platzierung von einfachen Gebäudeklötzchen auf die Höhenkarte, die aus der Dreiecksvermaschung der Punktwolke generiert wurde (links oben). Die verschiedenfarbigen Klötzchen stellen jeweils unterschiedliche Varianten des Stadthauses dar

Anschließend wurden diese einfachen Gebäudeklötzchen nach Autodesk Maya exportiert, um sie als Vorlagen für die Erstellung von detaillierten Architekturmodellen zu verwenden. Um die architektonischen Details für diese hochaufgelösten Modelle nachzubilden, war es notwendig, eine Menge an zusätzlichem Referenzdatenmaterial (Fotos, Abbildungen, Texte, usw.) zu sammeln. Die Details des physischen Stadtmodells selbst reichten nicht aus, um ein überzeugendes und realistisches, virtuelles Modell im Maßstab 1:1 zu erstellen. Darüber bot die Digitalisierung des Modells eine wertvolle Gelegenheit, das Modell durch zusätzliche historische Details zu ergänzen. Neben Fotos von der Stadtarchitektur, die in Stade und Lübeck aufgenommen wurden, wurden Referenzmaterial von Malern des niederländischen Goldenen Zeitalters (z.B. Jan van der Heyden, Cornelis Springer), von Online-Datenbanken (z.B. <http://www.fotografie-architektur.de/index.html>) sowie von weiteren digitalen 3D-Rekonstruktionen norddeutscher und niederländischer Architektur gesammelt. Archivmaterial, darunter alte Stadtpläne und Ansichten aus der Stadt des 16. und 17. Jahrhunderts, lieferte das Museum Schwedenspeicher (Abb. 5). Die Museumsausstellung selbst liefert zudem wertvolle Zusatzinformationen zu den architektonischen Stilen, die im 17. Jahrhundert in Stade präsent waren.

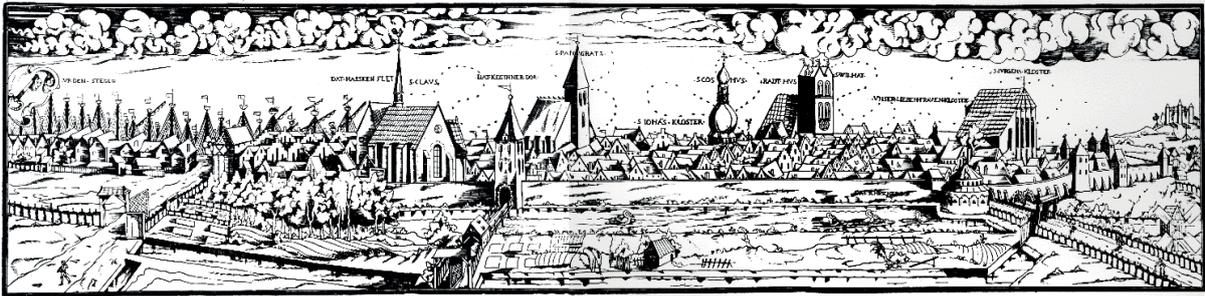


Abb. 5: Blick auf die Stadt Stade im Jahr 1550, nach einem Holzschnitt von Martin Weigel (Wirtgen & Bohmbach 1974)

Ein modularer Workflow wurde zur Modellierung der Stadthäuser und Verteidigungsstrukturen der Stadt (einschließlich der Mauern und der Verteidigungstürme) umgesetzt. Dieser Workflow beinhaltet die Zerlegung von Architekturobjekten in kleine, wiederholbare Elemente, die dann auf verschiedene Weise neu kombiniert werden können, um so mehrere Varianten von 3D-Modellen als Prototypen schnell zu erstellen. Nach der Modellierung der einzelnen Elemente können verschiedene Kombinationen für komplette Gebäude schnell getestet und fertig gestellt werden. Die Modellierung der oben aufgeführten Sondergebäude erfolgte individuell mit begrenzter Verwendung von verschiedenen Objektelementen, um die charakteristischen Besonderheiten dieser Gebäude herauszustellen. Nach und nach wurde jedes der detaillierten 3D-Modelle wieder in UE4 importiert, wobei die Modelle mit geringer Auflösung an den entsprechenden Stellen im Stadtmodell durch diese ersetzt wurden. Bei jedem Schritt der detaillierten Modellierung wurde darauf geachtet, dass die VR-Szene weiterhin die in der UE4-Dokumentation (<https://docs.unrealengine.com/en-US/Platforms/VR/DevelopVR/Profiling/Overview/index.html>) beschriebenen Performance-Benchmarks in Form von Frames per Second erfüllt.

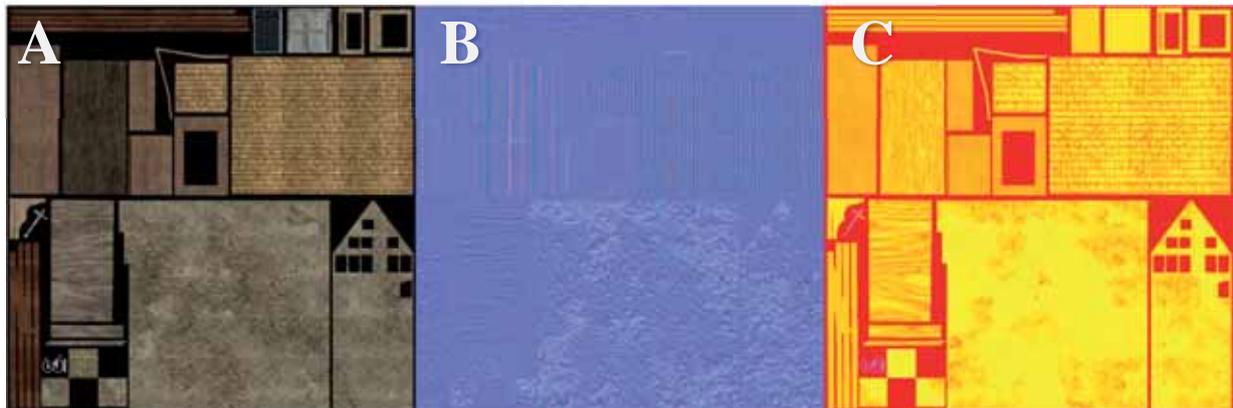


Abb. 6: Texturatlanen für ein Stadthaus: (A) Diffuse Map mit den Farbinformationen der Texturen; (B) Normal Map für den Tiefeneffekt; (C) Ambient Occlusion Map für die Umgebungsverdeckung, Metallic Map für glänzende metallische Objekte und Roughness Map für die Steuerung der Schärfe der Reflexionen

Nachdem die detaillierten 3D-Modelle in die Szene integriert worden waren, wurden die Modellflächen in Maya auf eine 2D-Bildfläche projiziert und in Substance Painter, einer Texturierungssoftware für 3D-Objekte, strukturiert. Substance Painter ermöglicht die Kombination von Bildtexturen und prozeduralen Designalgorithmen, um komplexere Texturen zu erstellen. Alle Texturen wurden für jedes Objekt in 4K-Texturatlanen zusammengefasst, um die Effizienz für die Renderzeit innerhalb der UE4 zu erhöhen. In den Texturatlanen sind folgende Informationen enthalten: Diffuse Map mit den Farbinformationen der Texturen (Abb. 6a), Normal Map für den Tiefeneffekt (Abb. 6b), Ambient Occlusion Map für die Umgebungsverdeckung, Roughness Map für die Steuerung der Schärfe der Reflexionen und Metallic Map für glänzende metallische Objekte (hier alle in einem Texturatlas zusammengefasst – Abb. 6c). Jedes Architekturmodell in der Szene erhielt ein Material (in UE4 wird ein Material verwendet, um sich auf die Kombination von 2D-Texturen und die Abfolge von Operationen zu beziehen, die der Game-Engine sagen, wie die Texturen dargestellt werden sollen), einschließlich der drei oben gezeigten Texturatlanen (Abb. 6). In Bereichen, in denen wiederholte Texturen offensichtlich vorlagen, wurden mehrere Texturen übereinandergelegt (Vertex Painting), um eine zusätzliche Variationsebene hinzuzufügen und um die Umgebung realistisch darzustellen. Für jedes Mesh wurden vier Level of Detail (LOD)-Modelle mit den entsprechenden UE4-Funktionen generiert, um die Computerperformance (Prozessorleistung) anzupassen, indem die Anzahl der Dreiecke in den Modellen reduziert wurde, die sich in einem bestimmten Bereich weiter entfernt von der Kamera befinden. Abb. 7 zeigt vier verschiedene Ansichten der Stadt Stade in der VR-Applikation, die mit dem Unreal-Editor erstellt wurden und die die Qualität der Modellierung und Texturierung der verschiedenen Stadtlandschaften zeigen.

Die ursprüngliche Landschaft in der 4K-Auflösung, die für die Platzierung der 3D-Modelle mit wenig Polygonen in der Szene verwendet wurde, wurde durch eine Landschaft mit geringerer 2K-Auflösung als endgültige Szene ersetzt, um die Geschwindigkeit für das Rendering zu erhöhen. Diese Landschaft wurde manuell mit dem UE4-Landscape-Editor bemalt bzw. texturiert, um so die Straßen, Wege und freien Vegetationsflächen zu gestalten. Zusätzlich wurden Pflanzen, darunter Laubbäume und Wiesengräser, ebenfalls manuell mit dem UE4-Editor Laubwerk in die Landschaft integriert. Eine zweite, niedrigpolare Landschaft wurde mit SRTM 1-Bogen-Sekunden-DEM-Daten erstellt, um der Stadt als niedrigpolare Kulisse zu dienen.

Um die Effizienz für das Rendering zu verbessern, wurde die Beleuchtung innerhalb der Engine in Vorberechnungen durchgeführt („Light Baking“) und sogenannte Lightmaps erstellt. Informationen über Licht- und Schattenverteilung auf einer Textur können somit schnell abgerufen werden und müssen nicht in Echtzeit berechnet werden, was die Rechenauslastung der Engine durch das Echtzeit-Rendering reduziert.

Schließlich wurde der Navigationsbereich in der virtuellen Umgebung erstellt, in dem sich der Anwender bewegen darf. Dazu wurden die grundlegende Bewegungssteuerung und Interaktionen über die knotenbasierte Programmierschnittstelle in UE4 durch Blueprints programmiert. Durch eine einfache Point-and-Click-Funktion am HTC Vive-Controller ist der Besucher dann in der Lage, sich durch Teleportation in der virtuellen Stadt selbstbestimmt fortzubewegen (Abb. 8).



Abb. 7: Ansichten des aktuellen Zustands der virtuellen Umgebung: (A) Blick in die Hökerstraße mit Rathaus auf der linken Seite; (B) Stadtsilhouette mit Platz am Sande im Vordergrund; (C) Stader Hafen mit Ladekran; (D) Stadtsilhouette mit Schiffen (Blickrichtung Südosten)

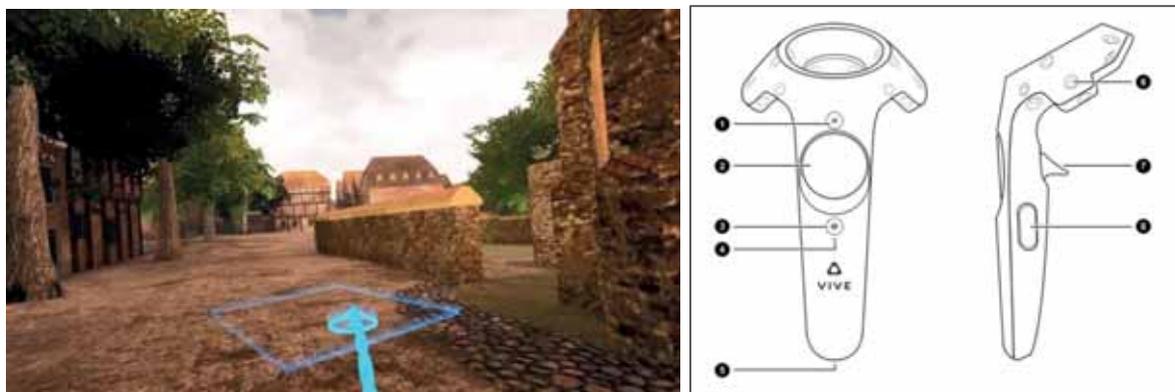


Abb. 8: Teleportation in der VR-Applikation durch Point-and-Click-Funktion am Controller (links) und programmierbare Tasten an den Controllern (rechts)

Nachdem die Grundlagen für die Visualisierung und Interaktionen gelegt wurden, bestand der nächste Schritt in diesem Projekt darin, Interaktionen zu entwickeln, mit denen Wissen spielerisch (Serious Games) und durch Animationen vermittelt werden kann, um in einer immersiven und interaktiven Welt Bildungsaspekte in das virtuelle Stadtmodell einzubringen. Ein Aspekt, der derzeit in der Erprobung ist, ist die Integration von 360°-Panorama-Fotos der heutigen Stadt Stade in die VR-Applikation, die in der virtuellen Realität durch Betreten einer Kreismarkierung an bestimmten Orten der Stadt und Klicken einer dafür vorgesehenen Taste am Controller zu betrachten

sind. Dieses Nebeneinander von historischer und heutiger Stadt lädt den Nutzer ein, die unterschiedlichen Stadtansichten direkt zu vergleichen und so die historischen Szenarien, die die Stadt prägten, zu erforschen.

Um diese Funktion zu testen, wurden am 9. September 2019 an verschiedenen Orten in der Stadt insgesamt acht 360°-Panoramafotos mit einer Kamera Samsung Gear 360 aufgenommen (Abb. 9 links). Um die Aufnahmeposition zu bestimmen, wurde die Gear 360 mit einem Smartphone verbunden, um so die GPS-Koordinaten als Geo-Tag zu erhalten. Damit wurde die spätere Platzierung der Panoramen in der virtuellen Umgebung erleichtert. Aus den Fotos wurde mit der Samsung Gear App für iOS je Standpunkt ein Panorama erzeugt und in UE4 importiert. Diese Panoramen wurden dann als Texturen auf die Innenseite von acht Hohlkugeln projiziert und an entsprechender Stelle in der Umgebung platziert. An jeder dieser Positionen wurde auch eine 3D-Animation (in diesem Fall ein animierter Kreis am Boden) platziert, der dazu dient, den Benutzer auf das Gebiet aufmerksam zu machen und ein verfügbares Panorama anzuzeigen (Abb. 9 rechts).



Abb. 9: 360°-Panoramafoto im Stader Hafen (links oben) und vom Stader Rathaus (links unten) mit der jeweiligen kreisförmigen 3D-Animation (rechts), als Information für den VR-Besucher über ein verfügbares Panoramafoto vom selben Ort zum Vergleich mit der heutigen Zeit

Die interaktiven Elemente dieser Funktion wurden mit Hilfe der visuellen UE4-Programmierschnittstelle als Blueprint implementiert. Sobald Benutzer den inneren Bereich des virtuellen Kreises betreten und anschließend die Trigger-Taste am Controller drücken (Nr. 7 in Abb. 8 rechts), wird auf das heutige Panorama umgeschaltet. Der Betrachter befindet sich dann in der 360°-Panoramakugel, die durch erneutes Drücken der Taste verlassen werden kann und anschließend wieder ausgeblendet wird. Zukünftig werden diese Panoramafotos, die nur zu Testzwecken aufgenommen und implementiert wurden, durch Panoramas mit höherer Auflösung ersetzt und ergänzt.

Darüber hinaus ist eine Reihe von Erweiterungen geplant, deren Umsetzung bereits in Zusammenarbeit mit der Stadt Stade begonnen hat:

- Integration von geschichtlich-korrekten Schiffen im Hafen und vor der Stadt
- Entwicklung eines in der VR-Applikation integrierten Tutorials zur Einweisung in die VR-Anwendung
- Integration einer Übersichtskarte von der Stadt (Abb. 10) mit der aktuellen Positionsanzeige des Anwenders zur besseren Orientierung innerhalb des Modells
- Integration von (aktuellen und/oder historischen) Straßennamen zur besseren der Orientierung
- Integration von Informationstafeln mit Text und Fotos für entsprechende Sehenswürdigkeiten bzw. Gebäude/Objekte
- Integration von Personen und Tieren (z.B. Pferde, Kühe, Schweine, Hühner, etc.) ggf. mit Audio sowie weitere Objekte (z.B. bewegliche Kisten, Fässer, Kutschen, etc.)
- Integration von Videos und/oder Hörtexten (MP3)
- Virtuell geführte Touren zu bestimmten Sehenswürdigkeiten des Stadtmodells als Stadtrundgang oder -führung
- Entwicklung eines Serious Games zur Entdeckung und Erkundung der Stadt in der Frühen Neuzeit (z.B. Handwerkerberufe, geschichtliche Aspekte zum 30-jährigen Krieg, etc.)



Abb. 10: Übersichtskarte der Stadt Stade im Jahr 1620 (als zukünftige Navigationshilfe für den Anwender)

## 8 Fazit & Ausblick

Die entwickelte VR-Applikation Stade 1620 wird im Jahr 2020 zum 400-jährigen Jubiläum der Stadtansicht mit der HTC Vive Pro als Head Mounted Display im Museum Schwedenspeicher in Stade integriert und auf der Plattform SteamVR für Präsentationen im Museum laufen. Mit diesem Projekt konnte ein Workflow entwickelt werden, der eine detailgetreue Visualisierung virtueller Stadtmodelle ermöglicht. Der hohe Detaillierungsgrad ist einerseits entscheidend für die Immersion sowie für eine hochauflösende Visualisierung und realistische Wiedergabe der virtuellen Umgebung, aber andererseits beansprucht die Modellierungs- und Texturierungsphase einen hohen zeitlichen Arbeitsaufwand. Für die erste Phase des Projektes bis zur Erstellung der ersten Version der VR-Applikation sind rund 450 Arbeitsstunden eingeflossen, von denen 275 Stunden für die detaillierte 3D-Modellierung und Texturierung verwendet wurden. Ein besonderer Fokus lag auf der manuellen Modellierung der individuellen Architektur-Objekte, um die Vielfalt in der räumlichen Stadtstruktur, die 1620 in der historischen Stadt Stade zu finden waren, besser wiederzugeben.

Darüber hinaus wurde wegen des hohen Detaillierungsgrades in der virtuellen Umgebung laufend die Prozessorleistung mit der eingesetzten Hard- und Software überprüft. Epic Games empfiehlt für UE4 eine Bildfolge rate von 90 Frames per Second (FPS) für das Rendering in Echtzeit für die HTC Vive Pro und SteamVR, um dem Benutzer beim Tragen der VR-Brille keine Unannehmlichkeiten (z.B. Motion Sickness) durch Latenzen zu bereiten. Eine hohe Bildrate für VR-Anwendungen ist sehr wichtig, um eine angenehme VR-Visualisierung ohne Latenz für die Benutzer zu ermöglichen. Animationen, Interaktionen und dynamische Elemente in der virtuellen Umgebung erhöhen die Prozessorleistung und reduzieren die Bildfolge rate, so dass die Nachführung der Bildsequenzen nicht flüssig genug ist und nicht zu den eigenen Kopfbewegungen passt. Dieser Effekt kann zum Unwohlsein des Betrachters führen. Um eine hohe Bildrate zu erreichen, wurden die Gebäudetexturen auf einen einzigen Texturatlas pro Gebäude mit 4K-Auflösung beschränkt, vier LoD-Modelle für jedes Mesh generiert und die Umgebung so gestaltet, dass die Anzahl der sichtbaren Gebäude in einem einzigen Frame reduziert wird. Dieser Workflow wird in Zukunft für ähnliche Projekte wieder angewendet, optimiert und weiterentwickelt.

## 9 Dank

Die Autoren danken dem Museum Schwedenspeicher in Stade, insbesondere Dr. Sebastian Möllers, sowie der Stadt Stade, insbesondere Dr. Andreas Schäfer, für ihre Unterstützung bei der Entwicklung der VR-Applikation Stade 1620. Darüber hinaus möchten die Autoren die finanzielle Unterstützung der Nico-Rüpke-Stiftung in Hamburg würdigen.

## 10 Literaturverzeichnis

- CHEVRIER, C., JACQUOT, K. & PERRIN, J.P., 2010: 3D Modelling of a Town Scale Model. EuroMed2010, 3rd International Conference dedicated on Digital Heritage, Short Papers, 99-107, Archaeolingua, Budapest.
- EDLER, D., HUSAR, A., KEIL, J., VETTER, M. & DICKMANN, F., 2018: Virtual Reality (VR) and Open Source Software: A Workflow for Constructing an Interactive Cartographic VR Environment to Explore Urban Landscapes. *Kartographische Nachrichten – Journal of Cartography and Geographic Information*, **68**(1), Bonn, Kirschbaum Verlag, 5-13.
- EDLER, D., KEIL, J., WIEDENLÜBBERT, T., SOSSNA, M., KÜHNE, O. & DICKMANN, F., 2019: Immersive VR Experience of Redeveloped Post-industrial Sites: The Example of “Zeche Holland” in Bochum-Wattenscheid. *Kartographische Nachrichten – Journal of Cartography and Geographic Information*, **69**(4), Bonn, Kirschbaum Verlag, 267-284.
- GAMESPARKS, 2019: Game Engine Analysis. <https://www.gamesparks.com/blog/game-engine-analysis/>, letzter Zugriff 21. Oktober 2019.
- HOLOPAINEN, J. & BJÖRK, S., 2005: *Patterns in Game Design*. Charles River Media Inc., Hingham, Massachusetts.
- JEDRZEJAS, T. & PRZYBILLA, H.-J., 2009: Aufbau historischer 3D-Szenarien am Beispiel der mittelalterlichen Stadt Duisburg. *PFG – Photogrammetrie, Fernerkundung, Geoinformation*, (3), 195-204.
- KELLER, F., SÄNGER, J., KERSTEN, T. & SCHIEWE, J., 2011: Historisches 4D-Stadtmodell der Freien und Hansestadt Hamburg - Automatisierte Generierung und Darstellung innerhalb der Google Earth Engine. *PFG – Photogrammetrie, Fernerkundung, Geoinformation*, (3), 155-169.
- KERSTEN, T., BÜYÜKSALIH, G., TSCHIRSCHWITZ, F., KAN, T., DEGGIM, S., KAYA, Y. & BASKARACA, A. P., 2017: The Selimiye Mosque of Edirne, Turkey - An Immersive and Interactive Virtual Reality Experience using HTC Vive. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, **42**(5/W1), 403-409, <https://doi.org/10.5194/isprs-archives-XLII-5-W1-403-2017>.
- KERSTEN, T., DEGGIM, S., TSCHIRSCHWITZ, F. & HINRICHSSEN, N., 2018a: Segeberg 1600 - Eine Stadtrekonstruktion in Virtual Reality. *Kartographische Nachrichten*, **68**(4), 183-191.
- KERSTEN, T., TSCHIRSCHWITZ, F., LINDSTAEDT, M. & DEGGIM, S., 2018b: The historic wooden model of Solomon’s Temple: 3D recording, modelling and immersive virtual reality visualisation. *Journal of Cultural Heritage Management and Sustainable Development, Special Issue: Digital workflows for the conservation and sustainability of historic places*, **8**(4), 448-464, <https://doi.org/10.1108/JCHMSD-09-2017-0067>.
- NEBIKER, S. & BARMETTLER, A., FISCHER, B., WEBER, E., 2009: Vom physikalischen Stadtmodell zum historischen 3D-GIS - Anforderungen und Lösungsansätze am Beispiel des historischen Solothurns. *PFG – Photogrammetrie, Fernerkundung, Geoinformation*, (3), 205-219.
- PAINTER, L., 2015: Hands on with HTC Vive virtual reality headset. <http://www.pcadvisor.co.uk/feature/gadget/hands-on-with-htc-vive-virtual-reality-headset-experience-2015-3631768/>, letzter Zugriff 21.10.2019.

- PFEIFFER, M., CARRÉ, C., DELFOSSE, V., HALLOT, P. & BILLEN, R., 2013: Virtual Leodium: From an Historical 3D City Scale Model to an Archaeological Information System. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **II-5/W1**, 241–246, <https://doi.org/10.5194/isprsannals-II-5-W1-241-2013>.
- SEDLACEK, D. & ZARA, J., 2009: Graph Cut Based Point-Cloud Segmentation for Polygonal Reconstruction. *Lecture Notes in Computer Science* 5876, 218-227, SpringerLink, [https://doi.org/10.1007/978-3-642-10520-3\\_20](https://doi.org/10.1007/978-3-642-10520-3_20).
- THOMPSON, E. M., HORNE, M. & FLEMING, D., 2006: Virtual Reality Urban Modelling - an Overview. *CONVR2006: 6th Conference of Construction Applications of Virtual Reality*, 3-4 August 2006, Florida, USA.
- TSCHIRSCHWITZ, F., RICHERZHAGEN, C., PRZYBILLA, H.-J. & KERSTEN, T., 2019: Duisburg 1566: Transferring a Historic 3D City Model from Google Earth into a Virtual Reality Application. *PFG – Photogrammetrie, Fernerkundung, Geoinformation*, **87**(1-2), 47-56.
- WARMOES, I., 2018: Le musée des Plans-reliefs. Diversité des collections (1668-1940) et enjeux contemporains. *Artefact* 7, 223-230.
- WIRTGEN, B. & BOHMBACH, J., 1974: Blick auf Stade. Ansichten und Pläne aus sieben Jahrhunderten. *Stadt-Sparkasse Stade*.

# Satellitenbild-basierte Geovisualisierung mittels immersiver VR

FLORIAN HRUBY<sup>1,2</sup> & RAINER RESSL<sup>2</sup>

*Zusammenfassung: Besieht man sich rezente Fachliteratur aus dem Bereich der Geoinformationswissenschaften, so ist ein wachsendes Interesse an Datenvisualisierungen mittels Technologien immersiver virtueller Realität (VR) unverkennbar. Diesem Interesse stehen jedoch limitierte VR-Fähigkeiten von Geoinformations- und Fernerkundungssoftware gegenüber, die eine direkte Datenausgabe auf stereoskopischen VR-Headsets nur sehr eingeschränkt unterstützen. Vor diesem Hintergrund diskutiert vorliegender Beitrag eine Arbeitsmethode, mit welcher Geodaten immersiv dargestellt und erfahrbar gemacht werden können. Diese Arbeitsmethode soll am Beispiel zweier immersiver VR-Szenarien näher ausgeführt und illustriert werden.*

## 1 Einleitung

### 1.1 Problemstellung

Während der vergangenen Jahrzehnte erlaubten die sich ständig weiterentwickelnden Fernerkundungstechnologien, Geodaten in immer größerer zeitlicher und räumlicher Auflösung zu erfassen und auszuwerten (FISHER et al. 2018). Satellitenbilder helfen also, Menschen räumlich entfernte Ereignisse und Plätze näher zu bringen. Gleichzeitig kann jedoch auch eine zunehmende Distanzierung und Entfremdung vieler Menschen von ihrer unmittelbaren und weiteren Umgebung beobachtet werden: Bereits 2002 stellten BALMFORM et al. in einer vielzitierten Studie fest, dass britische SchülerInnen mit Pokémon-Charakteren besser vertraut sind, als mit der heimischen Flora und Fauna. Dieses als *nature-deficit disorder* bekannt gewordene Phänomen (LOUV 2005) ist auch für den deutschsprachigen Raum dokumentiert: BRÄMER et al. (2016) konstatieren etwa im 7. *Jugendreport Natur*, dass auf die Frage nach heimischen Waldfrüchten auch Exotika wie Mangos genannt werden. Um diesen Widerspruch (unter Vorwegnahme auf Abschnitt 2) auf ein anschauliches Beispiel zu reduzieren, lassen sich also einerseits auf Grundlage hochaufgelöster Satellitenbilder Korallenriffe genau kartieren, während andererseits vielen Menschen völlig unbekannt ist, wie ein Riff aussieht und was seine ökosystemischen Funktionen sind.

Es kann nun als berechtigter Einwand vorgebracht werden, dass die Wissensvermittlung für ein breiteres Publikum kein primäres Ziel der Fernerkundung darstellt, und damit auch die eben genannte Distanzierung keine dringliche fachrelevante Fragestellung ist. Dennoch steht die Wissenschaft im Allgemeinen, und somit auch die Fernerkundung im Besonderen vor einer Reihe von Problemen, deren Lösung eine Einbeziehung der Öffentlichkeit voraussetzt, so dass aus vorhandenen Datenvolumina allgemein verständliche und zugängliche Information gewonnen werden

---

<sup>1</sup> National Commission for the Knowledge and Use of Biodiversity (CONABIO), Liga Periférico-Insurgentes Sur4903, MX-14010 Mexiko Stadt, E-Mail: [florian.hruby, rainer.ressl]@conabio.gob.mx

<sup>2</sup> Universität Wien, Institut für Geographie und Regionalforschung, Universitätsstr. 7, A-1010 Wien, Österreich, E-Mail: florian.hruby@univie.ac.at

muss. Der Klimawandel und seine verbreitete Wahrnehmung als räumlich, zeitlich und gesellschaftlich entferntes Phänomen ist das wohl prominenteste Beispiel einer solchen Frage, die einer erfolgreichen externen Wissenschaftskommunikation bedarf und zugleich ein auch für die Fernerkundung relevantes Problemfeld darstellt (SPENCE et al. 2012; LEISEROWITZ 2005).

Das zunehmende Interesse der Geoinformationswissenschaften an Datenvisualisierungen mit immersiver VR-Technologie stützt sich nun wesentlich auf die Annahme, räumliche Sachverhalte immersiv besser verständlich machen zu können als z.B. mit herkömmlichen desktop-basierten Anwendungen. Diese Annahme gründet sich wiederum auf Forschungen der Kognitions-wissenschaften, wonach NutzerInnen raumbezogene Probleme besonders gut begreifen, wenn sie sich im Problemraum anwesend fühlen und mit diesem so interagieren können, als ob sie wirklich dort wären. Man geht davon aus, dass ein solches Gefühl der Telepräsenz durch immersive VR-Technologie signifikant gesteigert werden kann (CUMMINGS & BAIENSON 2016).

Auf das bereits angesprochene und nachfolgend ausführlicher behandelte Thema eines Korallenriffs übertragen bedeutet das, dass wir selbiges auf Grundlage von Satellitenbilddaten nicht nur genau kartieren und durch Bodenbedeckungskarten beschreiben, sondern mithilfe dieser Informationen auch solche immersiven Visualisierungen erstellen können, die den NutzerInnen einen realistischen Eindruck vermitteln, in dem durch Fernerkundung erfassten Ökosystem anwesend zu sein.

## 1.2 Workflow

Vor dem Hintergrund der einleitend skizzierten Problemstellung soll im weiteren Verlauf des Artikels eine Arbeitsmethode präsentiert werden, mit welcher Fernerkundungsbasisdaten immersiv dargestellt werden können, d.h. in einem Format, das NutzerInnen erlaubt sich im visualisierten Raum anwesend zu fühlen. Die folgende Abbildung fasst diesen Arbeitsablauf graphisch zusammen und strukturiert somit auch den weiteren Aufbau des Beitrages.

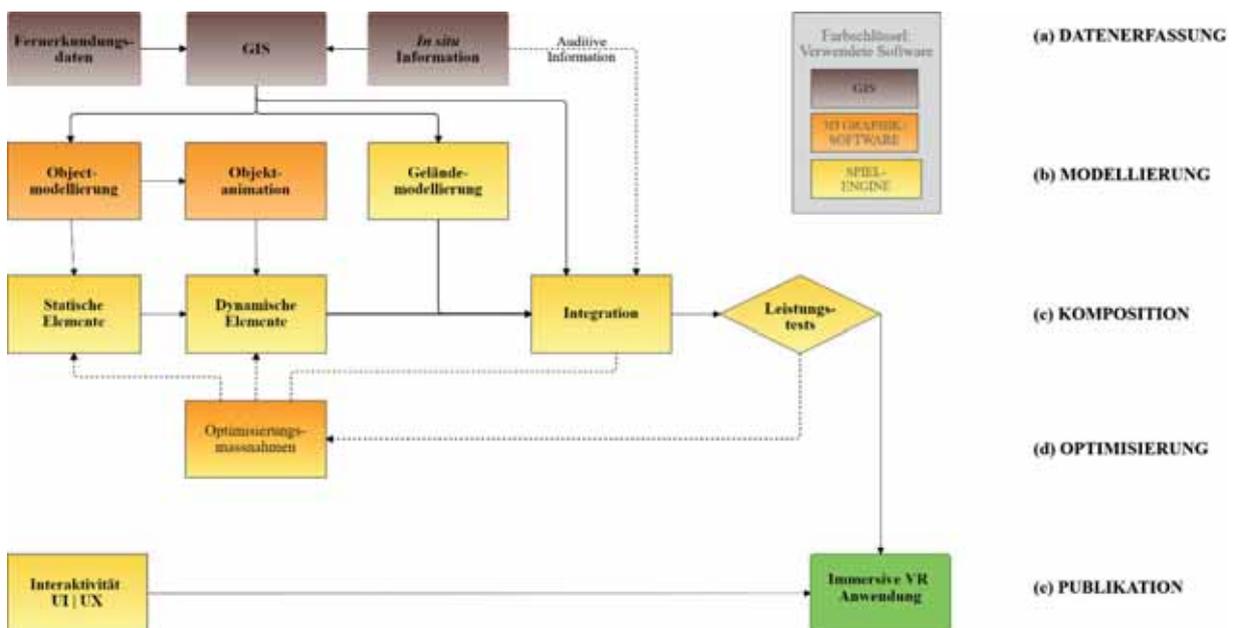


Abb. 1: Arbeitsschritte zur Erstellung satellitenbild-basierter Geovisualisierungen mittels immersiver VR

Auch wenn eine ausführliche Diskussion der Konzepte *Immersion* und *Telepräsenz* den Rahmen dieses Textes sprengen würde, sei explizit erwähnt, dass wir uns nur auf solche immersiven Geovisualisierungen im Maßstab 1:1 beziehen werden, in denen NutzerInnen Teil des dargestellten Raumes werden. Andere, kleinmaßstäbigere (1:x mit  $x > 1$ ) immersive Anwendungen, in denen etwa Karten oder Globen zwar via *head-mounted display* (HMD) genutzt werden können (YANG et al. 2018), NutzerInnen aber außerhalb des dargestellten Raumes bleiben, werden in unseren Überlegungen nicht berücksichtigt, auch wenn Teile der hier erörterten Methode auch für diese Art immersiver Geovisualisierung gelten. Ebenso wenig werden solche Applikationen diskutiert, welche die NutzerInnen im virtuellen Raum auf vorgegebene 360° Ansichten beschränken (ZHAO & KLIPPEL 2019), da in diesen Fällen das oben genannte Kriterium einer interaktiven Erfahrbarkeit nicht bzw. nur eingeschränkt erfüllt wird.

Schließlich sei auch noch darauf hingewiesen, dass wesentliche Teile des skizzierten Arbeitsablaufs bereits von anderen AutorInnen veröffentlicht bzw. besonders an bebauten Umgebungen erprobt wurden (EDLER et al. 2018, KERSTEN et al. 2018), nachfolgend jedoch ergänzt bzw. am Beispiel zweier immersiver VR-Szenarien unbebauter Ökosysteme konkret illustriert werden.

## 2 Datenerfassung

Immersive Anwendungen sind mittlerweile für verschiedene Ökosysteme entwickelt worden: Einem Großteil dieser VR-Applikationen ist gemeinsam, dass sie anstelle eines realen, georeferenzierten Ortes einen Prototypen repräsentieren, und relevante Flora und Fauna ohne den Anspruch möglichst realistischer Darstellung modellieren (vgl. AHN 2016). Um jedoch eine detailgenaue Beschreibung der naturräumlichen Gegebenheiten (z. B. Topographie/Bathymetrie, Flora und Fauna, atmosphärische Bedingungen) in Form eines digitalen Zwillinges (*digital twin*) zu erreichen, der den NutzerInnen erlaubt, sich im Sinne einer Telepräsenz eine angemessene Vorstellung davon zu machen, wie es an dem dargestellten Ort aussieht, ist ein Rückgriff auf Geodaten unerlässlich. Die nachfolgend genannten Anwendungen wurden mit dem Ziel einer realistischen Abbildung der beiden benachbarten Ökosysteme *Korallenriff* und *Mangrovenwald* in einer immersiven virtuellen Umgebung erstellt, und basieren daher auf Geodaten, die mittels Fernerkundungstechniken, GIS-Datensätzen und *in situ* erfasst wurden.

### 2.1 Datenerfassung zu Szenario A - Korallenriff

Auf atmosphärische Störungen korrigierte World-View-2-Satellitenbilder waren die Grundlage für die Ableitung genauer Bathymetriedaten sowie der wichtigsten benthischen Habitate im Bereich des mesoamerikanischen Riffsystems. Die relativ klaren karibischen Küstengewässer erlaubten eine Erfassung der Meeresböden bis zu einer Tiefe von ca. 20 m, womit das gesamte Korallenriff sowie die zur Küste anschließenden marinen Ökosysteme abgedeckt werden konnten. Mit Hilfe eines biooptischen Modells wurde der Einfluss der Wassersäule auf das vom Satelliten gemessene Rückstreuungssignal ermittelt, was zu einem „sauberen“ Signal für den nachfolgenden Klassifizierungsprozess führte, innerhalb dessen sieben benthische Bereiche abgegrenzt wurden.

Diese Klassifikation wurde abschließend auf Kleinstobjekte (*salt-and-pepper* Effekt) gefiltert, aggregiert und in einer thematischen Habitatkarte (Maßstab 1:8000) abgebildet. Interessierte LeserInnen seien auf CERDEIRA et al. (2012) für weiterführende Information zur Datenerfassung und -

klassifikation verwiesen. Abb. 2 zeigt die genaue Lage des modellierten Gebietes, welches sich über einen Bereich von  $3.5 \times 3.5$  km erstreckt, sowie die für diesen Bereich relevanten benthischen Klassen.

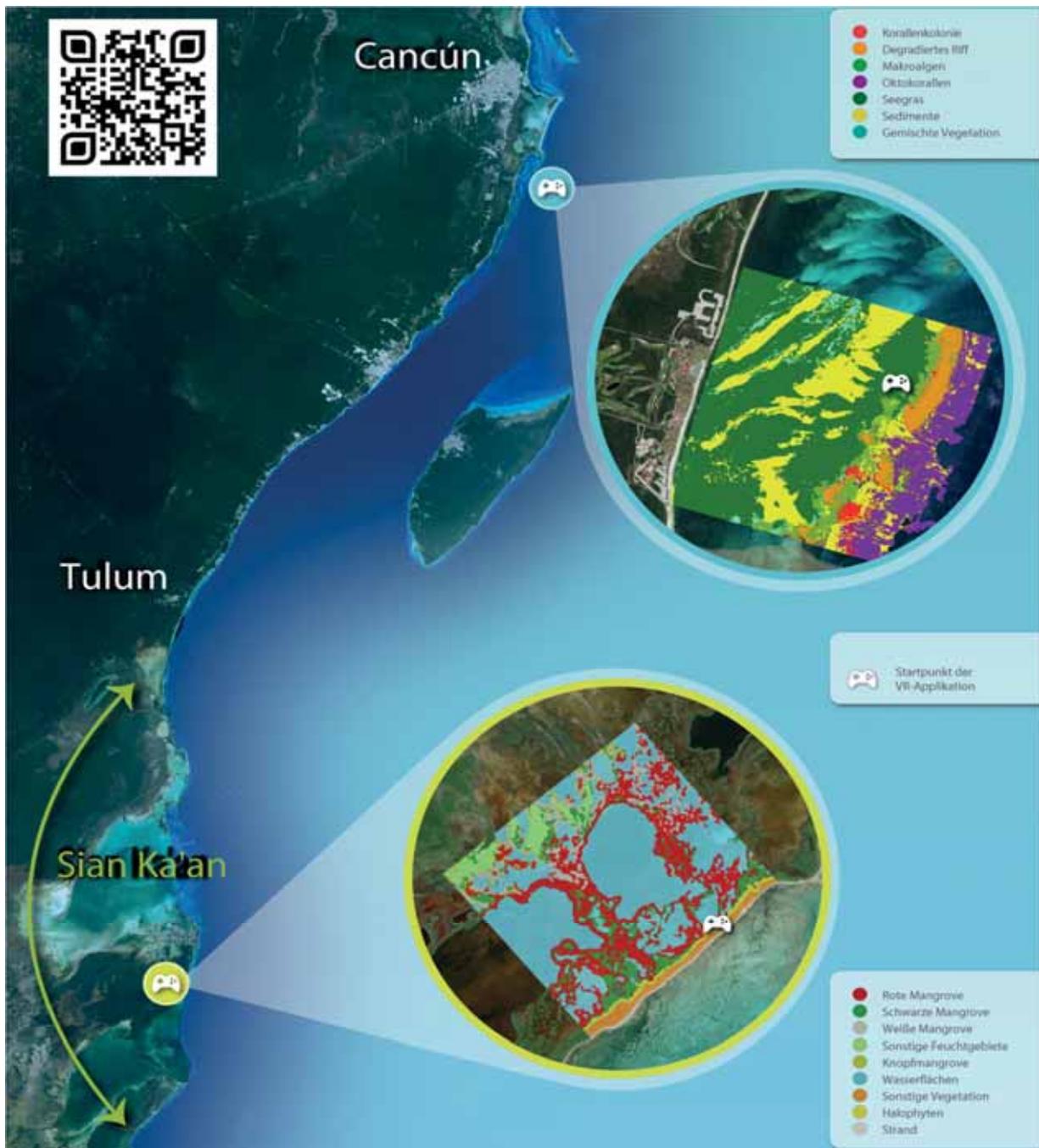


Abb. 2: Übersichtskarten zur Lage der beiden immersive visualisierten Ökosystemen samt der zugrundeliegenden Bodenbedeckungsklassen

Die Fernerkundungsbasisdaten bzw. die daraus abgeleitete Klassifizierung wurden mit georeferenzierten Einträgen aus dem *Nationalen Informationssystem über Biodiversität (SNIB)* angereichert, welches mehr als 12 Mio. Beobachtungen zu mexikanischer Flora und Fauna umfasst (SARUKHÁN & JIMÉNEZ 2016). Um ein möglichst vollständiges Bild der örtlichen Gegebenheiten zu bekommen, wurde auch auf georeferenzierte Fotos und Videos zurückgegriffen, mit deren Hilfe später orts- (z.B. Wassertrübung und -färbung) und spezienspezifische (z.B. Wuchshöhen von Seegrass und Korallen) Parameter realistisch simuliert werden konnten (vgl. Abschnitt 4).

## 2.2 Datenerfassung zu Szenario B - Mangrovenwälder

Grundlage des virtuellen Mangrovenwaldes sind Sentinel-2A Daten, aus denen atmosphärische Einflüsse mit der *Sen2Cor*-Software der Europäischen Raumfahrtagentur (ESA) entfernt wurden. Diese Daten wurden anschließend mit *in situ* erhobenen Referenzwerten zur Bodenbedeckung trainiert und in einem überwachten Klassifikationsverfahren (*nearest neighbour*) zu den wesentlichen Vegetationstypen zusammengefasst. Bathymetrie bzw. Mikrotopographie konnten im Falle des virtuell modellierten Mangrovegebietes (2.5 × 2.5 km) aufgrund der geringen Reliefenergie vernachlässigt werden (Abb. 2).

Wiederum wurde diese Datengrundlage um SNIB-Einträge ergänzt bzw. durch *in situ* gewonnene Information angereichert. Aufgrund der terrestrischen Umgebung konnten, im Gegensatz zum Unterwasserszenario, auch hochaufgelöste Photoaufnahmen für Texturierung und Farbgebung von 3D-Modellen aus Fauna und Flora verwendet werden. Durch Audioaufnahmen wurde zudem versucht, das vielfältige akustische Ambiente eines Mangrovenwaldes wiederzugeben.

## 3 Modellierung

Da vektorbasierte GIS versuchen, räumliche Phänomene klassifiziert als zweidimensionale Punkte, Linien und Polygone zu beschreiben, müssen diese geometrischen Basiselemente für eine immersive Geovisualisierung in konkrete (d.h. nicht-klassifizierte) 3D-Objektmodelle der für die zugrundeliegende Klasse typischen Elemente übersetzt werden. Diese 3D-Objekte sind danach lagerichtig einem digitalen Geländemodell (DGM) zu überlagern, welches die reale Topographie in möglichst hoher vertikaler und horizontaler Auflösung wiedergibt, und meist in einem Rasterformat vorliegt. Entsprechend ergeben sich als zentrale Aufgaben der Gestaltung einer VR-Umgebung die 3D-Modellierung von Einzelobjekten sowie des zugrundeliegenden Geländes. Während die Geländemodellierung ein relativ einfacher Prozess ist, welcher direkt in der Spiele-Engine erfolgen kann und im wesentlichen aus den Rasterzellwerten des DGM ein dreidimensionales Dreiecksnetz erstellt, ist für die Modellierung von Einzelobjekten ein Rückgriff auf 3D-Graphiksoftware (in den vorliegenden Szenarien *Blender* und *SpeedTree*) erforderlich, welche nicht nur die Wiedergabe der Objektform, sondern im Falle dynamischer Agenten auch eine entsprechende Animation ermöglicht (Abb. 1).

### 3.1 Objektmodellierung

Die unmittelbarste Form der Modellierung eines 3D-Objektes im virtuellen Raum beruht auf einer (semi-)automatischen Vermessung des realen Referenten selbst. Eine solche kann beispielsweise durch Methoden der Nahbereichsphotogrammetrie erfolgen, wobei Oberflächen aus 3D-Punkt-

wolken (*Structure from Motion*, Laserscanning) berechnet und zugleich hochaufgelöste Texturen erfasst werden können (CARRIVICK et al. 2016; KERSTEN et al. 2018; ZAMBRANO et al. 2017). Andere bildgebende Verfahren (z.B. Tomographien) erlauben zusätzlich, auch die innere Struktur des abzubildenden Objektes dreidimensional zu beschreiben (TRAN & VU 2016).

Die eben genannten Methoden eignen sich für die Erfassung unbelebter bzw. unbewegter Räume (z.B. Gebäude); in belebten bzw. dynamischen Umgebungen, wie es die beiden hier beispielgebenden Ökosysteme sind, können jedoch die benötigten stabilen Aufnahmebedingungen oft nicht hergestellt werden. Unter solchen Umständen bieten sich (semi-)manuelle Modellierungsverfahren an, bei denen auf Grundlage von Referenzen des realen Objektes (z.B. Photos oder Zeichnungen) dessen Aussehen im digitalen Raum durch verschiedene Methoden (z.B. *box-modelling*, *sculpting*) rekonstruiert werden kann (Abb. 3).

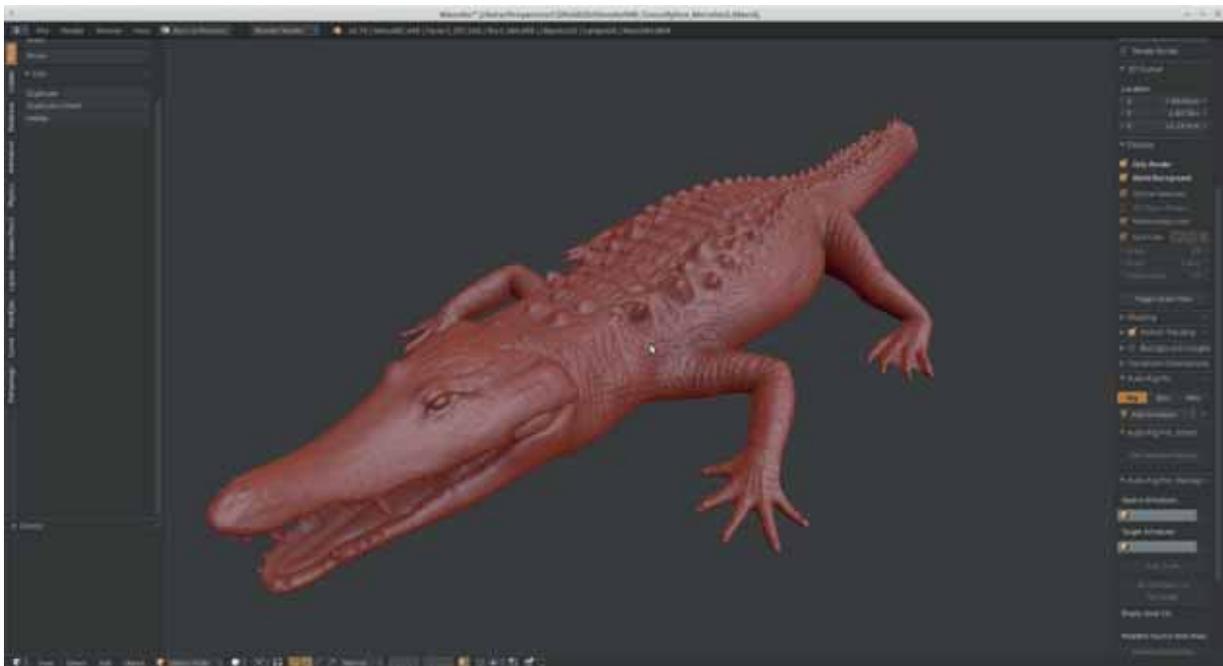


Abb. 3: 3D-Modellierung mittels Sculpting in Blender (Beulenkrokodil - *Crocodylus moreletii*)

### 3.2 Objektanimation

Vor dem Anspruch, einen geographischen Raum so zu repräsentieren, dass die NutzerInnen der entsprechenden Applikation das Gefühl der Anwesenheit am dargestellten Ort entwickeln können, ist es notwendig, nicht nur die entsprechenden Elemente dieses Raumes realistisch in Bezug auf ihre Form zu modellieren, sondern auch entsprechende motorische Muster von Fauna und Flora glaubhaft zu simulieren. Computeranimationstechniken (z.B. *rigging*, *vertex animation*) erlauben, die erstellten Objektmodelle auch komplexen Bewegungsabläufen zu unterwerfen (Abb. 4).

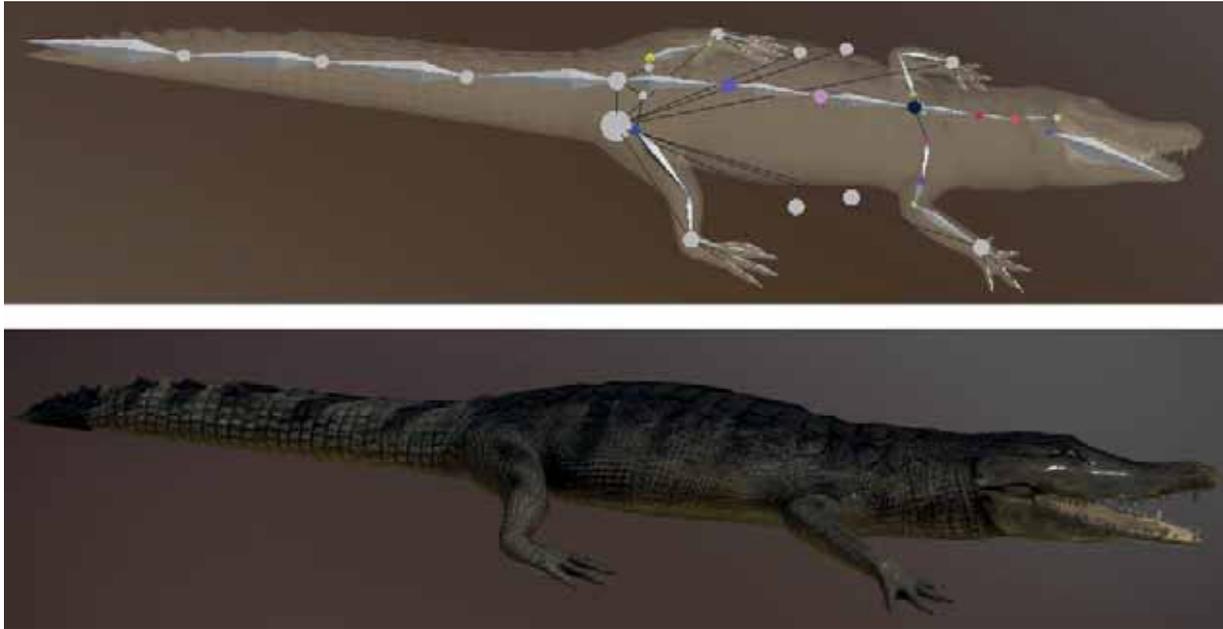


Abb. 4: Animationsrig (oben) und Render (unten) – Beulenkrokodil (*Crocodylus moreletii*)

## 4 Komposition

In den vorangegangenen Arbeitsschritten wurden aus Fernerkundungsbasisdaten sowohl Geländehöhenmodelle (Rasterformat) als auch Bodenbedeckungsklassen (Vektorformat) abgeleitet, und letztere in statische bzw. dynamische 3D-Modelle der wichtigsten Elemente von Fauna und Flora disaggregiert. Diese Bausteine können nun zu einem vollständigen virtuellen Ökosystem zusammengefasst werden. Um den Anforderungen einer interaktiven und immersiven Visualisierung gerecht zu werden, kann in diesem Arbeitsteil auf Spiele-Engines (im vorliegenden Fall: *Unreal Engine*) zurückgegriffen werden.

### 4.1 Distribution statischer Elemente

Auf dem bereits erstellten digitalen Geländemodell (vgl. Abschnitt 3) können alle statischen Objektmodelle lagerichtig verteilt werden. Dazu werden in einem Zwischenschritt die genannten Bodenbedeckungskarten dem Geländemodell als Texturen aufgebracht, um die für jede Klasse modellierten Objekt im VR-Raum entsprechend zu positionieren. Jede Bodenbedeckungsklasse wird also durch eine Reihe von Modellen der entsprechenden Flora (z.B. Mangrovenarten) und Fauna (z.B. Korallenarten) repräsentiert.

Um den unnatürlichen Eindruck gleichförmiger Objektverteilung zu vermeiden, werden neben einer Randomisierung von Speziesdichte und -häufigkeit auch zusätzliche Verteilungsparameter berücksichtigt, etwa um in Abhängigkeit von Seehöhe bzw. Wassertiefe die Wuchsgrößen und -formen relevanter Arten realistisch zu variieren. Weiters werden die in Bodenbedeckungskarten meist scharf abgegrenzten Klassen durch die Festsetzung von Übergangsgredienten zwischen benachbarten Kategorien in ein kontinuierliches und damit natürlicheres Verteilungsmuster gebracht (Abb. 5).

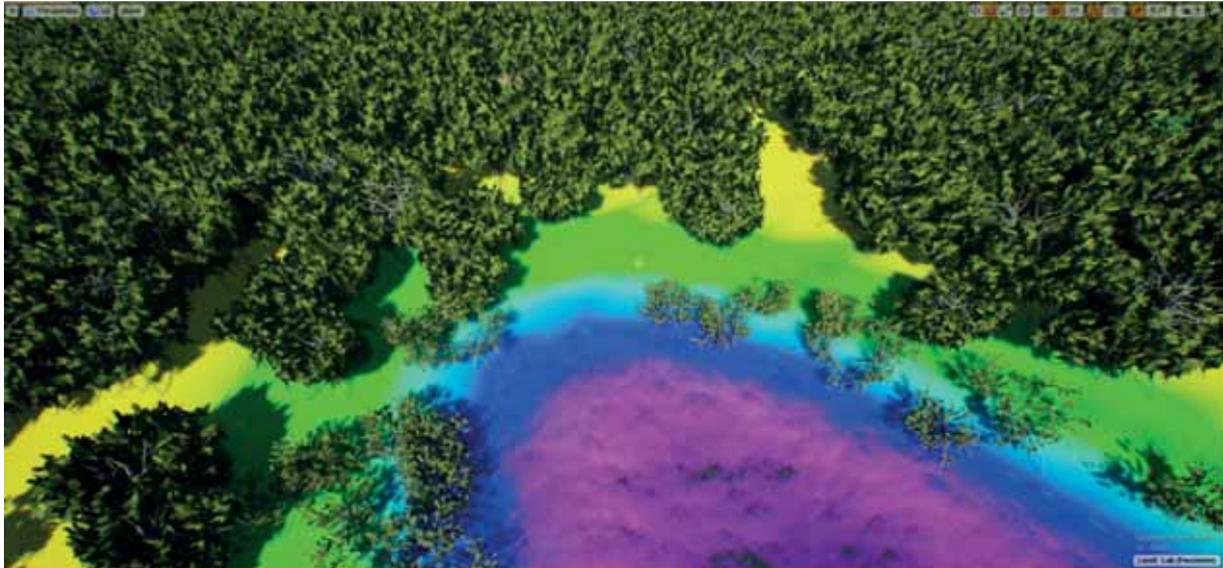


Abb. 5: Distribution von 3D-Modellen der Roten Mangrove (*Rhizophora mangle*) in Funktion der unterlegten Bodenbedeckungskarte. Der Übergang zu Wasserflächen (blau und violett gefärbt) wird durch entsprechend abnehmende Wuchsdichte und -höhe visualisiert

## 4.2 Distribution dynamischer Elemente

Während die Zuordnung statischer VR-Objekte auf Grundlage von GIS-Daten relativ einfach und im Grunde mit georeferenzierter Genauigkeit möglich ist, kann für die Verteilung dynamischer Agenten nicht ohne weiteres auf eine entsprechende Datengrundlage zurückgegriffen werden. Stattdessen lassen sich typische Lebensräume (z.B. von Fischen oder Vögeln) unter Zuhilfenahme von Fachliteratur und *in situ* gewonnener Information sowohl vertikal als auch horizontal (z.B. bekannte Flughöhen und Schwimmtiefen) abgrenzen.

Eine besondere Herausforderung besteht hierbei darin, nicht nur Bewegungen sondern auch Verhaltensweisen zu modellieren, im konkreten Fall etwa das Schwarmverhalten von Fisch- oder Vogelarten. Dies kann im einfachsten Fall durch manuell angelegte Animationspfade erfolgen, doch erfordern umfassendere Ökosysteme auch die Implementierung regel-basierter Bewegungsmuster, um komplexere Beziehungen (z.B. Räuber-Beute-Schemata) abbilden zu können.

## 4.3 Atmosphärische Gestaltung

Eine realistische Wiedergabe im virtuellen Raum ist nicht nur eine Frage von Objektmodellierung, -animation und lagetreuer -verteilung, sondern umfasst auch atmosphärische Faktoren, deren Berücksichtigung von GIS-Software nicht bzw. nur sehr eingeschränkt vorgesehen ist. Das Konzept einer Atmosphäre kann hier zunächst im engeren Sinn verstanden werden, so dass also auch Lichtverhältnisse, Wasser- und Lufttrübung, Wind- und Wellenbewegungen, etc. mit den Werkzeugen der Spiele-Engine modelliert werden.

Atmosphäre kann jedoch im Falle immersiver Geovisualisierung auch in einem weiteren Sinn berücksichtigt werden, nämlich in Bezug auf die Vermittlung einer raumspezifischen Stimmung. So können etwa verschiedene Lichtverhältnisse und Tageszeiten simuliert, Tierstimmen und Umweltgeräusche (auch raumakustisch) implementiert bzw. auch tageszeitenspezifisch variiert werden, um somit das angestrebte Gefühl der Telepräsenz noch konkreter zu vermitteln.

## 5 Optimierung & Publikation

Aktuellen Marktanalysen zufolge erreichen immersive Systeme ein immer noch kleines, aber beständig wachsendes Publikum, welches sich v.a. der HMD von *Oculus Rift* und *HTC Vive* bedient (STEAM 2019). Beide VR-Visoren sind mit den Ausgabeformaten aktueller Spiele-Engines kompatibel. Neben der Verbreitung geeigneter HMD stellen die Anforderungen an Rechen- und Graphikleistung des von den NutzerInnen verwendenden Computers einen wesentlichen Faktor für eine erfolgreiche immersive Geovisualisierung dar. Werden die entsprechenden Anforderungen (im Falle der vorliegenden Applikationen mindestens 6 GB expliziter Graphik-speicher) nicht erfüllt, fällt die Bildfrequenz der in Echtzeit zu rendernden Stereoszenen in einen Bereich, in dem eine nur verzögert oder ungleichmäßig laufende Anwendung den Eindruck von Telepräsenz (zer)stören oder gar Unwohlsein (*motion sickness*) hervorrufen kann.

Entsprechend kann bereits in der Entwicklungsphase (v.a. im Rahmen der Modellierung) versucht werden, die VR-Anwendung hinsichtlich der geforderten Rechenleistung möglichst zu optimieren. Beispielsweise können die Anzahl von Polygonen eines 3D-Objekts bereits während dessen Modellierung minimiert (z.B. durch manuelle Nachbearbeitung automatisch erstellter Geometrien), sowie unterschiedliche Detaillierungsgrade (*levels of detail*) definiert werden, so dass in Abhängigkeit der jeweiligen Distanz zwischen NutzerIn und Objekt nur diejenige Anzahl von Polygonen gerendert wird, die für den Eindruck realistischer Darstellung notwendig ist. Unterschiedliche Ökosysteme können diesbezüglich verschiedene Anforderungen stellen: Während im Falle eines Unterwasserambientes die natürliche Wassertrübung auch den einsehbaren Bereich und damit die Anzahl der jeweils zu berechnenden Objekte reduziert, kann die große Sichtweite im Falle eines Überblicks auf einen Mangrovenwald die Menge der darzustellenden Modelle so sehr erhöhen, dass entfernte Objekte idealerweise nur mehr durch zweidimensionale Geometrien (*bill boards*) visualisiert werden, ohne dass die NutzerInnen diese Reduktion visuell wahrnehmen können.

Einige dieser möglichen Optimierungsmaßnahmen werden auch vom jeweils gewählten Publikationsformat abhängig sein: Sind die Hardware-Settings des geplanten VR-Systems vorab bekannt (etwa im Falle einer Installation innerhalb eines Science Centres), kann die Anwendung entsprechend entwickelt werden. Soll hingegen ein breites Publikum mit unterschiedlichen Hardwarekonfigurationen angesprochen werden, können etwa verschieden große geographische Ausdehnungen angeboten, oder aber auch kostengünstigere Hardware (z.B. *Oculus Go* oder *Cardboard*-artige Systeme) mit graphisch entsprechend reduzierten Applikationen bedient werden.

In Abhängigkeit des Publikationsformates ergeben sich schließlich auch unterschiedliche Kriterien zur Gestaltung von User Interface (UI) und User Experience (UX): Während die beiden hier besprochenen VR-Ökosysteme mit jeder handelsüblichen Xbox-Steuerungskonsole bedienbar sind, stellen die beiden Marktführer *HTC* und *Oculus VR* auch produktspezifische Controller zur Verfügung, die innerhalb des jeweiligen Systems getrackt und daher mit erweiterten Funktionalitäten ausgestattet werden können. Augensteuerung ist z.B. in smartphone-basierten VR-Systemen eine weitere wichtige Interaktionsform, die entsprechend spezifische Anforderungen an UI und UX stellt.



Abb. 6: Immersive VR-Geovisualisierung am Beispiel der karibischen Ökosysteme *Mangrovenwald* (oben) und *Korallenriff* (unten). Zur besseren Anschaulichkeit werden beide Szenen für diesen Beitrag hier als nicht-stereoskopische Render dargestellt

Das im Beitrag vorgestellte virtuelle Korallenriff wurde bereits kostenlos unter einer CC BY-NC-ND 4.0 Lizenz veröffentlicht (<http://www.biodiversidad.gob.mx/geoviz>), eine Beta-Version des virtuellen Mangrovenwaldes stellen wir auf Anfrage gerne zur Verfügung. Bislang wurden mehr als 100 Elemente aus Flora und Fauna (semi)manuell modelliert und animiert (Abb. 3 & 4), die größtenteils über die 3D-Web-Bibliothek *Sketchfab* (<https://sketchfab.com/conabio>) einsehbar sind. Entsprechende Geometrien und Texturen geben wir wiederum auf Anfrage gerne weiter.

## 6 Fazit & Ausblick

Vorliegender Beitrag beschreibt eine Arbeitsmethode, mit der kleinmaßstäbige Fernerkundungsdaten in 1:1 skalierte, immersive VR-Geovisualisierungen transformiert, und somit den NutzerInnen so nahe gebracht werden können, dass diese sich im dargestellten Raum anwesend fühlen. Zentrale Arbeitsschritte wurden anhand je eines terrestrischen und marinen Ökosystems ausgeführt und illustriert. Auch wenn die vorgestellte Methode einen gangbaren Weg aufzeigt, können wir damit nicht beanspruchen, die einfachste oder gar einzige Möglichkeit einer satellitenbild-basierten Geovisualisierung in VR präsentiert zu haben. Zukünftige Verbesserungen des hier skizzierten Prozesses sind bei allen Arbeitsschritten zu erwarten: Beispielsweise gestatten Softwareprodukte wie ESRI's City Engine bereits eine Plugin-gestützte Ausgabe von Geodaten in Spiele-Engines, so dass für zukünftige Generationen von GIS eine größere Kompatibilität (z.B. stereoskopisches Rendern) mit immersiven VR-Systemen erwartet werden kann.

Unabhängig der gewählten Erstellungsmethode bieten realistische immersive Geovisualisierungen eine Vielzahl von Anwendungsmöglichkeiten, die mit aktuellen GIS weder angedacht wurden noch umsetzbar scheinen. Als Beispiel sei hier die (derzeit auch unter Mitarbeit der Autoren) untersuchte Möglichkeit genannt, einen auf die selbständige Erkennung invasiver Spezien hin entwickelten Roboter nicht in einer physisch-realen, sondern stattdessen in einer (systematischer modulierbaren) VR-Replika dieses Raumes maschinell zu trainieren. Als weiteres Anwendungsbeispiel, in dem immersive VR verwendet wird, um eine reales räumliches Problem zu erklären, ist der Versuch, kollektives Verhalten von Tieren und Menschen zu simulieren und zu verstehen (KLEIN et al. 2019; MAGALLANES et al. 2018).

Um abschließend auf die eingangs skizzierte Problematik einer erfolgreichen externen Wissenschaftskommunikation zurückzukommen, sei auch noch auf die Möglichkeit hingewiesen, mit immersiven Geovisualisierungen nicht nur gegenwärtige Räume zu visualisieren, sondern auch vergangene und mögliche zukünftige Szenarien, z.B. die Folgen eines sich wandelnden Klimas, nicht nur numerisch-abstrakt zu beschreiben, sondern auch anschaulich erfahrbar zu machen.

## 7 Literaturverzeichnis

- AHN, S., BOSTICK, J., OGLE, E., NOWAK, K., MCGILICUDDY, K. & BAILENSON, J., 2016: Experiencing nature: embodying animals in immersive virtual environments increases inclusion of nature in self and involvement with nature. *Journal of Computer-Mediated Communication* **21**(6), 399-419.
- BALMFORD, A., CLEGG, L., COULSON, T. & TAYLOR, J., 2002: Why conservationists should heed Pokémon. *Science* **295**(5564), 2367.
- BRÄMER, R., KOLL, H. & SCHILD, H.-J., 2016: 7. Jugendreport Natur 2016. [https://www.wanderforschung.de/files/jugendreport2016-web-final-160914-v3\\_1903161842.pdf](https://www.wanderforschung.de/files/jugendreport2016-web-final-160914-v3_1903161842.pdf).
- CARRIVICK, J. L., SMITH, M. W. & QUINCEY, D. J., 2016: Structure from Motion in the Geosciences. John Wiley & Sons, Chichester.
- CERDEIRA-ESTRADA, S., HEEGE, T., KOLB, M., OHLENDORF, S., URIBE, A., MÜLLER, A., GARZA, R., RESSL, R., AGUIRRE, R., MARIÑO-TAPIA, I., SILVA, R. & MARTELL DUBOIS, R., 2012: Benthic habitat and bathymetry mapping of shallow waters in puerto morelos reefs using

- remote sensing with a physics based data processing. IEEE International Geoscience and Remote Sensing Symposium, 4383-4386.
- CUMMINGS, J. J. & BAILENSON, J. N., 2016: How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, **19**(2), 272-309.
- EDLER, D., HUSAR, A., KEIL, J., VETTER, M. & DICKMANN, F., 2018: Virtual reality (VR) and open source software: a workflow for constructing an interactive cartographic VR environment to explore urban landscapes. *Kartographische Nachrichten* **68**(1), 3-11.
- FISHER, J. R., ACOSTA, E. A., DENNEDY-FRANK, P. J., KROEGER, T. & BOUCHER, T. M., 2018: Impact of satellite imagery spatial resolution on land use classification accuracy and modeled water quality. *Remote Sensing in Ecology and Conservation*, **4**(2), 137-149.
- KERSTEN, T.P., TSCHIRSCHWITZ, F., DEGGIM, S. & LINDSTAEDT, M., 2018: Virtual Reality for Cultural Heritage Monuments – from 3D Data Recording to Immersive Visualisation. Euro-Mediterranean Conference 2018, Ioannides, M., Fink, E., Brumana, R., Patias, P., Doulamis, A., Martins, J. & Wallace, M. (Hrsg.), *Lecture Notes in Computer Science* **11197**, Springer, Cham, 74-83.
- KLEIN, K., SOMMER, B., NIM, H., FLACK, A., SAFI, K., NAGY, M., FEYER, S., ZHANG, Y., REHBERG, K., GLUSCHKOW, A. & QUETTING, M., 2019: Fly with the flock: immersive solutions for animal movement visualization and analytics. *Journal of the Royal Society Interface*, **16**(153), 20180794.
- LEISEROWITZ, A., 2005: American risk perceptions: is climate change dangerous? *Risk Analysis* **25**(6), 1433-1442.
- LOUV, R., 2005: Last child in the woods: saving our children from nature-deficit disorder. Algonquin Books, Chapel Hill.
- MAGALLANES GUIJÓN, G., HRUBY, F., RESSL, R., AGUILAR SIERRA, V., DE LA BORBOLLA DEL VALLE, G. & RODRÍGUEZ PAZ, O., 2018: Modelling of collective movement in immersive environments. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* **42**(4), 397-402.
- SARUKHÁN, J. & JIMÉNEZ, R., 2016: Generating intelligence for decision making and sustainable use of natural capital in Mexico. *Current opinion in environmental sustainability* **19**, 153-159.
- SPENCE, A., POORTINGA, W. & PIDGEON, N., 2012: The psychological distance of climate change. *Risk Analysis* **32**(6), 957-972.
- STEAM, 2019: Steam Hardware & Software Survey. <https://store.steampowered.com/hwsurvey>
- TRAN, M. & VU, H., 2016: A Research on 3D Model Construction from 2D DICOM. IEEE Int. Conference on Advanced Computing and Applications (ACOMP), 158-163.
- YANG, Y., JENNY, B., DWYER, T., MARRIOTT, K., CHEN, H. & CORDEIL, M., 2018: Maps and globes in virtual reality. *Computer Graphics Forum* **37**(3), 427-438.
- ZAMBRANO, A., PADILLA, O., TOULKERIDIS, T., ZAPATA, J., ORDOÑEZ, E. & MATO, F., 2017: Software constraints for caves' virtual environments modeling. IEEE Symposium Series on Computational Intelligence (SSCI), 1-5.
- ZHAO, J. & KLIPPELT, A., 2019: Scale-unexplored opportunities for immersive technologies in place-based learning. IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 155-162.

# Herausforderungen bei der Integration von 3D: Daten, Devices und Dimensionen im AR-VR-Kontinuum

**CLAUDIA LINDNER<sup>1</sup>, JOHANNES SCHULTZ<sup>1</sup>, ANNETTE ORTWEIN<sup>1</sup>, KILIAN STAAR<sup>1</sup>,  
HENRYK HODAM<sup>1</sup>, CARSTEN JÜRGENS<sup>1</sup>, FABIAN SELG<sup>1</sup> & ANDREAS RIENOW<sup>1</sup>**

*Für den Einsatz von Geodaten im Schulbetrieb verfügen die wenigsten Schulen über ausreichende Hard- oder Software. Auch die mobilen Endgeräte der Schülerinnen und Schüler sind hierzu nur bedingt fähig, hier können jedoch zusätzliche Funktionen der Augmented und Virtual Reality (AR und VR) genutzt werden. In beiden Fällen liegt eine heterogene Ausstattung vor. Geodaten müssen daher aufwendig in GIS selektiert, vorprozessiert, exportiert, in Spieleentwicklungsumgebungen importiert und in 2D- und 3D-Umgebungen implementiert werden. Aufgrund von Datenformatinterkompatibilitäten und Dateigrößenbegrenzungen, Versionsunterschieden und variierenden Verfügbarkeiten von Daten müssen zielführende Lösungswege entwickelt werden.*

*Einige Lösungsmöglichkeiten bezüglich der genannten Daten, Devices und Dimensionen, die im Rahmen des Projektes KEPLER ISS in der Entwicklung von AR- und VR-Applikationen für den Schulunterricht gefunden und entwickelt wurden, werden in diesem Beitrag vorgestellt.*

## 1 Einleitung

Der Digitalpakt zwischen Deutschem Bund und Ländern ist nach langjährigen Verhandlungen endlich zu Stande gekommen. Ziele des Digitalpaktes sind der Aufbau schulischer Serverlösungen, W-LAN, digitale Lehr-Lern-Infrastrukturen, digitale Arbeitsgeräte und schlussendlich schulgebundene mobile Endgeräte, sofern das W-LAN bereits fertig ausgebaut ist (BMBF 2019). Bis zur endgültigen Umsetzung verbleibt die IT-Infrastruktur an vielen Schulen jedoch mangelhaft, mit einer heterogenen Ausstattung, was Hard- und Software angeht, denn an Schulen ist die hohe Diskrepanz zwischen dem aktuellen Stand der Technik und dem tatsächlichen Ausstattungsniveau besonders ausgeprägt (SÖBKE & STREICHER 2016).

Die Nutzung von Satellitendaten im Geographie-Unterricht wird von Bildungsstandards seit Jahren gefordert (DGFG 2008, DGFG 2014). Doch für die Verarbeitung oder auch nur Visualisierung von Geodaten müssen Rechner Anforderungen bezüglich Arbeitsspeicher, Prozessoren, Speicherplatz und einer Grafikkarte erfüllen, die deutlich über die Grundausstattung hinausgehen. Gleichzeitig erfordert das Lehren von Geoinformationssystemen (GIS) ein Vorwissen, das nicht in jedes Lehramtsstudium integriert ist, und das Erlernen mehr Zeit, als in regulären Stundenplänen und Schuljahren vorhanden ist. Um also Geodaten sinnvoll in den regulären Schulunterricht einzubringen, ist es von Vorteil, diese soweit zu reduzieren, dass sie auf mobilen Endgeräten funktionieren, und zwar auch auf denen der Schülerinnen und Schüler (SuS), während der notwendige Ausbau der IT-Infrastruktur der Schulen noch nicht abgeschlossen ist (LINDNER et al. 2019a).

---

<sup>1</sup> Ruhr-Universität Bochum, Geographisches Institut, Arbeitsgruppe Geomatik, Universitätsstr. 150, D-44801 Bochum, E-Mail: [claudia.lindner, johannes.schultz, annette.ortwein]@rub.de, kstaar@uni-bonn.de, [henryk.hodam, carsten.juergens, fabian.selg, andreas.rienow]@rub.de

Augmented Reality (AR) und Virtual Reality (VR) auf von SuS mitgebrachten Geräten („Bring your own device“, BYOD) gewinnen daher zunehmend an Bedeutung, auch zur Vermittlung wissenschaftlicher Inhalte. Die Voraussetzungen dieser mitgebrachten Geräte variieren jedoch stark, was bspw. das Betriebssystem oder dessen Version angeht. Apps für den Schulunterricht müssen dies mit einbeziehen, ebenso wie unterschiedliches Nutzerverhalten. Auf der anderen Seite werden Geodaten mit immer höherer raumzeitlicher Auflösung bereitgestellt, die theoretisch in den Unterricht einfließen könnten, die besagten mitgebrachten mobilen Geräte der LuL und SuS jedoch Hard- und Softwaretechnisch überfordern und auch ihre Nutzerinnen und Nutzer (NuN) vor Herausforderungen stellen. Individuelle Lösungsansätze müssen geschaffen werden, die auf so vielen Geräten und von so vielen NuN wie möglich angewendet werden können.

Im Rahmen der Projekte KEPLER ISS (ehemals Columbus Eye) und Fernerkundung in Schulen (FIS), die an den Universitäten Bochum und Bonn durchgeführt und vom Deutschen Zentrum für Luft- und Raumfahrt (Förderkennzeichen 50JR1701 & 50EE1703) gefördert werden, entstehen verschiedene AR- und VR-Lerneinheiten rund um die Themenfelder Erdbeobachtung für die Fächer Geographie und Physik. Durch das European Space Education Resource Office (ESERO) Germany werden sie in ganz Deutschland in die Schulen integriert und ebenfalls europaweit disseminiert. Anhand von zwei Beispielen für interdisziplinäre Lernmaterialien für den Schulunterricht mit Geodaten wird verdeutlicht, wie die o.g. Herausforderungen die Entwicklung von immersiven Applikationen beeinflussen und welche Lösungsansätze zielführend sind. Die hierzu verwendeten Programme sind Unity und Unreal, bei denen es sich um 3D-Spiele-Entwicklungsumgebungen handelt, Vuforia, welches die Bilderkennung für die AR bereitstellt, und AgiSoft, eine Photogrammetrie-Software.

## **2 Ausgangssituation: Raum für Verbesserung - Raum für Neues**

### **2.1 Nutzung von Synergieeffekten zwischen diversen AR-EO-Apps**

Im Projekt Columbus Eye wurden insgesamt sechs alleinstehende AR-Apps mit Erdbeobachtungsdaten entwickelt, davon vier auf Deutsch und zwei komplementäre englische Versionen (siehe Tabelle 1). Alle Apps sind jeweils einem Arbeitsblatt zugeordnet, das die Funktionen sinnvoll in den Unterricht einbringt (LINDNER et al. 2019a). Die Gesamtgröße aller sechs Apps beträgt fast 200 MB. Darin sind zahlreiche Redundanzen inkludiert: Die Inhalte und Funktionen des Nutzerinterfaces (User Interface, UI), das über die App und das Projekt informiert, Grundfunktionen für 3D-Animation, zusätzliche Funktionen, etc. werden zu jedem einzelnen App-Paket hinzugefügt. Auf den mobilen Endgeräten der Lehrerinnen und Lehrer (LuL) sowie der SuS sind auch zahlreiche andere Applikationen für verschiedenste Zwecke installiert. Manches beliebte Spiel nimmt allein 1-2 GB an Speicherplatz ein. Hinzu kommen Fotos, Videos, Musikdateien, System- und voreingestellte Apps des Herstellers, die sich zu mehreren GB aufsummieren und selbst großzügige Speicherplätze füllen. Für den Schulbetrieb ist es daher zwingen notwendig, die Apps kompakt zu implementieren. Während das W-LAN noch nicht flächendeckend ausgebaut ist, müssen sie zudem ohne Streaming auskommen: Nicht alle SuS haben unbegrenzten Internetzugang oder Netzabdeckung in ihrem Klassenzimmer. Sämtliche Teile einer modularen App müssen daher vorbereitend im heimischen W-LAN herunterladbar sein und sollten sich für den Fall, dass doch im Unterricht etwas nachgeladen werden muss, auf wenige MB beschränken.

Tabelle 1: Im Projekt Columbus Eye veröffentlichte alleinstehende Apps. Diese wurden am 27.06.2019 aus dem Store entfernt und durch die Gesamt-App ersetzt (vgl. Kap. 4.2) (Datenquelle: Google Play Console des Projekt-Accounts, Stand: 14.01.2020).

Bezeichnung	Veröffentlichung	Aquirierte neue Nutzer (kumuliert) bis 27.06.2019	Dateigröße (letzte Version)	Details
Im Auge des Sturms	19.07.2016	414	14,3 MB	ORTWEIN et al. 2017
(EN) Eye of the Cyclone	02.11.2016	97	24,9 MB	s.o.
Aralkum	22.06.2017	399	30,5 MB	SCHULTZ et al. 2017
(EN) Aralkum	15.05.2018	13	30,5 MB	s.o.
Erde bei Nacht	15.05.2018	44	61,9 MB	LINDNER et al. 2018
Erde-Mond-System	21.06.2018	152	37,4 MB	LINDNER et al. 2019b

Auf Fortbildungen wurde von teilnehmenden LuL wiederholt eine vereinigende App erbeten. Dabei erwarteten die LuL sowohl eine Reduzierung der Gesamdateigröße, als auch eine vereinfachte Nutzung. Dies ist auch im Sinne des Projektes, da diverse Lizenzmodelle sich inzwischen nach der Anzahl der veröffentlichten Apps richten. Weiterhin wünschten sich die LuL eine Version für iOS.

Die Überarbeitung und Zusammenfassung der sechs Apps soll zusammen mit einem Versionsupdate der verwendeten Software einhergehen. Es soll ein gemeinsames UI zur Übersicht entstehen, über die jeder App-Part, der eine der vorherigen individuellen Apps darstellt, einzeln heruntergeladen und gelöscht werden kann. Weiterhin soll es nicht mehr eigene Apps auf Englisch geben, sondern die gesamte App mit allen Parts lokalisiert, d.h. eine Übersetzung zur Verfügung gestellt werden.

## 2.2 Migration und Integration von 3D Modellen in AR- und VR-Applikationen

Die virtuelle Abbildung von komplexen dreidimensionalen Objekten ist heute eine gängige Methode, um Simulationen durchzuführen, wissenschaftliche Inhalte zu vermitteln, oder Daten im alltäglichen Leben zu visualisieren. Dass dreidimensionale Modelle auch in Zukunft eine immer wichtigere Rolle spielen werden, ist unbestreitbar: Umso wichtiger ist es, dass SuS bereits in der Schule mit der Thematik der 3D-Modellierung und den damit oft verbundenen Visualisierungstechniken für die Erstellung von VR- und AR-Applikationen in Kontakt kommen.

Im Zuge des KEPLER-ISS-Projektes werden zwei konkrete Maßnahmen zur Verbesserung des Verständnisses der 3D-Modellierung bei SuS verfolgt: Zum einen die 3D-Modellierung markanter Strukturen der Erdoberfläche und die anschließende Bereitstellung mittels VR-Anwendungen, zum anderen die Vermittlung von Kompetenzen zur 3D-Modellierung mittels Arbeitsblätter, die

über Smartphone-basierte AR-Apps die Inhalte anschaulich vermitteln. Dieser integrative Ansatz impliziert, dass sowohl PC-, Smartphone- als auch Tablet-basierte Anwendungen entwickelt werden müssen.

Aufgrund der Heterogenität der Endgeräte und vor allen Dingen der verwendeten Betriebssysteme ist es erstrebenswert, dass Anwendungen und die damit verbundenen Datensätze möglichst interoperabel in Bezug auf Hard- und Software entwickelt werden. Universelle Datenaustauschformate oder direkte Schnittstellen zwischen den benötigten Programmen sind jedoch nur bedingt vorhanden. Eine sehr große Anzahl von verschiedenen Standards und die fehlende Interkompatibilität erschweren insbesondere den Austausch und die Integration von 3D-Geodaten in VR-Umgebungen. Besonders die steigende raumzeitliche Auflösung von Geodaten, einhergehend mit zunehmender Datengröße, sowie unterschiedliche Datenformate und die unterschiedliche Leistungsfähigkeit der Endgeräte stellen hierbei große Herausforderungen dar. Je nach Anwendung müssen die Programme und auch Datensätze individuell auf das verwendete Endgerät angepasst werden. Basierend auf Daten des NASA-HDEV-Experiments und Astronautenbildern werden die dabei auftretenden Herausforderungen im Folgenden dargelegt.

### **3 Herausforderungen der Aufbereitung von 2D- und 3D-Daten**

Für die einzelnen Module müssen die jeweiligen Erdbeobachtungsdaten so in GIS-Programmen vorprozessiert werden, dass die benötigte Rechenkapazität für die Anzeige und Verarbeitung auf ein für Endgeräte nutzbares Maß reduziert wird. Gleichzeitig muss eine möglichst große Detailtreue in Bezug auf die räumliche und die spektrale Auflösung erhalten bleiben. Es werden die Prozessierung von hypso- und bathymetrischen Daten zur simplifizierten Simulation von Gezeiten in der Deutschen Bucht unter einer variablen Gravitationskraft und von Hyperspektraldaten des ISS-gebundenen Sensors Hyperspectral Imager for the Coastal Ocean (HICO) zur Unterscheidung potentiell giftiger und ungiftiger Algenblüten im Lake Erie vorgestellt. Beide Datensätze finden Verwendung in AR-Apps zum Einsatz von ISS- und Satellitenbildern im Schulunterricht. Die photogrammetrische Aufbereitung von Astronautenbildern in 3D-Modelle dient als Beispiel für die Vorprozessierung von Daten, die als 3D-Modelle in VR-Anwendungen Verwendung finden. Durch die Verbindung der Faszinationsfelder „Immersion“ und „Erdbeobachtung“ entstehen so Synergieeffekte, die didaktisch in Wert gesetzt werden, um die SuS für die MINT-Bildung begeistern, LuL in Bereich Virtual und Mobile Learning fortzubilden und zusätzlich Themen der nachhaltigen Entwicklung in den alltäglichen Unterricht zu integrieren.

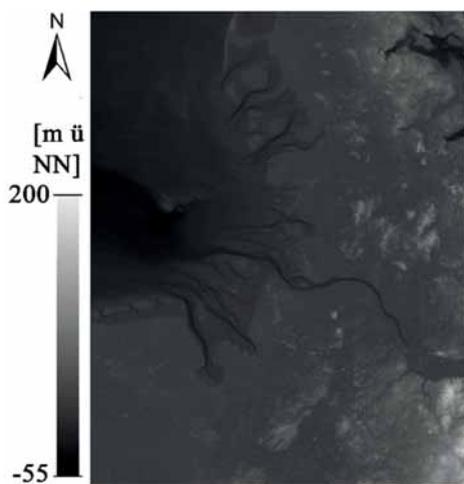
#### **3.1 Zweidimensionale Daten**

##### **3.1.1 Höhenmodell**

Für die Erde-Mond-System-App wurde ein Ausschnitt des ASTER GDEMs (METI & NASA 2011) der Deutschen Bucht als Basis genutzt. Vektordaten der Bathymetrie der Deutschen Nordsee (GEOPOTENTIAL DEUTSCHE NORDSEE 2013) wurden zu Rasterdaten mit gleichen Georeferenz-Daten umgewandelt und in das GDEM eingefügt. Rasterdaten des Baltikums waren nur mit erheblich geringerer räumlicher Auflösung zu erhalten (BALTIC HYDROGRAPHIC COMMISSION 2014), was

aufgrund der schlussendlichen Bildgröße in der Anwendung vernachlässigbar ist. Nach der Integration der drei Datensätze verblieben zahlreiche kleine Teile des Wattenmeers ohne Höhen- oder Tiefendaten. Diese wurden in originaler Auflösung der umgebenden Daten von 30m bilinear interpoliert. Aus dem Geotiff-Datensatz von 196 MB wurde zur Verarbeitung in der App ein 24-bit PNG von 61 KB generiert. Dabei entspricht jeder Graustufenwert von 0-255 einem Meter Höhen- bzw. Tiefenunterschied, beginnend beim tiefsten Punkt im Kartenausschnitt von 55 m unter NN (direkt südlich von Helgoland). Dies ist nur möglich, weil der Gradient im Bildausschnitt weniger als 256 m beträgt.

In Unity sind beim Einladen von Bilddaten Einstellungen vorzunehmen, die sich auf die Speicherung und Darstellung beziehen. Sie werden auf „Sprite (2D and UI)“ eingestellt, welches in der Entwicklungsumgebung für 3D-Welten die für 2D-Abbildungen optimierte Einstellung ist. Eine weitere mögliche Einstellung ist, die Pixeldaten der Bilder auszulesen und zu verändern („Read/Write enabled“). Dies zu aktivieren, bedeutet, dass stets eine Kopie des jeweiligen Bildes vorliegt; die vorgehaltene Datenmenge wird also verdoppelt. Angesichts der Anforderung an die App, Speicher- und Rechenleistungsverbrauch zu reduzieren, ist die Aktivierung dieser Einstellung nur bei absoluter Notwendigkeit vorzunehmen.



In der Erde-Mond-System-App ist daher nur ein Bild „Read/Write enabled“: Das finale Höhen- und Tiefenmodell, auf dem die Berechnungen für die Tiden simulation basieren (Abb. 1). Mit knapp 88.000 Pixeln und 61 KB hat es gerade so eine ausreichende Auflösung für die Betrachtung auf Tablets. Die Berechnung und Erstellung der überlagernden Wasserspiegel-Ebene bringt einige moderne Geräte an ihre Grenzen, die gleichzeitig auch noch die Bilderkennung mit der Kamera durchführen müssen. Daher wird nicht in jedem Frame (ca. 30/Sekunde), sondern nur in jedem dritten der Tidenhub neu berechnet und überlagert.

Abb. 1: Digitales Höhen- und Tiefenmodell der Deutschen Bucht (Eigene Darstellung in ArcGIS, Datenquellen: METI & NASA 2011, Geopotential Deutsche Nordsee 2013, Baltic Hydrographic Commission 2014)

### 3.1.2 Hyperspektraldaten

Die neueste App dreht sich um eine massive Algenblüte im Erie-See in Nordamerika im Jahr 2011. Aus einer HICO-Szene, die im Original über 128 Bänder von 353 bis 1080 nm mit den Ausmaßen von 512×2000 Pixeln verfügt, wurden die 87 Bänder zwischen 400 und 900 nm, welche über ein ausreichendes Signal-to-Noise-Ratio verfügen (LUCKE et al. 2011), ausgewählt und entsprechend den Anforderungen der Problemstellung zurechtgeschnitten. Die Daten enthalten eine Georeferenzierung, die sie jedoch aufgrund der Flugbahn vergleichsweise stark verzerrt und auf einer genordeten Karte sehr schief darstellt. Da den SuS die Daten so original wie möglich zur Verfügung gestellt werden sollen, aber auch aufgrund der begrenzten Größe und Auflösung von Smartphone-Bildschirmen, wurde stattdessen ein 1-dimensionales Resampling durchgeführt, das das Bild nicht

verzerrt aber die der Aufnahme inhärente Streckung der Pixel in Reihenrichtung annähernd ausgleicht.

Die resultierende Szene ist nur noch  $491 \times 725$  Pixel groß, passt in Originalgröße auf den Smartphone-Bildschirm und als 24-bit-PNG ist jedes einzelne Band nur 200-230 KB groß. Bei 87 Bändern summiert sich dies jedoch auf fast 19 MB.

Keines der HICO-Bänder ist Read/Write enabled, d.h., sämtliche Einfärbungen müssen ohne das Auslesen der Bilder funktionieren. Die Lösung hierfür sind spezielle Shader, die lediglich die Darstellung auf dem Bildschirm definieren.



Abb. 2: HICO-Szene des Westendes des Erie-Sees vom 03.09.2011 in Echtfarben, wie sie in der App eingesetzt wird: nicht genordet, in horizontaler Richtung eindimensional resampled (Eigene Darstellung, Datenquelle: NASA)

## 3.2 Dreidimensionale Modelle

### 3.2.1 3D-Modelle aus HDEV-Daten

Von April 2014 bis Oktober 2019 waren an der Außenseite des Columbus-Modul der ISS im Rahmen des NASA-HDEV-Experiments vier HD-Videokameras installiert, die die Erde aus verschiedenen Blickwinkeln durchgehend aufgenommen haben (RUNCO 2015). Bei einer durchschnittlichen Flughöhe der ISS von circa 400 km (STEFANOV et al. 2011), ergibt sich im Nadir eine Ground Sampling Distance (GSD) von 500 m bei einer Videoauflösung von  $1280 \times 720$  Pixeln.

Der ISS-Orbit ist sonnenasynchron und in Bezug auf Flughöhe und Lage hoch variabel (STEFANOV et al. 2011). Da HDEV über keinen eigenen Sternensensor verfügt und die Aufnahmezeit nicht gespeichert wird, ist eine automatische Georeferenzierung der Videos nicht möglich. Aufgrund der geringen Auflösung der Bilddaten und der fehlenden Georeferenzierung ist die Erstellung von 3D-Modellen möglich, aber aufwändig (vgl. hierzu SCHULTZ et al. 2018a; SCHULTZ et al. 2019). Die Videos werden im Super-Resolution-Verfahren verbessert, bei welchem die Pixelinformationen aus 15 Einzelbildern in einem Bild kombiniert werden, was die Auflösung der Bilder und somit die Bodenauflösung des Modells von 500 m (Nadir) auf 150 m erhöht. Es handelt sich dabei um ein Multiframe-Verfahren (IRANI 2002), welches mit mehreren niedrigauflösenden Bildern

derselben Szene arbeitet. Hierzu wurden in einem Abstand von 200 Frames 15 Frames aus dem Videomaterial (60 fps), mit MATLAB extrahiert und eine einfache Atmosphärenkorrektur (vgl. hierzu SCHULTZ et al. 2018) durchgeführt. Da sich die ISS zwischen den einzelnen Frames fortbewegt, sind die Einzelbilder um Subpixel verschoben und enthalten geringfügig veränderte Informationen zum Folgebild, die sich das Verfahren zu Nutze macht. Von jeder ausgewählten Szene wurden mittels Agisoft Photoscan Professional ein Höhenmodell, ein Orthomosaik und ein 3D-Modell erstellt und anschließend georeferenziert (AGISOFT PHOTOSCAN 2019). Dabei konnte die Ungenauigkeit der Referenzierung bei bisherigen Modellen auf etwa ein Drittel der Bodenaufklärung, also circa 50 m reduziert werden. Das in Abb. 3 dargestellte Modell wurde aus HDEV-Daten eines Überfluges über Chile am 29.08.2018 erstellt hat eine Ausdehnung von 1.500 km in der Länge und einer Fläche von circa 346.500 km<sup>2</sup>.

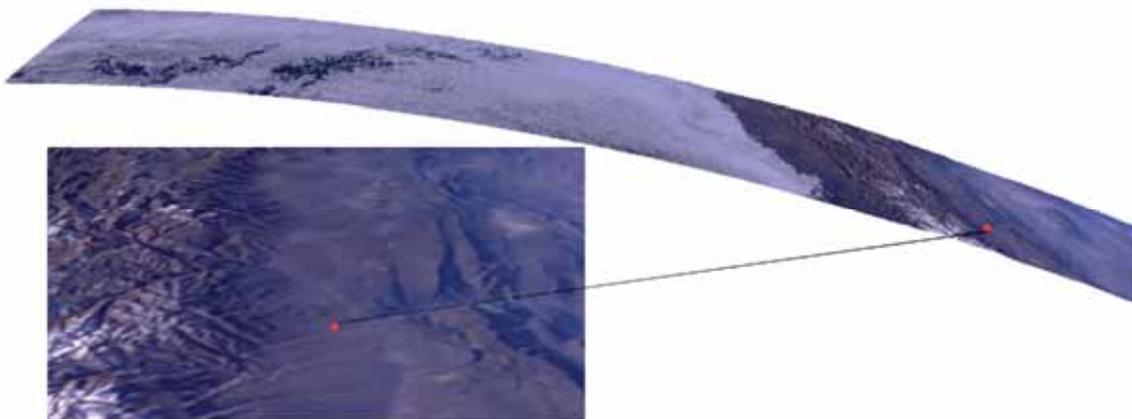


Abb 3: 3D-Modell abgeleitet aus HDEV Daten. 1500 km langer ISS Überfluges über Chile (Eigene Darstellung, Datenquelle: NASA)

Da die Höhenunterschiede im Modell schwierig zu erkennen sind, wurde ein weiteres 3D-Modell, basierend auf dem Orthomosaik und SRTM-Daten (Shuttle Radar Topography Mission, <https://www2.jpl.nasa.gov/srtm>), erstellt. Das photogrammetrische erstellte Höhenmodell und das auf SRTM-Daten basierende Höhemodell wiesen große Unterschiede auf. Diese Differenzen sind hauptsächlich auf Schatteneffekte und die geringe Auflösung der HDEV-Daten zurückzuführen. Allein der Vergleich der beiden Modelle gestaltet sich kompliziert, da sowohl im Im- als auch Export nur eine kleine und nicht notwendigerweise mit anderer verwendeter Software kompatible Auswahl von Dateiformaten zur Verfügung steht. So unterstützt ArcGIS nicht ohne Weiteres den Import von 3D-Modellen aus dem Wavefront OBJ Format, besonders Texturen und auch die Projektion führen zu Problemen. Desweiteren erschwert die ungenaue Georeferenzierung des aus HDEV-Daten abgeleiteten Modells den Vergleich mit den SRTM-Daten. Relativ einfach kann jedoch das aus HDEV-Daten mit Agisoft Photoscan Professional abgeleitete DGM mit den SRTM Daten verglichen werden.

### 3.2.2 3D-Modelle aus Astronautenbildern

Alle von Astronauten aufgenommenen Bilder von 1961 bis heute sind im „Gateway to Astronaut Photography of Earth“ der NASA frei verfügbar (<https://eol.jsc.nasa.gov>). Die hochauflösenden

Astronautenbilder der Erde eignen sich besonders gut für die Modellierung einzelner Strukturen in hoher Auflösung, wie z. B. Vulkane oder Bergzüge. Für die Entwicklung eines 3D-Modelles des 3.776 m hohen in Japan gelegenen Mt. Fuji wurden 27 Bilder (ISS046-E-35801 bis ISS046-E-35828), die am 08.02.2016 von der ISS aufgenommen wurden verwendet (SCHULTZ et al. 2019). Astronautenbilder eignen sich für die Erstellung eines 3D-Modells unter den Voraussetzungen, dass die Bilder in einem geringem zeitlichen Abstand aufgenommen wurden, sodass sie über einen großen Überlappungsbereich verfügen, einen geringfügig unterschiedlichen Aufnahmewinkel aufweisen und eine gleichbleibende Kameraeinstellung verwendet wurde (SCHULTZ et al. 2018b). Gleichzeitig haben die hier verwendeten Astronautenbilder eine hohe Auflösung von  $4.928 \times 3.280$  Pixel. Durch die Verwendung eines 1.150 mm Objektivs liegt die GSD bei circa 4 m. Das in Abb. 4 dargestellte 3D-Modell des Mt. Fuji wurde mit Agisoft Photoscan Professional erstellt. Agisoft Photoscan errechnet basierend auf der Überlappung der Bilder die genaue Position der Kamera. Zur Verbesserung des Modells wurden in den Ausgangsbildern 10 Marker manuell gesetzt, die auch die Georeferenzierung des 3D-Modells ermöglichen. Aus diesen Daten wird eine 3D-Punktwolke und ein Mesh berechnet. Hierbei führen besonders Bildbereiche mit Wolken zu großen Fehlern, die manuell korrigiert werden müssen. Auf das Mesh werden die einzelnen Bilder projiziert und daraus die Textur erzeugt. Die Ergebnisse können als einfaches, frei drehbares 3D-Modell (Abb. 4) dargestellt werden. Ferner stehen verschiedene Austauschformate in Agisoft Photoscan zur Verfügung. Ergänzend kann auch ein animiertes PDF oder ein Anaglyphenbild (MICHEL 2013) erstellt werden, welches einen besseren dreidimensionalen Seheindruck vermittelt. Durch die Vielzahl der verwendeten Bilder konnte die GSD von 4 m auf 2.68 m verbessert werden. Das Modell hat  $6804 \times 6628$  Pixel und besteht aus 6.105.526 Punkten.



Abb. 4: 3D-Modell des Mt. Fujis, erstellt aus Aufnahmen von der ISS (Eigene Darstellung, Datenquelle: NASA)

### 3.3 Aufbereitung für VR

Für die Verarbeitung eines AgiSoft-Modells in Unreal Engine bestehen grundsätzlich zwei Möglichkeiten; das Modell kann entweder als Mesh oder als Landscape mit direkter Einbindung in übergeordnete Landschaft eingeladen werden. Die Einbettung als Landscape hat einige Vorteile, darunter eine einfachere Bearbeitung direkt im Editor, bessere Performance und standardisierte

Werte für Interaktion und Kollisionsabfragen (UNREAL 2020a). Hierbei entstehen jedoch Skalierungsprobleme beim Import in die Engine, da metrische Einheiten nicht 1: 1 übertragen werden. Um die Größenverhältnisse zwischen Hintergrund, Spielerfigur und Landscape korrekt darzustellen, werden die Elemente relativ zueinander durch einen Skalierungsfaktor festgesetzt. Dies muss bei Anwendungen (messen etc.) berücksichtigt werden, da die Einheiten nicht den realistischen Entfernungs- und Größenmaßen entsprechen und dementsprechend a priori umgerechnet werden sollten.

Eine weitere Herausforderung ist die Anpassung der Datenkomplexität des zu Grunde liegenden Modells. Ein extrem hochaufgelöstes Höhenmodell ist zwar für die NuN der VR-Anwendung realistischer, erfordert aber entsprechende Hardware (und damit Zeit) zur Berechnung; für detaillierte Anwendungen können daher verschiedene Level of Detail eingebracht werden (UNREAL 2020a). Die Vereinfachung des Modells kann z.B. indirekt durch Umwandlung in PNG (256 - 766 Höhenabstufungen) erfolgen, dies macht jedoch eine manuelle Glättung der entstandenen Kanten für realistischere Darstellung ohne Terrassen notwendig, da, anders als im GDEM in Kapitel 3.1.1 der Gradient im Bild deutlich größer ist als die Anzahl der zur Verfügung stehenden Graustufen. Bessere Auflösungen können durch r16-Rohdaten (raw-Format 16bit) erreicht werden, die jedoch nur durch wenige Bildverarbeitungsprogramme überhaupt generiert, und nur eingeschränkt in die Engine importiert werden können – sie führen häufig zu lückenhafter Darstellung, große Höhenunterschiede werden z.T. nicht korrekt übertragen.

Daher hat sich im Anwendungsfall des Mt. Fuji eine manuelle Anpassung des aus AgiSoft im PNG-Format exportierten Graustufen-Höhenmodells durch Smoothing und Sculpturing im Editor der Unreal Engine als praktikablere Alternative herausgestellt. Bei dieser Vorgehensweise werden die Konvertierungsschritte zwischen den einzelnen Dateiformaten auf ein Minimum beschränkt und somit der Workflow effizient gestaltet, ohne zu viele Details des zu Grunde liegenden Modells zu verlieren. Der Informationsverlust ist hierbei abhängig von der Spannweite der Werte des generierten Höhenmodells, daher ist dieses Vorgehen für höher aufgelöste Höhenmodelle mit großen Höhenunterschieden eher nicht zu empfehlen – hier bietet sich ein Tiling der hochaufgelösten Höhenwerte an (UNREAL 2020a).

## 4 Herausforderungen des Anwendungsdesigns

### 4.1 Entwicklungsumgebungs-Versions-Upgrade

Die AR-Apps werden in der 3D-Spieleentwicklungsumgebung Unity erstellt. Die eigentliche AR entsteht durch die Bilderkennung der Software Vuforia, welche vordefinierte Zielbilder anhand von Kantendetektion identifiziert. In Unity können beliebig Videos, 3D-Elemente usw. durch die Zielbild-Erkennung aktiviert werden. Hier werden auch zweidimensionale Inhalte im UI implementiert.

Zu Beginn des Unterfangens war Vuforia in die neueste Unity-Version, 2017, integriert und beide Programme umfassend überarbeitet worden. Dabei wurde auch eine für die Columbus-Eye-Apps wichtige Schlüsselfunktion überarbeitet: In der AR dargestellte Videos nutzten zuvor instanziierte Funktionen, i.e. mehrere Videos konnten eingeladen und individuell auf ihren zugehörigen Zielbildern abgespielt werden. Nach der Überarbeitung wurde jedoch die Funktion, die das Abspielen der Videos steuert, statisch, und so pro Szene nur noch ein Video möglich. Da jede der Apps über

mindestens zwei abzuspielende Videos verfügt, war das Upgrade daher nicht möglich und es wird die aktuelle Version von Unity 5 genutzt. Diese wird vom Hersteller aktuell noch insofern unterstützt, als dass Sicherheitslücken geschlossen werden. Die Vorgabe des Google Play Store, allen Apps eine 64-bit-Version beizufügen, kann jedoch nicht eingehalten werden. Es existiert eine Ausnahmeregelung für Apps, die mit derartigen Unity-Versionen erstellt wurden, sodass diese noch bis 01.08.2021 als reine 32-bit-Version im Store veröffentlicht werden dürfen. Ob bis dahin Unity 5 entsprechend überarbeitet wird, um auch 64-bit-App-Versionen generieren zu können, ist nicht geklärt (UNITY TECHNOLOGIES 2019).

Ähnliche Probleme ergeben sich auch bei der Nutzung anderer Entwicklungsumgebungen wie der Unreal Engine; ein Level erfordert die Darstellung eines Videos auf einem Laptopbildschirm. Hierfür wurde ein entsprechendes Mesh erstellt und der „virtuellen Laptopbildschirm“ mit einer Funktion versehen, welche eine Videodatei (.mp4) abspielt, sobald bestimmte Voraussetzungen im Spiel erfüllt wurden. Ein Problem dieser angebotenen Funktionen ist, dass diese versionsabhängig sein können. In der Folge mussten sowohl die Videoabspieelfunktion als auch die Hierarchien der Objekte mehrfach angepasst werden, um die Kompatibilität mit der neuesten Unreal-Version sicherzustellen.

## 4.2 Gemeinsames Nutzerinterface

Das neue UI der AR-App hat die Hauptaufgabe, den NuN übersichtlich Informationen bereitzustellen. Beim erstmaligen Öffnen der App müssen die NuN darüber informiert werden, dass die App nur mit den entsprechenden Arbeitsblättern funktioniert. Danach wird dieses Wissen als gegeben angesehen und soll nicht mehr bei jedem Öffnen der App erscheinen (dies war ein expliziter Wunsch von LuL auf Fortbildungen). Beim Wegklicken der Nachricht wird eine Einstellung in den Android PlayerPrefs vorgenommen. Dadurch bleibt der Wert auch beim Neustart oder der Neuinstallation der App erhalten. Die Nachricht kann jedoch über einen der obersten Buttons (Abb. 5, „Willkommen!“) wieder abgerufen werden.

Weitere Buttons enthalten die gleichen Inhalte oder Links, wie schon in den individuellen Apps, führen zu den Arbeitsblättern und den rechtlichen Hinweisen. Nahe dem unteren Bildrand findet sich auch ein Button zum Beenden der App, der äquivalent ist mit dem „Stopp erzwingen“ im Android Anwendungsmanager oder dem Wegwischen in der App-Fenster-Ansicht, die App aber auch im Hintergrund direkt wieder startet. Dies ist erforderlich, wenn die App unerwünschtes Verhalten zeigt („Bugs“), und beendet werden muss. Nach Erfahrung in LuL-Fortbildungen und SuS-App-Tests kennen manche Smartphone-NuN die Funktion des „Stopp erzwingen“ bzw. den Unterschied zwischen dem in Android üblichen Verhalten, die Apps im Hintergrund weiterlaufen zu lassen, um sie schneller wieder starten zu können, und dem tatsächlichen Beenden der Apps, nicht. Auch der „Beenden“-Button war daher ein expliziter Wunsch von LuL.

Wie auch auf zahlreichen Webseiten üblich, sind die Spracheinstellungen am rechten oberen Rand des UIs zu finden. Zwei Flaggen stehen repräsentativ für die genutzten Sprachen: Die deutsche Flagge für die deutsche Sprache und der Union Jack des Vereinigten Königreiches für die britisch-englische Sprache (Abb. 5, siehe Kap. 4.4).

Die Haupt-App „Columbus Eye“ hat eine Dateigröße von 40,9 MB. Weitere 61,6 MB können hinzugeladen werden, um sämtliche App-Teile zur Verfügung zu stellen. Davon entfallen allein 30 MB auf den Part „Algenblüte“ mit den HICO-Daten und einem Landsat-Szenenausschnitt in Originalauflösung. „Columbus Eye“ wurde seit ihrer Veröffentlichung am 27.06.2019 bisher 122-Mal heruntergeladen (Stand: 14.01.2020, vgl. Tab 1).



Abb. 5: Ausschnitte des Nutzerinterfaces des Hauptmenüs der „Columbus Eye“-App (Eigene Darstellung)

### 4.3 Dynamisches Zu- und Entladen der App-Inhalte

Es sind eine Reihe zusätzlicher Skripte notwendig, die sich mit dem Finden, Herunterladen, Speichern und Löschen der Daten befassen. Eine Liste mit den Download-Links auf einem sftp-Server und ihrer Zugehörigkeit zu den einzelnen App-Parts ist in einer separaten Klasse gespeichert. Beim Betätigen des Download-Buttons für den jeweiligen Part wird abgefragt, welche Links heruntergeladen werden müssen, und diese, sofern nicht schon vorhanden, mit den jeweiligen Unity-Funktionen heruntergeladen. Bildsammlungen wie die HICO-Bänder sind dabei in AssetBundles zusammengefasst, einem Unity-eigenen Archivdateiformat. Videos werden einzeln heruntergeladen. Alles wird zusammen in einem gesonderten Ordner auf dem Smartphone gespeichert, der in Abhängigkeit von Hersteller und Betriebssystem variiert (bspw. unter Android 9: /Interner Speicher/Android/data/com.ColumbusEye.Main/files). Bei Bedarf können NuN so auch außerhalb der App bzw. ohne die Marker auf die Videos zugreifen und lernen dabei noch, durch die Eigenen Dateien auf dem mobilen Endgerät zu navigieren. Es wurde für den Download in Einzelteilen

gegenüber dem Download ganzer App-Parts entschieden, da sich diese einzelnen Parts zahlreiche Funktions-Skripte und sonstige Dateien teilen und es erneut zu Redundanzen gekommen wäre. Es ist möglich, den Download für mehrere Parts gleichzeitig durchzuführen, da es sich um eine asynchrone Funktion handelt, jedoch kann während eines aktiven Downloads kein anderer App-Part geöffnet werden. Dies würde den Download unterbrechen. Ein Download im Hintergrund ist nicht erwünscht, da dies die Apps verlangsamen würde. Sind alle Einzelteile eines App-Parts heruntergeladen wird der Download-Button durch einen Öffnen-Button ersetzt und der nebenstehende Info-Button, über den die NuN über das dynamische Zu- und Entladen informiert werden, durch einen Löschen-Button (vgl. Abb. 5). Bei Löschen der Dateien werden diese aus dem Speicher gelöscht und die Herunterladen- und Info-Buttons wieder aktiviert.

#### **4.4 Lokalisation**

Vorgefertigte Lokalisationseinstellungen von Unity sind erst ab der 2018er-Version verfügbar und können daher in der Columbus-Eye-App nicht genutzt werden (UNITY TECHNOLOGIES 2019). Die in den individuellen Apps enthaltenen Texte und Beschriftungen waren über das Unity-Developer-UI zunächst als konstante Werte hinterlegt. Für die Lokalisation wurde jedes textenthaltende UI-Element mit einem Skript versehen, das die Beschriftung aus einer Liste ausliest. Aufgrund der Simplizität wurde für diese Liste je eine JSON-Datei pro Sprache angefertigt, in dem sämtliche Texte sowie ihre Formatierung in html-Format hinterlegt sind, sofern etwas vom Standardformat abweicht. Dies wird auch die Übersetzung in weitere Sprachen erleichtern, die so auch von Personen ohne Programmierkenntnisse durchgeführt werden kann. Die zusätzlich notwendigen Skripte sind zudem sehr kompakt und benötigen insgesamt nur 11 KB an Speicherplatz. Die Umstellung von einer Sprache zur anderen erfolgt instantan beim Drücken des entsprechenden Flaggen-Buttons im Hauptmenü.

Innerhalb der App-Parts reicht die Methode jedoch nicht immer aus. Sobald die Zielbilder oder Bildinhalte sprachlich eindeutigen Text enthalten, müssen sie separat implementiert werden. Gleiches gilt für fest in Videos eingeblendeten Text (im Gegensatz zu einem in der App übergelagerten Text). Davon sind nur zwei Videos mit einer Größe von insgesamt ca. 3 MB betroffen. Zwar enthalten noch mehr Videos Text, dieser besteht jedoch überwiegend aus Zahlen und Einheiten und enthält wenige englische Worte, die für die deutschsprachigen SuS im Zielgruppen-Alter unproblematisch sind. Bildinhalte der App werden grundsätzlich ohne Text entwickelt, welcher dann Lokalisationsabhängig eingeblendet wird. Für Zielbilder der AR ist dies jedoch nicht möglich. Fast alle App-Parts benötigen daher einen separaten „übersetzten“ Part, in dem die Bilderkennung nach den lokalisierten Zielbildern sucht. In Abhängigkeit der gewählten Lokalisation wird beim Drücken des „Öffnen“-Buttons der deutsche oder englische Part geladen und ein Ladebildschirm mit ebenfalls lokalisiertem Text angezeigt. Um die Sprache zu wechseln, müssen NuN den Part wieder verlassen. Die meisten Zielbilder sind sich zwischen ihren jeweiligen Lokalisationen jedoch so ähnlich, dass sie auch in der anderssprachigen Szene funktionieren.

#### **4.5 Texturierung, Materialien und Interaktion in der VR**

Eine zusammengesetzte Luftbildaufnahme dient als Grundlage für die Textur des virtuellen Mount Fuji. Damit der visuelle Eindruck des Mount Fuji aus den Astronautenbildern erhalten bleibt, wurde eine große Gesamttextur erstellt, welche das Luftbild 1:1 ohne Kachelungseffekt auf die

erstellte Landschaft projiziert. Im Nahbereich führt diese Herangehensweise jedoch zu verwaschenen Texturen, da die Auflösung des Luftbildes zu gering ist. Durch Überblenden der Gesamttextur mit künstlichen, hochauflösenden Einzeltexturen kann ein realistischerer Eindruck erzeugt werden. Hierzu wurde ein Material erstellt, welches das Verhalten der einzelnen Texturen („texture blending“) in Bezug auf Entfernung zum Betrachter, Beleuchtungseffekte und die Lokation der Einzelkomponenten auf dem Modell steuert (UNREAL 2020b). Grundlage des letzten Punktes ist eine Minimum-Distance-Klassifikation der Oberflächen in die Klassen Schnee, Stein, und Vegetation, aus der eine RGB-Karte erstellt wurde. Für einen dreidimensionalen Eindruck der kleinteiligen Texturen wurde zudem eine Normal Map mit der Software Crazybump erstellt (<http://www.crazybump.com/>), um das Gelände auch im großen Maßstab realistisch wirken zu lassen. Sowohl für Materialien als auch Texturen sind frei verfügbare Bibliotheken erhältlich, die es ermöglichen, in kurzer Zeit auf das eigene Vorhaben angepasste eigene Materialzusammenstellungen zu entwickeln (UNREAL 2020c). Für eine stimmige Einbettung des Fuji-Modells in eine Szenerie wurden eine Umgebungslandschaft in geringerer Auflösung sowie eine Skybox erstellt. Hierbei sind Tag-Nacht-Wechsel ebenso möglich wie Wettereffekte.

Ein Vorteil der Verwendung von Spieleengines ist, dass sich vielfältige Eingabemethoden einrichten und gleichzeitig verwenden lassen. Somit lassen sich bereits vorhandene (und vertraute) Eingabemethoden wie Tastatur und Maus, Gamepad, etc. genauso verwenden wie VR-Motion Controller, was das Zurechtfinden innerhalb der Anwendung erleichtert. Die Bewegung innerhalb der VR-Anwendung erfolgt als Freiflug in Ich-Perspektive. Für den Actor werden verschiedene Funktionen definiert, unter anderem das Verhalten bei Kollision mit Landschaftselementen, und Steuerungsparameter wie Geschwindigkeit, Beschleunigung und Lage. Die Ausdehnung der Spielszene wird durch eine Blocking Volume begrenzt, welche den Spieler in seinem Bewegungsradius einschränkt und somit die Erstellung einer unendlichen Welt obsolet macht.

Die interaktiven Elemente der Anwendungen wurden in ein separates Level eingebunden, welches durch eine Teleportfunktion angesprochen wird. Fliegt der Spieler auf den Gipfel des Mt. Fuji, löst er eine Triggerfunktion aus, wird in das o.g. Level transportiert und wechselt die Perspektive von „fliegend“ auf „laufend“, was die Interaktion (in diesem Fall mit einem Laptop und einem Tastenfeld) intuitiver gestaltet. Diese Lösung wurde gewählt, um die Flugsimulation so authentisch wie möglich zu gestalten und die Aufmerksamkeit der SuS in dieser Anwendungsphase nicht durch interaktive Elemente abzulenken.

#### **4.6 Plattformübergreifendes Design**

Das verwendete Fuji-Modell der Desktopversion stellt, aufgrund der hohen Auflösung, zu hohe Ansprüche an die Leistung von aktuellen Smartphones und Tablets. Um das Problem zu lösen, wurde das Modell vereinfacht, indem die Anzahl an Vertices stark verringert und das Material auf die Verwendung von nur einer Textur (die des zusammengesetzten Astronautenbilds) begrenzt wurde. So konnte ein sinnvolles Gleichgewicht zwischen Informationsgehalt und Performance gewahrt werden (GOOGLE DEVELOPER 2020; UNREAL 2020e). Da es nicht ohne externe Eingabe möglich ist, mit der Anwendung zu interagieren, wurde auch die Interaktionskomponente eingeschränkt. Die Flugsteuerung wurde auf automatische Fortbewegung umgestellt, bei der Richtungsänderungen durch Kopfbewegung eingeleitet werden. Ein Stoppen ist somit nicht möglich. Auch die Interaktion mit Spielelementen erfolgt nun ausschließlich durch Fokussierung eines Objekts

für mehrere Sekunden. Dies schränkt bislang die Möglichkeiten ein, immersive Anwendungen mit interaktiven Elementen auszustatten. Bluetooth-Controller können diese Limitation beheben, sind dann jedoch unabdingbar und müssen von den Anwendern bereitgestellt werden. Dies widerspricht der Idee der ubiquitären Verfügbarkeit von Smartphones, die der Grund für die Portierung auf mobile Endgeräte ist.

## 5 Integration ins Klassenzimmer

SuS der Klassen 10-11 testeten die Erde-Mond-System-App auf ihre Funktionalitäten. Reaktionen auf die App fielen positiv aus und brachten selbst Physik-desinteressierte SuS dazu, sich für Themen wie Gravitation und Astronomie zu interessieren. Im Nachgang wurden zahlreiche Verbesserungsvorschläge implementiert, die sich auf Versionsunterschiede, Datengröße, Zielbilder und Übersichtlichkeit des UIs beziehen (LINDNER et al. 2019b). Dabei wurde auch klar, dass SuS sehr unterschiedliche Erwartungen an Lern-Apps haben. In der Diskussion mit der Klasse wurde bspw. der Wunsch nach einer Vertonung einer Funktion laut, bei der der Mond in die Erde stürzt: Während manche SuS sich eine comic-artige Explosion wünschten, diskutierten andere realistische Alternativen und kamen zu dem Schluss, dass sich im annähernden Vakuum des Weltraums kein Schall ausbreitet und in der Perspektive der NuN von hinter dem Mond daher kein Geräusch zu hören wäre.

Die erste lauffähige Version der VR-Anwendung wurde auf Publikumstagen ebenfalls von SuS getestet. Ihr mündliches Feedback fiel überwiegend positiv aus. Gelobt wurde die Verwendung von VR-Elementen im Bildungsbereich und den Vorstoß neuer Technologien in das Klassenzimmer. Das Thema, das als eher trocken empfunden wurde, konnte durch die Anwendung anschaulicher dargestellt werden und wurde mit größerem Enthusiasmus aufgenommen. Bemängelt wurde die Verfügbarkeit der benötigten Hardware in der Schule sowie die limitierten Interaktionsmöglichkeiten innerhalb der Anwendung. Um diesen Anregungen Rechnung zu tragen, wird die Anwendung aktuell auf mobile Endgeräte portiert. Um den immersiven VR-Charakter zu erhalten, können VR-Halterungen für Smartphones (z.B. Google Cardboard) eingesetzt werden. Hierbei treten einige Hürden auf, insbesondere hinsichtlich der Komplexität des Modells und der Heterogenität der Smartphone-Hard- und -Software, u.a. unterschiedliche Programmierschnittstellen wie OpenGL ES2 vs. 3.1, sowie unterstützte Multitexturformate (ETC1/2, DXT, ASTC, ...) (GOOGLE DEVELOPER 2020, UNREAL 2020d). Die Lösung liegt hierbei (noch) in der Bereitstellung multipler Formate für die unterschiedlichen Endgerätagenerationen.

## 6 Fazit und Ausblick

Die Entwicklung von AR- und VR- Anwendungen für den Schulunterricht steht in der Praxis besonders an den Schnittstellen von Daten, Devices und Dimensionen vor großen Herausforderungen. Für den sinnvollen Einsatz von Erdbeobachtungsdaten im Schulunterricht reicht die aktuelle Schul-IT-Infrastruktur nur in Ausnahmefällen aus und wird es auch in Zukunft aufgrund der erhöhten Anforderungen an Hard- und Software nicht flächendeckend. Eine Möglichkeit, derartige Daten trotzdem einzubringen, sind die mobilen Endgeräte von SuS im BYOD-Verfahren. Auch diese sind für herkömmliche Geodaten aufgrund der immer höheren raumzeitlichen Auflösung nur

bedingt geeignet, verfügen jedoch über zusätzliche Möglichkeiten, die Daten SuS-gerecht aufzuarbeiten und darzustellen.

Hierzu müssen Geodaten sorgfältig individuell selektiert und prozessiert werden, um sie in ihrer Datengröße zu reduzieren und in Dateiformate zu überführen, die von den weiterprozessierenden Programmen gelesen werden können. Oftmals geschieht dies über Umwege, da GIS- und Photogrammetrieanwendungen für die Visualisierung und Auswertung der Daten nur über wenige Schnittstellen-Formate mit Spielentwicklungsumgebungen verfügen, die für die Einbindung in Unterrichtsaufgaben genutzt werden. Eines der wenigen, nahezu universell verfügbaren Formate ist das PNG. So können Farbabstufungen unter bestimmten Voraussetzungen zur Übertragung von Höhendaten, auch im 3D-Format, genutzt werden.

Im Projekt KEPLER ISS wurden eine Reihe existenter AR-Applikationen für Android-Smartphones zusammengefügt, um Redundanzen zu reduzieren und die Bedienbarkeit zu verbessern. Dabei wird den NuN die Möglichkeit gegeben, die Anzeigesprache instantan zu wechseln, einzelne Inhalte dynamisch hinzuzuladen oder zu entfernen sowie auf einzelne Inhalte auch außerhalb der App zuzugreifen. Weiterhin wurde eine VR-Anwendung basierend auf Bildaufnahmen von Astronauten an Bord der ISS entwickelt, für die die besagten Aufnahmen zu einem hochauflösenden 3D-Modell prozessiert und für eine 3D-Umgebung exportiert wurden. Auch dieses soll auf den mobilen Endgeräten der SuS anwendbar sein, was jedoch zu Einschränkungen in den Bedienmöglichkeiten führt.

Im AR-Bereich ist in naher Zukunft die Migration der App in iOS vorgesehen, damit sie einer breiteren Masse an LuL und SuS zur Verfügung stehen. Als nächster Inhalt sind Interferometrie-Daten in einer Lektion über Vulkane vorgesehen, bei dem auch das 3D-Fuji-Modell zum Einsatz kommen wird. Bis dahin werden weiter die Möglichkeiten der Übertragung des VR-Erlebnisses auf Smartphones in Cardboards exploriert. Weitere 3D-Modelle für die Anwendung im Schulunterricht werden kontinuierlich erstellt. Eine formative Evaluation dieser Lernmaterialien in AR und VR findet kontinuierlich statt.

## 7 Danksagungen

Die Projekte Kepler ISS und FIS werden vom Deutschen Zentrum für Luft- und Raumfahrt e.V. mit Mitteln des Bundesministeriums für Wirtschaft und Energie aufgrund eines Beschlusses des Deutschen Bundestages gefördert (Förderkennzeichen 50JR1701 & 50EE1703).

Weiterhin danken wir der Remote Sensing Research Group der Universität Bonn, in der die Projekte für viele Jahre ausgeführt wurden, für den andauernden hervorragenden fachlichen Austausch.

## 8 Literaturverzeichnis:

AGISOFT PHOTOSCAN, 2019: Agisoft Metashape, Professional Edition, Version 1.5  
[https://www.agisoft.com/pdf/metashape-pro\\_1\\_5\\_en.pdf](https://www.agisoft.com/pdf/metashape-pro_1_5_en.pdf)

BALTIC HYDROGRAPHIC COMMISSION, 2014: Baltic Sea Bathymetry Database. <http://data.bshc.pro>

- BUNDESMINISTERIUM FÜR BILDUNG UND FORSCHUNG (BMBF), 2019: Verwaltungsvereinbarung: DigitalPaktSchule 2019 bis 2024. [https://www.bmbf.de/files/VV\\_DigitalPaktSchule\\_Web.pdf](https://www.bmbf.de/files/VV_DigitalPaktSchule_Web.pdf)
- DEUTSCHE GESELLSCHAFT FÜR GEOGRAPHIE (DGFG), 2008: Bildungsstandards im Fach Geographie für den Mittleren Schulabschluss, Bonn.
- DEUTSCHE GESELLSCHAFT FÜR GEOGRAPHIE (DGFG), 2014: Bildungsstandards im Fach Geographie für den Mittleren Schulabschluss, Bonn.
- GEOPOTENTIAL DEUTSCHE NORDSEE, 2013: Bathymetrie. [www.gpdn.de/](http://www.gpdn.de/)
- GOOGLE DEVELOPER, 2020: Unreal: Development for Android. <https://developers.google.com/ar/develop/unreal>
- IRANI, M., 2002: Multi-Frame Correspondence Estimation Using Subspace Constraints. International Journal of Computer Vision **48**(3), 173-194, doi: 10.1023/A:1016372015744.
- LINDNER, C., JÜRGENS C. & RIENOW, A., 2018: Fernerkundung und Raumfahrt in Schulen unter besonderer Berücksichtigung des Ruhrgebiets. In: Schulbuchinformationsdienst Ruhrgebiet **91**, 2-3.
- LINDNER, C., MÜLLER, C., ORTWEIN, A., HODAM, H., JÜRGENS, C., SCHULTZ, J., SELG, F. & RIENOW, A., 2019a: Das Klassenzimmer im Weltraum – Anwendungen zur Erdbeobachtung zwischen Realität und Virtualität. In: T. Kersten (Hrsg.): Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V., Band **28**, 241-255.
- LINDNER, C., RIENOW, A. & JÜRGENS, C., 2019b: Augmented Reality applications as digital experiments for education – An example in the Earth-Moon System. In: Acta Astronautica **161**, 66-74, doi: 10.1016/j.actaastro.2019.05.025 .
- LUCKE, R.L., CORSON, M., MCGLOTHLIN, N.R., BUTCHER, S.D., WOOD, D.L., KORWAN, D.R., LI, R.R., SNYDER, W.A., DAVIS, C.O. & CHEN, D.T., 2011: Hyperspectral Imager for the Coastal Ocean: instrument description and first images. Applied optics **50**(11), 1501-1516, doi: 10.1364/AO.50.001501.
- MICHEL, B., 2013: Digital Stereoscapy: Scene to Screen 3D Production Workflows. Stereoscapy News, Sprimont-Belgium.
- MINISTRY OF ECONOMY, TRADE, AND INDUSTRY & UNITED STATES NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (METI & NASA), 2011: ASTER Global Digital Elevation Model V002. doi: 10.5067/ASTER/ASTGTM.002
- ORTWEIN, A., KREBS, B., MEHRING, J., SCHULTZ, J. & Rienow, A., 2017: Neue Dimensionen der Erdbeobachtung im Schulunterricht: Raumfahrt, 3-D und Augmented Reality. In: T. Kersten (Hrsg.): Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V., Band **26**, 332-341.
- RUNCO, S., 2015: International Space Station – High Definition Earth Viewing (HDEV), [http://www.nasa.gov/mission\\_pages/station/research/experiments/917.html](http://www.nasa.gov/mission_pages/station/research/experiments/917.html)
- SCHULTZ, J., ORTWEIN, A., LINDNER, C., HODAM, H., WEPPLER, J. & RIENOW, A., 2017: Augmenting Pupil's Reality from Space – Learning with Digital Media based on Earth Observation Data from the ISS. Proceedings of the 68th International Astronautical Congress 2017 25-29 September in Adelaide, Australia, 1-6.

- SCHULTZ, J.A., ORTWEIN, A. & RIENOW, A. 2018a: Technical Note: Using ISS Videos in Earth Observation - Implementations for Science and Education. *European Journal of Remote Sensing* **51**(1), 28–32. doi:10.1080/22797254.2017.1396880
- SCHULTZ, J., HODAM, H., LINDNER, C., ORTWEIN, A., SELG, F. & RIENOW, A., 2018b: Ableitung von 3D-Modellen aus Daten des High Definition Earth Viewing-Experiments (ISS) –Anwendungen für den Schulunterricht. In: T. Kersten, Gülch, E., Schiewe.J., Kolbe, T., Stilla, U. (Hrsg.): *Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V.*, Band **27**, 129-140.
- SCHULTZ, J.A., HARTMANN, M., HEINEMANN, S., JANKE, J., JÜRGENS, C., OERTEL, D., RÜCKER, G., THONFELD, F., & RIENOW, A. 2019: DIEGO: A Multispectral Thermal Mission for Earth Observation on the International Space Station. *European Journal of Remote Sensing*, Online-Ausgabe, 1-11, doi: 10.1080/22797254.2019.1698318
- SÖBKE H. & STREICHER A., 2016: *Serious Games Architectures and Engines*. Dörner R., Göbel S., Kickmeier-Rust M., Masuch M., Zweig K. (Hrsg.): *Entertainment Computing and Serious Games*. *Lecture Notes in Computer Science* **9970**, Springer.
- UNITY TECHNOLOGIES, 2019: Unity User Manual. <https://docs.unity3d.com/Manual/index.html>
- UNREAL, 2020a: Unreal Engine 4 Documentation: Landscape Outdoor Terrain. <https://docs.unrealengine.com/en-US/Engine/Landscape/index.html#landscapefeatures>
- UNREAL, 2020b: Unreal Engine 4 Documentation: Landscape Materials. <https://docs.unrealengine.com/en-US/Engine/Landscape/Materials/index.html>
- UNREAL, 2020c: Unreal Engine 4 Documentation: Materials. <https://docs.unrealengine.com/en-US/Engine/Rendering/Materials/index.html>
- UNREAL, 2020d: Unreal Engine 4 Documentation: Android Game Development. <https://docs.unrealengine.com/en-US/Platforms/Mobile/Android/index.html>
- UNREAL, 2020e: Mobile Performance Tips and Tricks. <https://docs.unrealengine.com/en-US/Platforms/Mobile/Performance/TipsAndTricks/index.html>

# Analyse der Auswirkung der Wasseroberflächenmodellierung auf die Genauigkeit von Gewässerbodenpunkten in der Laserbathymetrie

KATJA RICHTER<sup>1</sup>, DAVID MADER<sup>1</sup>, PATRICK WESTFELD<sup>2</sup> & HANS-GERD MAAS<sup>1</sup>

*Zusammenfassung: Für die Refraktionskorrektur in der Laserbathymetrie werden genaue Informationen über die Geometrie der Wasseroberfläche benötigt. Grundlage für die Bestimmung des Einfallswinkels des Laserstrahls bilden Modelle der lokalen wellen-induzierten Wasseroberfläche, die aus den gescannten Wasseroberflächenpunkten abgeleitet werden. Konventionelle Methoden zur Modellierung der Wasseroberfläche vernachlässigen die lokale Krümmung der Wasseroberfläche. Die Abweichungen zwischen Wasseroberflächenmodell und tatsächlicher Wasseroberfläche verursachen einen Koordinatenversatz am Gewässerboden, der das Genauigkeitspotential des Messverfahrens limitiert.*

*Im Beitrag werden strengere Ansätze zur Modellierung der Wasseroberfläche vorgestellt, die auf einem parametrischen Modell sowie auf Freiformflächen basieren. Die genauere Repräsentation der Wasseroberflächengeometrie resultiert in einem geringeren Koordinatenversatz am Gewässerboden. Die Auswirkung der Wasseroberflächenmodellierung auf die Genauigkeit der Gewässerbodenpunkte wird mit Hilfe eines Laserbathymetrie-Simulators analysiert.*

## 1 Einleitung

Die Bestimmung geometrisch korrekter Gewässerbodenpunkte in der Laserbathymetrie erfordert die Korrektur der Refraktion auf Basis des Snellius'schen Gesetzes. Grundlage für die Refraktionskorrektur ist ein Wasseroberflächenmodell, aus dem genaue Informationen über die lokale Wasseroberflächenneigung abgeleitet werden können. Durch Verschneidung des einfallenden Laserstrahls mit dem Wasseroberflächenmodell kann der lokale Inzidenzwinkel zwischen Laserstrahl und Wasseroberfläche ermittelt werden. Das Wasseroberflächenmodell wird aus allen verfügbaren Oberflächenreflexionen abgeleitet. Bei signifikant ausgeprägten Wasseroberflächenwellen dürfen allerdings nur zeitlich eng benachbarte Wasseroberflächenpunkte einbezogen werden. Wie gut das Wasseroberflächenmodell die wahre Wasseroberfläche repräsentiert, ist einerseits von den Gegebenheiten bei der Datenerfassung (Scanauflösung, Komplexität des Wellenmusters, Datenlücken) und andererseits von der Methode der Modellbildung abhängig.

Konventionelle Methoden zur Modellierung der Schnittstelle Luft-Wasser basieren auf einer starken Vereinfachung der Wasseroberflächengeometrie. Am simpelsten ist die Annahme einer horizontalen ebenen Wasseroberfläche, an welcher der Laserstrahl gebrochen wird. Schon bei kleineren Wellen kann die Abweichung von der Ebenheit zu einem signifikanten Lage- und Höhenversatz am Gewässerboden führen. In vorangegangenen Studien wurde gezeigt, dass der Lageversatz am Gewässerboden, je nach Seegang, Strahldivergenz und Wassertiefe, einige Dezimeter bis hin

---

<sup>1</sup> Technische Universität Dresden, Institut für Photogrammetrie und Fernerkundung, Helmholtzstraße 10, D-01069 Dresden, E-Mail: [david.mader, katja.richter1, hans-gerd.maas]@tu-dresden.de

<sup>2</sup> Bundesamt für Seeschifffahrt und Hydrographie, Sachgebiet Geodätisch-hydrographische Verfahren und Systeme, Neptunallee 5, D-18057 Rostock, E-Mail: Patrick.Westfeld@bsh.de

zu mehreren Metern betragen kann (WESTFELD et al. 2017). Darüber hinaus muss mit Höhenfehlern im Dezimeterbereich gerechnet werden, die sich bei nicht-horizontalem Gewässerboden durch Anteile des Lageversatzes vergrößern. Komplexere Korrekturmethode versuchen daher, die tatsächliche Geometrie der Wasseroberfläche zu berücksichtigen, indem die detektierten Wasseroberflächenpunkte zu einem Dreiecksnetz vermascht werden (ULLRICH & PFENNIGBAUER 2011). Auch hier können Abweichungen zwischen der wahren Wasseroberfläche und der Repräsentation durch das Dreiecksnetz zu Koordinatenfehlern am Gewässerboden führen. Der Versatz in Lage und Höhe fällt jedoch geringer aus, als bei der ersten Methode (RICHTER et al. 2018). Effekte, die sich durch die Refraktion an einer gekrümmten Wasseroberfläche ergeben, werden bisher vernachlässigt.

In diesem Beitrag werden alternative Modellierungsansätze vorgestellt, die eine bessere Berücksichtigung der wahren Wasseroberflächengeometrie ermöglichen (Kapitel 2). Die Analyse der Auswirkung der Wasseroberflächenmodellierung auf die Genauigkeit der Gewässerbodenpunkte erfolgt mit Hilfe eines Laserbathymetrie-Simulators (Kapitel 3). Die Ergebnisse der Untersuchungen zeigen, dass die Komplexität der Wasseroberflächenmodellierung direkten Einfluss auf die Genauigkeit der Gewässerbodenpunkte hat (Kapitel 4).

## 2 Modellierung der Wasseroberfläche

Die Modellierung der Wasseroberfläche erfolgt auf Basis der gescannten Wasseroberflächenpunkte in unterschiedlichen Komplexitätsstufen. Die verschiedenen Methoden zur Wasseroberflächenmodellierung sind in Abbildung 1 dargestellt. Für den Vergleich mit konventionellen Refraktionskorrekturmethode wird die Wasseroberfläche durch horizontale sowie lokal geneigte ebene Wasseroberflächenelemente modelliert.

Darüber hinaus werden strengere Modellierungsansätze genutzt, mit denen die lokale Krümmung der Wasseroberfläche berücksichtigt werden kann. Eine Möglichkeit stellt die lokale Modellierung der Wasseroberfläche durch ein parametrisches Modell dar. Eine weitere Möglichkeit ist die Modellierung durch eine Freiformfläche. Die Bewegung der Wasseroberfläche während der sequentiellen Abtastung eines Oberflächenelements durch den Laserscanner kann vernachlässigt werden. In Hinblick auf die hohe Pulswiederhol- und Scanrate moderner Flugzeuglaserscanner (PRR > 100 kHz, Scanrate > 100 Hz) ist die Betrachtungsweise als lokale Momentaufnahme gerechtfertigt.

### Horizontale ebene Wasseroberflächenelemente

Zur Modellierung der Wasseroberfläche wird an jedem Wasseroberflächenpunkt ein lokales horizontal orientiertes Wasseroberflächenelement mit wellenabhängiger Höhenlage definiert (Abb. 1 (a)). Die Höhe des Wasseroberflächenelements entspricht der Höhe des zugehörigen Wasseroberflächenpunktes.

### Lokal geneigte ebene Wasseroberflächenelemente

Zunächst wird aus allen Wasseroberflächenpunkten mittels Delaunay-Triangulation ein Dreiecksnetz erstellt (Abb. 1 (b), schwarz). Die Dreiecke repräsentieren die geneigten Wasseroberflächenelemente. Da es sich bei den Eckpunkten der Dreiecke um unstetige Punkte handelt, können keine

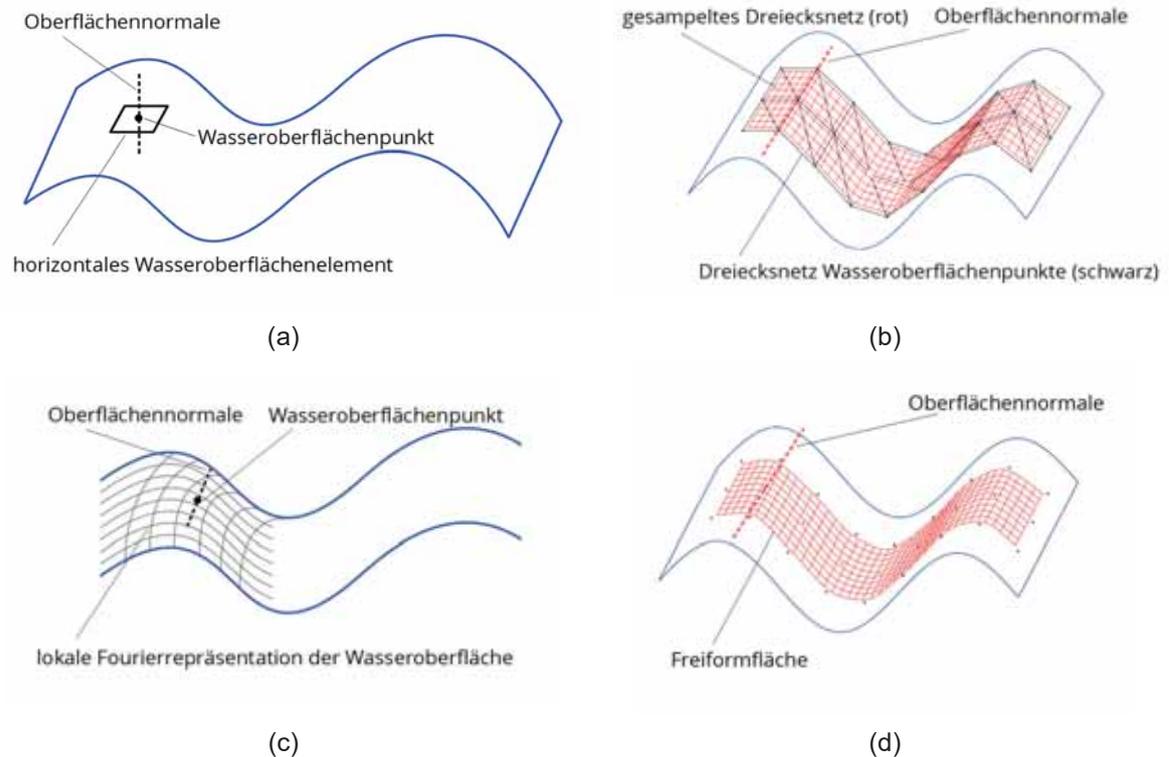


Abb. 1: Wasserflächenmodelle: horizontale Wasserflächenelemente (a), lokal geneigte Wasserflächenelemente (b), parametrisches Modell (c), Freiformfläche (d)

lokalen Neigungen berechnet werden. Aus diesem Grund wird das Dreiecksnetz in einem regelmäßigen Raster hochauflösend abgetastet (Abb. 1 (b), rot), wobei die Koordinaten der Rasterpunkte über lineare Interpolation im Dreiecksnetz bestimmt werden. Die lokale Neigung der Wasserfläche an der Stelle des Wasserflächenpunktes können anschließend im gesampelten Dreiecksnetz bestimmt werden.

### Parametrisches Modell

Als Grundmodell für die lokale Modellierung der Wasserfläche bietet sich eine Fourier-Repräsentation an (Abb. 1(c)). Jede stetige zweidimensionale Funktion mit der Periode  $T_x$  in x-Richtung und  $T_y$  in y-Richtung kann in eine Fourier Reihe entwickelt werden (CSANYI & TOTH 2005):

$$f(x) = \sum_{u=0}^{\infty} \sum_{v=0}^{\infty} \left( a_{uv} \cos \frac{2\pi ux}{T_x} \cos \frac{2\pi vy}{T_y} + b_{uv} \cos \frac{2\pi ux}{T_x} \sin \frac{2\pi vy}{T_y} + c_{uv} \sin \frac{2\pi ux}{T_x} \cos \frac{2\pi vy}{T_y} + d_{uv} \sin \frac{2\pi ux}{T_x} \sin \frac{2\pi vy}{T_y} \right) \quad \text{Gl. 1}$$

Die Parameter der Fourier Reihe werden durch eine Kleinste-Quadrate-Ausgleichung aus den Wasserflächenpunkten einer begrenzten Nachbarschaft bestimmt. Die Größe der Nachbarschaft richtet sich nach der Komplexität des Wellenmusters, welche die Anzahl der Harmonischen in x- und y-Richtung ( $u, v$ ) bestimmt.

## Freiformflächen

Freiformflächen werden aus abschnittsweise definierten Polynomfunktionen (z. B. B-Splines) generiert und erlauben die mathematische Beschreibung einer beliebigen Oberflächengeometrie (Abb. 1 (d)). Die Fläche wird durch folgende Gleichung mit den Punkten  $P_{i,j}$  des Kontrollrasters und den B-Spline-Funktionen  $N_{i,p}(u)$  und  $N_{j,q}(v)$  beschrieben (GRIMM-PITZINGER & RUDIG 2005):

$$S(u, v) = \sum_{i=0}^n \sum_{j=0}^m N_{i,p}(u) \cdot N_{j,q}(v) \cdot P_{i,j} , \quad \text{Gl. 2}$$

Die Parameter  $p$  und  $q$  beschreiben den Grad der Basisfunktionen und müssen nicht identisch sein. In einer Kleinste-Quadrate-Ausgleichung werden die Kontrollpunkte  $P_{i,j}$  bestimmt, so dass die resultierende Freiformfläche die gegebenen Wasseroberflächenpunkte durch eine doppelt gekrümmte Fläche annähert.

## 3 Numerische Simulation von Laserbathymetriedaten

In vorangegangenen Studien wurde ein Laserbathymetrie-Simulator entwickelt, mit dem die Datenerfassung für beliebige Wellenverhältnisse, Aufnahmekonstellationen und Sensoren nachgebildet werden kann (WESTFELD et al. 2017; RICHTER et al. 2018). Die Simulation umfasst die Modellierung der Wasseroberfläche, des Gewässerbodens, der Plattformbewegung, des Scanmechanismus, der Strahldivergenz und der Refraktion an der Wasseroberfläche.

Für die Modellierung der wellen-induzierten Wasseroberfläche wird das von TESSENDORF (2001) entwickelte Modell für Ozeanwellen verwendet. Es stellt eine gute Basis für die Simulation einer realistischen Ozeanoberfläche dar. Da es sich aber nicht ohne Weiteres invertieren lässt, ist es prinzipiell für die Erzeugung des Wasseroberflächenmodells aus gescannten Wasseroberflächenpunkten ungeeignet. Mit Hilfe des Modells wird für jeden Zeitpunkt  $t$  ein Höhenfeld in Form eines dichten regelmäßigen Rasters generiert. Das Höhenfeld repräsentiert eine realistische Ozeanoberfläche und kann über verschiedene Parameter (z. B. Gitterweite der Fourier Transformation, Windgeschwindigkeit, Windrichtung) modifiziert werden. Für die nachfolgenden Untersuchungen wurden Wellen mit einer Amplitude von ca. 2 m erzeugt. Als Grenzfall wurde im Wellenmuster außerdem ein größerer Wellenberg mit 4 m Amplitude integriert. Abbildung 2 zeigt eine Momentaufnahme der resultierenden Wasseroberfläche.

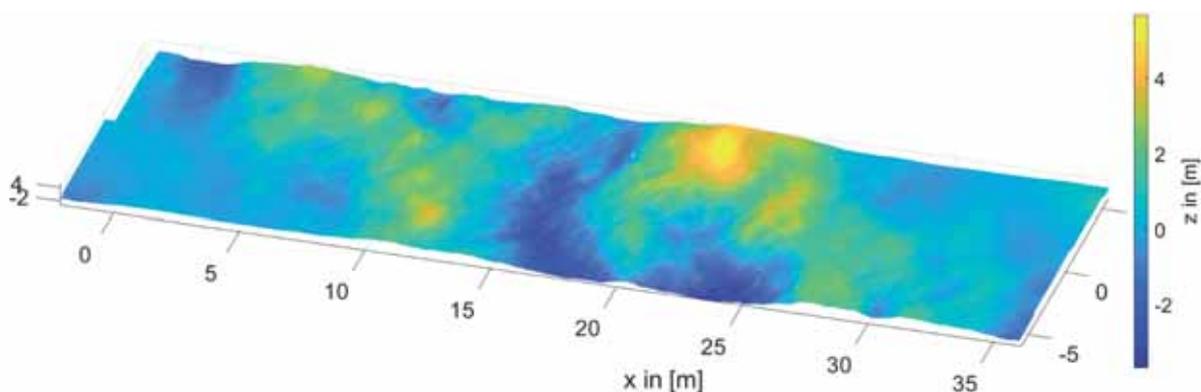


Abb. 2: Simulierte Wasseroberfläche nach Tessendorf

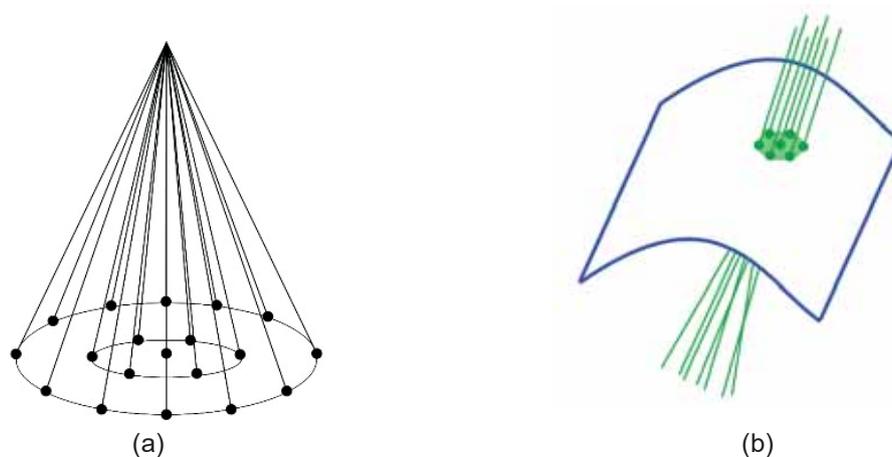


Abb. 3: Modellierung der Strahldivergenz (a) und der Refraktion (b)

Die Modellierung des Gewässerbodens erfolgt als horizontale Ebene mit einer vom Nutzer festgelegten Rauigkeit. Die Plattformbewegung wird als gleichförmige geradlinige Translation mit konstanten Kurswinkeln modelliert. Der Scanmechanismus wird entsprechend den Spezifikationen gängiger Laserbathymetriesysteme als Palmer-Scan simuliert.

Zur Modellierung der Strahldivergenz wird der Laserstrahl in eine große Anzahl von Teilstrahlen aufgespalten (insgesamt 41 Stück), die auf der Wasseroberfläche einen Laserfootprint mit endlichem Durchmesser repräsentieren (Abb. 3 (a)). Der Intensitätsverlauf innerhalb des einfallenden Laserpulses folgt einer gaußförmigen Intensitätsverteilung. Die Modellierung der Refraktion erfolgt differentiell, indem jeder Teilstrahl separat betrachtet wird (Abb. 3 (b)). Zunächst werden die Schnittpunkte mit dem Wasseroberflächenmodell und die lokalen Wellenneigungen bestimmt. Anschließend wird für jeden Teilstrahl die Refraktion an der Wasseroberfläche nach dem Snellius'schen Brechungsgesetz modelliert und der Auftreffpunkt auf dem simulierten Gewässerboden berechnet. Der intensitätsgewichtete Schwerpunkt der Gewässerbodenpunkte aller Teilstrahlen repräsentiert die finale Bodenpunktcoordinate für den gesamten Laserstrahl. Die Simulation basiert auf der Annahme eines identischen Hin- und Rückwegs. Die Auswirkungen von Mehrfachstreuung und Dispersion werden vernachlässigt.

Die Simulation geht zunächst von idealen Bedingungen aus, d.h., dass jeder emittierte Laserpuls ein Echo an der Wasseroberfläche und am Gewässerboden erzeugt. In der Realität ist das allerdings je nach Ausrichtung der Wellen nicht der Fall. Abbildung 4 verdeutlicht die Problematik an einem Beispieldatensatz. Abbildung 4 (a) zeigt die Dichte der detektierten Wasseroberflächenpunkte in einem 1m-Raster. Um den Zusammenhang zwischen Punktdichte und Ausrichtung der Wasseroberfläche zum Laserscanner zu untersuchen, wurde der Winkel zwischen der Richtung des Laserstrahls und der Oberflächennormalen berechnet (Abb. 4 (b)). Bei ebener Wasseroberfläche beträgt der Winkel durch die bei einem Palmer-Scanner geneigte Scanachse  $20^\circ$ . Ein kleinerer Winkel bedeutet, dass das Wasseroberflächenelement dem Scanner zugewandt ist, ein größerer, dass es abgewandt ist. Im konkreten Beispiel liegen die Werte zwischen  $10^\circ$  und  $30^\circ$ . Der Vergleich von Punktdichte und Ausrichtung der Wellen zeigt, dass auf der dem Scanner zugewandten Seite der Wellen mehr Wasseroberflächeninformationen verfügbar sind.

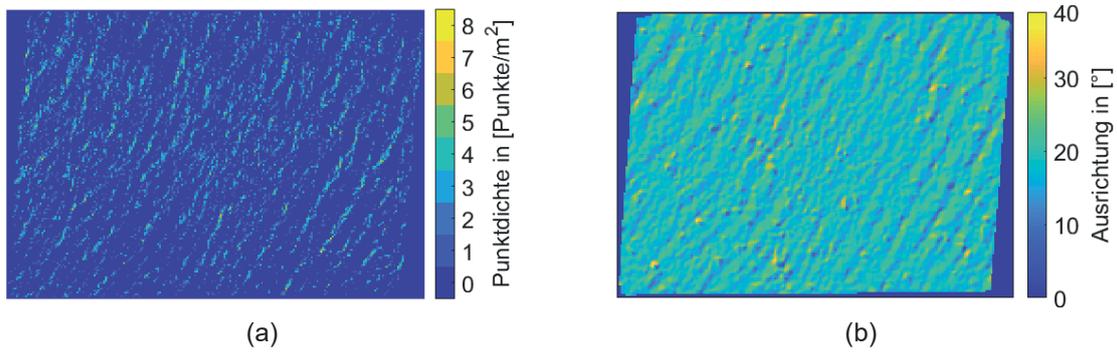


Abb. 4: Beispieldatensatz: Punktdichte im 1m-Raster (a) und Ausrichtung Rasterzellen (d)

Für eine möglichst realistische Simulation der Laserbathymetriedatenerfassung soll die ausrichtungsabhängige Variation der Punktdichte berücksichtigt werden. Zu diesem Zweck wird die modellierte Wasseroberfläche (s. Abb. 2) in der Simulation zunächst vollständig abgescannt. Abbildung 5 (a) zeigt die resultierenden Wasseroberflächenpunkte (schwarz) sowie die 1m-Rasterzellen. Anschließend werden die simulierten Wasseroberflächenpunkte in Abhängigkeit von der Ausrichtung der Wasseroberfläche zum Laserscanner ausgedünnt. Grundlage für diese Filterung ist die Verteilung der Punktdichte im realen Beispieldatensatz (Abb. 6). Zunächst wird die Ausrichtung der Rasterzellen zum Laserscanner berechnet. Nachfolgend wird die Anzahl der Punkte pro Rasterzelle in Abhängigkeit von der Ausrichtung so reduziert, dass die Häufigkeitsverteilung dem realen Datensatz entspricht. Im konkreten Fall wird beispielsweise ein Großteil der Rasterzellen mit ungünstigster Ausrichtung komplett geleert, so dass die Punktdichte 0 Pkt./m<sup>2</sup> wie im realen Datensatz zu 60 % vertreten ist. Abbildung 5 (b) zeigt die verbleibenden Wasseroberflächenpunkte. Der auf diese Weise erzeugte Simulationsdatensatz dient als Grundlage für die nachfolgenden Untersuchungen.

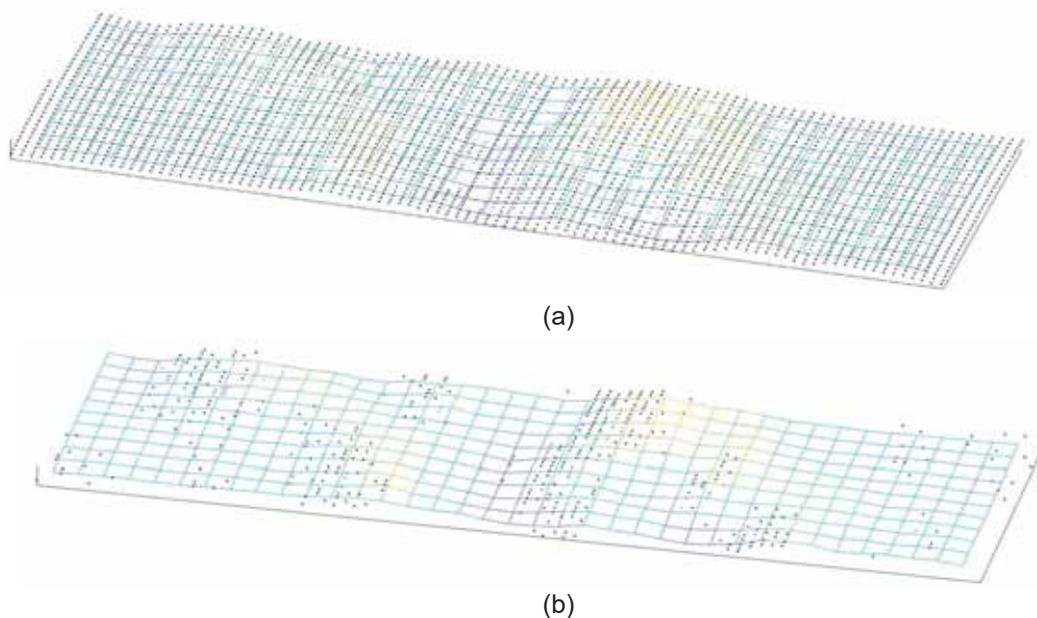


Abb. 5: Simulation der gescannten Wasseroberflächenpunkte unter idealen (a) und realen (b) Bedingungen

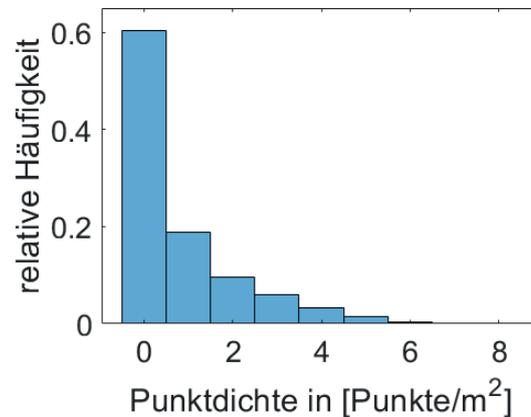


Abb. 6: Punktdichte im Beispieldatensatz

#### 4 Analyse der Auswirkung der Wasseroberflächenmodellierung

Die Modellierung der Wasseroberfläche erfolgt auf Basis der simulierten Wasseroberflächenpunkte in unterschiedlichen Komplexitätsstufen. Im Gegensatz zu einer realen Messkampagne ist die lokale wellen-induzierte Wasseroberfläche in der Simulation exakt bekannt und kann für die Evaluierung der Wasseroberflächenmodellierung genutzt werden. Die nach TESSENDORF (2001) simulierte Wasseroberfläche dient als Referenz für den Vergleich mit der Wasseroberfläche, die auf Basis der in der Simulation gescannten Wasseroberflächenpunkte mit unterschiedlichen Methoden modelliert wurde. Da sich die Wellen während des Scanvorgangs in der Simulation weiterbewegen, muss die simulierte Wasseroberfläche für den Vergleich aus den einzelnen Simulationsschritten zusammengesetzt werden. Dazu werden die simulierten Wasseroberflächen in einer lokalen Umgebung um jeden gescannten Wasseroberflächenpunkt zu einer Referenzfläche kombiniert. In Abbildung 7 (a) sind die simulierte Referenzfläche (grau schattiert) und die modellierte Wasseroberfläche (farbcodiert) für die Methode mit lokal geneigten Wasseroberflächenelementen dargestellt. Die Abweichungen zwischen simulierter Wasseroberfläche und auf simulierten Scandaten basierendem Wasseroberflächenmodell können leicht berechnet werden. Für die Modellierung der Refraktion sind allerdings nur die Bereiche des Wasseroberflächenmodells relevant, in denen Messdaten vorliegen. Die Untersuchung der Abweichungen erfolgt demnach nur in diesen Bereichen (Abb. 7 (b)).

Darüber hinaus wird die Auswirkung der Wasseroberflächenmodellierung auf die Genauigkeit der Gewässerbodenpunkte untersucht. Die Simulation erlaubt die Prognose von wellen-induzierten Koordinatenverschiebungen am Gewässerboden. Dazu wird die Refraktion eines infinitesimal schmalen Laserpulses an den verschiedenen Wasseroberflächenmodellen simuliert. Der Vergleich der resultierenden Gewässerbodenpunktkoordinaten erlaubt eine Analyse der Auswirkung der Wasseroberflächenmodellierung auf den Lage- und Höhenversatz am Gewässerboden.

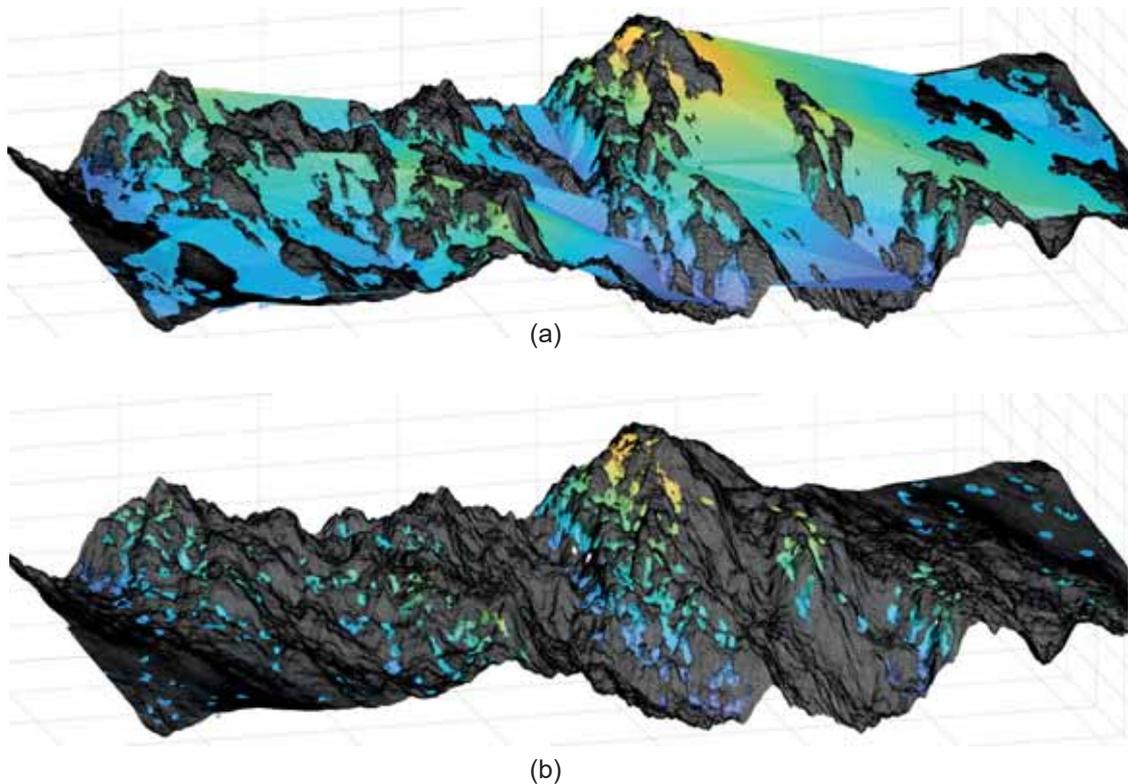


Abb. 7: Simulierte Wasseroberfläche (grau schattiert) und farbcodiertes Wasseroberflächenmodell in überhöhter Darstellung, (a) gesamtes Wasseroberflächenmodell, (b) Wasseroberflächenmodell in den relevanten Bereichen der Laserfootprints

## 5 Ergebnisse

### 5.1 Wasseroberflächenmodellierung

Abbildung 8 zeigt am Beispiel der Freiformfläche die Abweichungen zwischen Referenzwasseroberfläche und Wasseroberflächenmodell im Bereich der Laserfootprints. Ein Großteil der Abweichungen variiert zwischen  $-20$  mm und  $20$  mm. Wie zu erwarten, treten die größten Werte am Rand der Bereiche ohne Wasseroberflächeninformationen sowie in Bereichen mit komplexerem Wellenmuster auf.

In Tabelle 1 sind die Ergebnisse für horizontale und lokal geneigte Wasseroberflächenelemente, parametrisches Modell und Freiformflächen zusammengefasst. Die RMS-Werte (root mean square) deuten zum einen auf eine geringe Differenz zwischen den Wasseroberflächenmodellen und der Referenzwasseroberfläche und zum anderen auf einen geringfügigen Unterschied zwischen den Modellen hin. Die Analyse der minimalen und maximalen Abweichungen zeigt jedoch große Diskrepanzen zwischen den Modellierungsansätzen. Für das parametrische Modell variieren die minimale Abweichung und maximale Abweichung zwischen  $-150$  mm und  $279$  mm. Die Werte für die Wasseroberflächenmodelle mit horizontalen bzw. lokal geneigten Elementen und die Werte für die Freiformfläche sind in ihrer Größenordnung vergleichbar und mit  $-107$  mm bis  $110$  mm deutlich geringer. Die größeren Abweichungen bei der Wasseroberflächenmodellierung mit dem

parametrischen Modell deuten auf ein lokales Ausschwingen der Fourier-Repräsentation am Übergang zu Bereichen mit wenig Wasseroberflächeninformation hin. Im konkreten Fall wird die Referenzwasseroberfläche durch die Freiformfläche am besten approximiert.

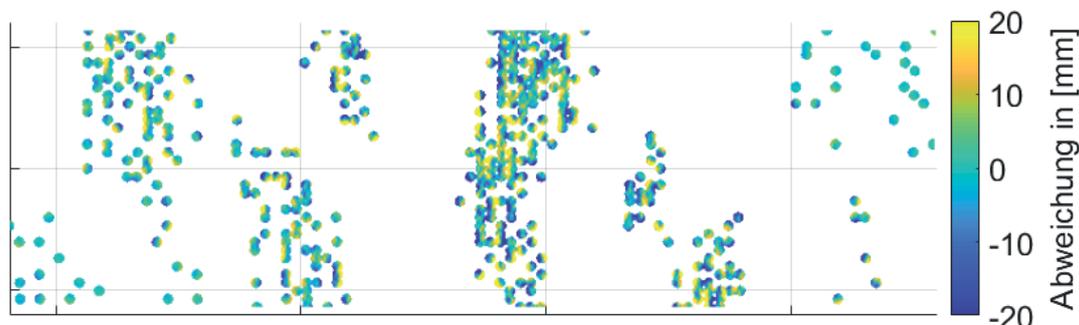


Abb. 8: Abweichung zwischen Wasseroberflächenmodell nach Tessendorf (2001) und Freiformfläche im Bereich der gescannten Wasseroberflächenpunkte

Tab. 1: Abweichungen zwischen Referenz und Wasseroberflächenmodell in mm

	Horizontale Elemente	Lokal geneigte Elemente	Parametrisches Modell	Freiformflächen
RMS	21	14	16	12
min	-104	-107	-150	-74
max	110	87	279	101

## 5.2 Gewässerbodenpunkte

Die Auswirkung der Wasseroberflächenmodellierung auf die Genauigkeit der Gewässerbodenpunkte wurde für alle Wasseroberflächenmodelle untersucht. Die Ergebnisse sind in Tabelle 2 zusammengefasst. Der Koordinatenversatz am Gewässerboden setzt sich aus den lateralen Komponenten  $dX$  und  $dY$  sowie der Tiefenkomponente  $dZ$  zusammen. Da die Auswirkungen der Wellen auf die Refraktion linear mit der Wassertiefe zunehmen, werden alle Ergebnisse in Prozent der Wassertiefe angegeben.

Für die Modellierung der Refraktion mit horizontalen Wasseroberflächenelementen ergibt sich ein RMS-Wert von 3,9% für die X-Komponente und 2,0% für die Y-Komponente. Bei einer Wassertiefe von 5 m entspricht das 19,5 cm für die X-Komponente und 10 cm für die Y-Komponente. Die Modellierung der Refraktion mit lokal geneigten Wasseroberflächenelementen resultiert in einem RMS-Wert von 2,3% für  $dX$  (11,5 cm bei 5 m Wassertiefe) bzw. 1,5% für  $dY$  (7,5 cm bei 5 m Wassertiefe). Wird das parametrische Wasseroberflächenmodell für die Refraktionskorrektur verwendet, betragen die RMS-Werte für die X- bzw. Y-Komponente 2,9% (14,5 cm bei 5 m Wassertiefe) bzw. 2,4% (12 cm bei 5 m Wassertiefe). Bei der Refraktionskorrektur mit der Freiformfläche ergibt sich ein RMS-Wert von 1,7% (8,5 cm bei 5 m Wassertiefe) für die X-Komponente und 1,1% (5,5 cm bei 5 m Wassertiefe) für die Y-Komponente. Die Werte zeigen, dass sich der laterale Koordinatenversatz mit steigender Komplexität der Wasseroberflächenrepräsentation verringert, wobei auch die minimalen und maximalen Abweichungen (Tabelle 2) diese Tendenz bestätigen. Eine Ausnahme bildet das parametrische Modell, welches die Wasseroberfläche im konkreten Fall nur ungenügend repräsentiert.

Tab. 2: Koordinatenversatz am Gewässerboden in Prozent der Wassertiefe

	dX			dY			dZ		
	min	max	RMS	min	max	RMS	min	max	RMS
Horizontale Wasseroberflächenelemente	-6,7	12,3	3,9	-4,5	6,4	2,0	-3,2	2,2	1,0
Lokal geneigte Wasseroberflächenelemente	-8,4	8,2	2,3	-7,0	5,1	1,5	-2,5	2,5	0,7
Parametrisches Modell	-11,4	10,5	2,6	-7,4	15,5	2,4	-3,0	3,2	0,8
Freiformflächen	-6,8	4,3	1,7	-3,4	0,1	1,1	-2,0	2,2	0,6

Die Tiefenkomponente dZ fällt grundsätzlich kleiner aus als die beiden Lagekomponenten (RICHTER et al. 2019). Der RMS-Wert variiert je nach Wasseroberflächenmodell zwischen 0,6 % (Freiformfläche) und 1,0 % (horizontale Wasseroberflächenelemente) (3 cm bzw. 5 cm). Die Zahlenwerte beziehen sich auf einen ebenen Gewässerboden. Bei bewegtem Gelände kann der Einfluss auf die Z-Komponente aufgrund von Strahldivergenz und Geländerepräsentationsfehler deutlich höher ausfallen.

Der absolute Koordinatenversatz am Gewässerboden wird in den Abbildungen 9 und 10 beispielhaft für die Refraktionskorrektur an lokal geneigten Wasseroberflächenelementen sowie an der Freiformfläche visualisiert. Der Koordinatenversatz wird jeweils getrennt nach den beiden Lagekomponenten und der Tiefenkomponente dargestellt. Jeder Punkt repräsentiert einen simulierten Wasseroberflächenpunkt. Die Farbcodierung entspricht dem Koordinatenversatz am Gewässerboden. Die simulierte Wasseroberfläche beinhaltet sowohl vergleichsweise glatte Wasserflächen (links und rechts) als auch einen Wellenberg (Mitte) mit einer Amplitude von 4 m (vgl. Abb. 2). Die Punktdichte variiert zwischen 0 Punkten/m<sup>2</sup> und 6 Punkten/m<sup>2</sup>. Der Farbverlauf in Abbildung 9 zeigt deutlich die Auswirkung der Wasseroberflächenmodellierung auf die resultierenden Bodenpunktkoordinaten. Der Einfluss auf die Lagekomponenten ist dabei größer als der Einfluss auf die Tiefenkomponente. In Bereichen mit ausreichend Wasseroberflächeninformationen liegt der Koordinatenversatz in der Lage zwischen  $\pm 3$  mm. Außerhalb dieser Bereiche ist der Koordinatenversatz vom lokalen Wellenmuster abhängig und nimmt Werte von bis zu  $\pm 12$  mm an. Die größten Abweichungen treten am Wellenberg auf, dessen Geometrie durch das Wasseroberflächenmodell nur unzureichend repräsentiert wird.

Abbildung 10 zeigt den absoluten Koordinatenversatz am Gewässerboden für die Refraktionskorrektur an der Freiformfläche. Die Werte variieren zwischen  $\pm 6$  mm für die Lagekomponente und  $\pm 2$  mm für die Tiefenkomponente. Der Vergleich zwischen beiden Wasseroberflächenmodellen zeigt, dass der Koordinatenversatz am Gewässerboden durch Verwendung des komplexeren Wasseroberflächenmodells deutlich verringert wird.

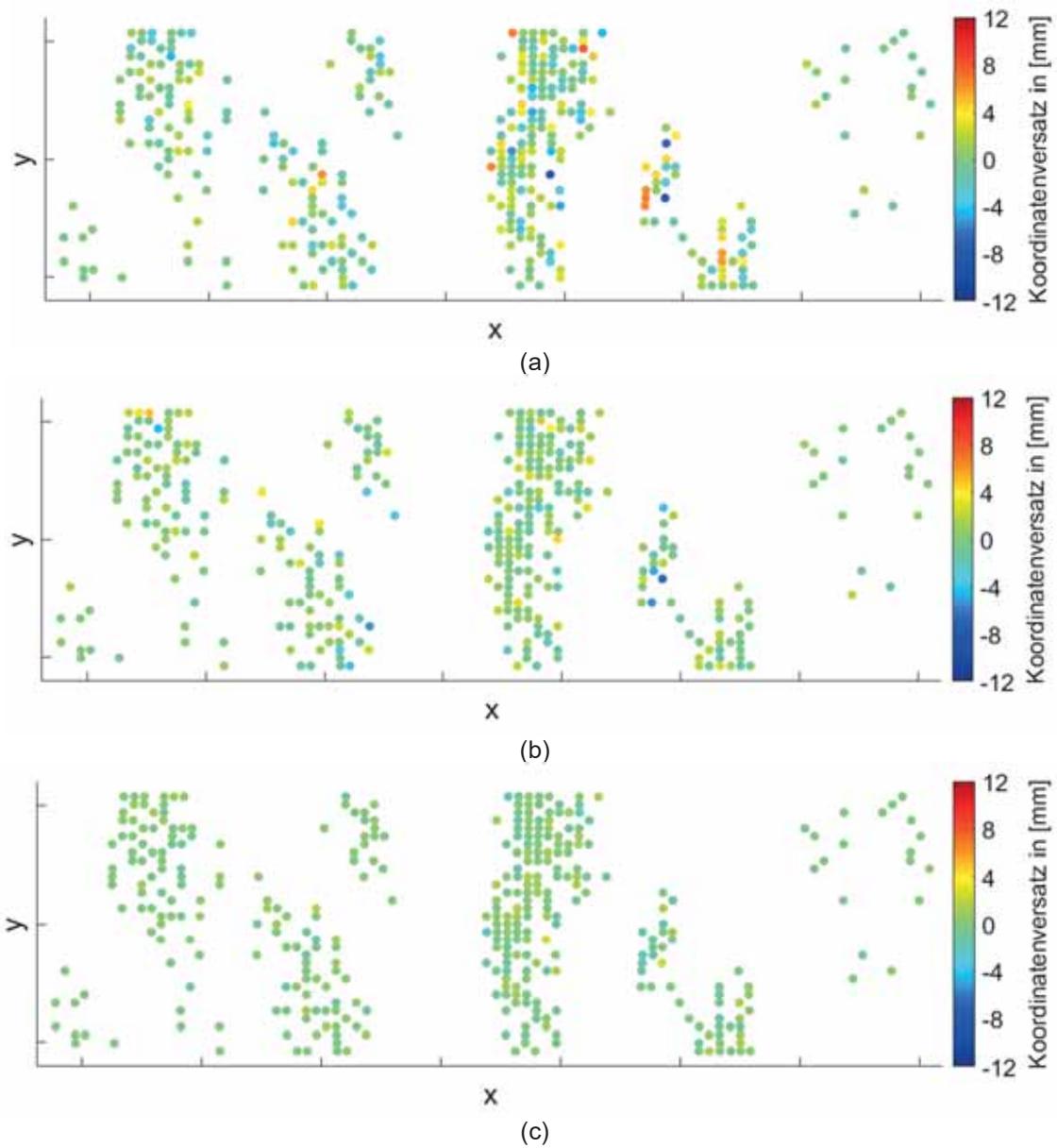


Abb. 9: Koordinatenversatz am Gewässerboden in x-Richtung (a), y-Richtung (b) und z-Richtung (c) für lokal geneigte Wasseroberflächenelemente

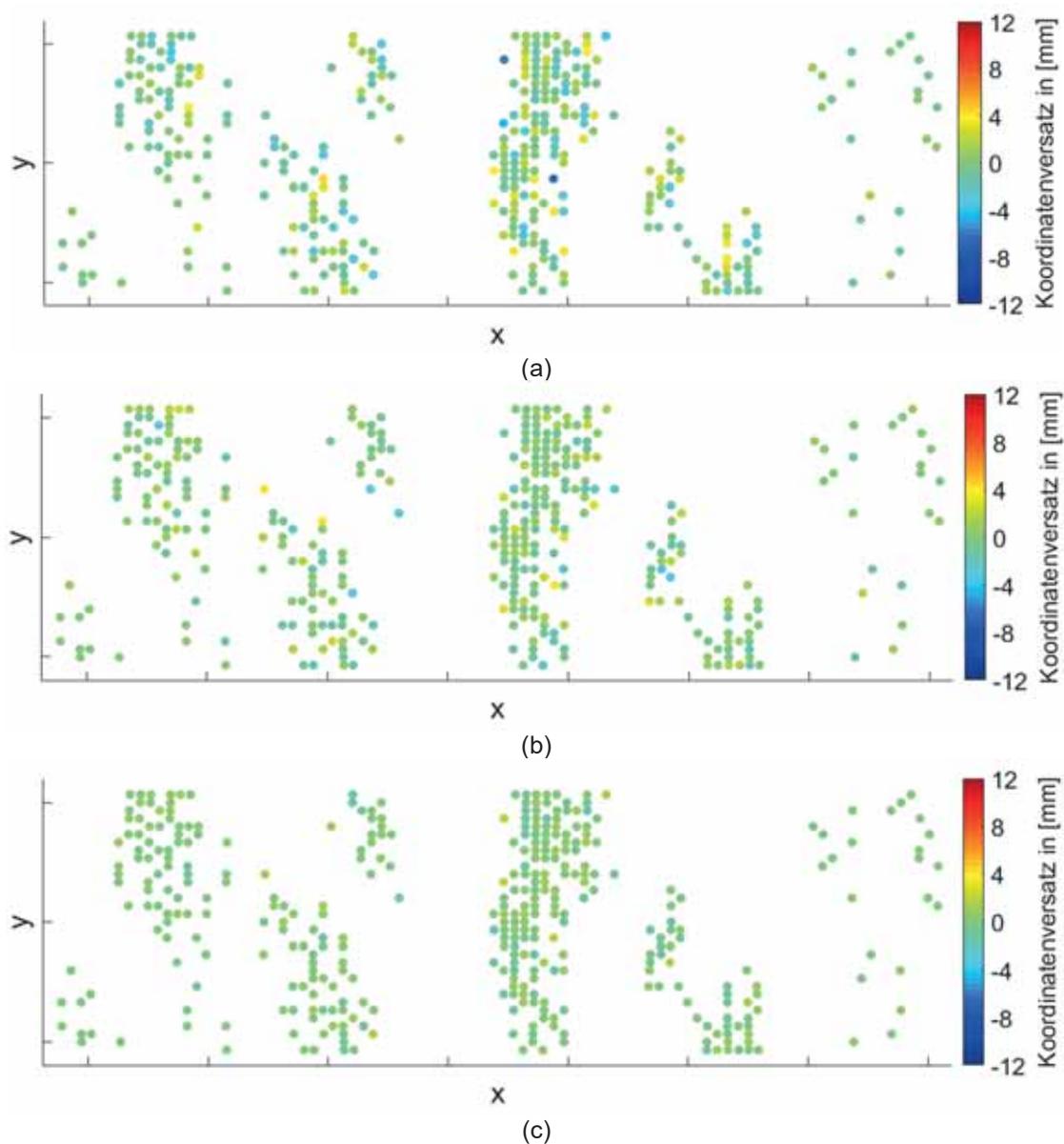


Abb. 10: Koordinatenversatz am Gewässerboden in x-Richtung (a), y-Richtung (b) und z-Richtung (c) für Freiformfläche

## 6 Zusammenfassung und Ausblick

Der Beitrag analysiert die Auswirkung der Wasseroberflächenmodellierung auf die Genauigkeit der Gewässerbodenpunkte in der Laserbathymetrie. Neben den konventionellen Methoden zur Modellierung der Wasseroberfläche werden zwei alternative Ansätze vorgestellt, welche die lokale Krümmung der Wasseroberfläche berücksichtigen. Die Evaluierung der Modellierungsansätze erfolgt mit Hilfe eines Laserbathymetrie-Simulators. Die Abweichungen zwischen modellierter Wasseroberfläche und tatsächlicher Wasseroberfläche beeinflussen die Refraktionskorrektur und resultieren in einem Lage- und Höhenversatz am Gewässerboden. Je besser die wahre

Wasseroberfläche durch das Modell angenähert wird, desto geringer fällt der Koordinatenversatz am Gewässerboden aus. Die Analyse der Simulationsergebnisse zeigt, dass sich die Wasseroberflächenmodellierung mit einer Freiformflächen am besten für die Approximation einer gekrümmten Wasseroberflächengeometrie eignet. Während der Koordinatenversatz für die konventionelle Wasseroberflächenmodellierung mit horizontalen Wasseroberflächenelementen in Lagerichtung zwischen 2 % und 4 % und in Tiefenrichtung 1 % beträgt, verringern sich die Werte für die Wasseroberflächenmodellierung mit Freiformflächen auf 1 % für die Lagekomponente und 0,6 % für die Tiefenkomponente.

Die Erkenntnisse werden in zukünftigen Arbeiten Eingang in die differentielle Modellierung der Refraktion eines divergenten Laserpulses mit endlichem Durchmesser an einer zeitlich veränderlichen wellen-induzierten Wasseroberfläche finden und damit zur Steigerung des Genauigkeitspotentials moderner Laserbathymetriesysteme beitragen.

## 7 Literaturverzeichnis

- CSANYI, N. & TOTH, C.K., 2005: Some aspects of using fourier analysis to support surface modelling. *Global Priorities in Land Remote Sensing*.
- GRIMM-PITZINGER, A. & RUDIG, S., 2005: Freiformflächen zur Modellierung von Deformationsmessungen. *zfv – Zeitschrift für Geodäsie, Geoinformation und Landmanagement (ZfV)* **130**(3), 180-183.
- RICHTER, K., MADER, D., WESTFELD, P., & MAAS, H.-G., 2018. Numerical simulation and experimental validation of wave pattern induced coordinate errors in airborne Lidar bathymetry. *The International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* **42**(2).
- RICHTER, K., MADER, D., WESTFELD, P., & MAAS, H.-G., 2019. Analyse der Auswirkung von Wellen auf konventionelle Refraktionskorrekturmethode in der Laserbathymetrie. Dreiländertagung der DGPF, der OVG und der SGPF in Wien, Österreich – Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V., Band **28**, 334-346.
- TESSENDORF, J., 2001: Simulating ocean waters. *ACM SIGGraph Course Notes*.
- ÜLLRICH, A. & PFENNIGBAUER, M., 2011: Laser-Hydrographieverfahren. Patent WO 2011137465 A1. Riegler Laser Measurement Systems GmbH.
- WESTFELD, P., MAAS, H. G., RICHTER, K., & WEIß, R., 2017: Analysis and correction of ocean wave pattern induced systematic coordinate errors in airborne LiDAR bathymetry. *ISPRS Journal of Photogrammetry and Remote Sensing* **128**, 314-325.

# Bathymetrieableitung aus multispektralen Luftbildern über Convolutional Neural Networks

HANNES NÜBEL<sup>1</sup>, GOTTFRIED MANDLBURGER<sup>1</sup>, MICHAEL KÖLLE<sup>1</sup> & UWE SÖRGEL<sup>1</sup>

*Zusammenfassung: Zur Ableitung von Gewässertiefen werden aufgrund der einfachen Datenerfassung in den letzten Jahren vermehrt optische Ansätze basierend auf der Bildradiometrie eingesetzt. In flachen Bereichen kann die Tiefe dabei durch die Modellierung der Signaldämpfung in verschiedenen Bändern ermittelt werden. Im Rahmen des im Beitrag vorgestellten Ansatzes soll untersucht werden, inwieweit ein Convolutional Neural Network in der Lage ist, Wassertiefen aus multispektralen Luftbildern zu schätzen. Um auf die in den Bildpixel abgebildeten schrägen Wasserdistanzen zu trainieren, werden als Basis nicht das Orthophoto, sondern die originalen weitwinkeligen Luftbilder herangezogen. Der verwendete Datensatz enthält neben RGB-Bildern auch synchron erfasste panchromatische Bilder mit einem Coastal Blue Filter, dessen Nutzen in der CNN-basierten Regression analysiert wird. Die Auswertung zeigt, dass Trends für unterschiedliche Tiefen und Bodenbedeckungen erkannt werden können und eine Standardabweichung von 0.3-0.4 m erreicht werden kann. Limitiert wird der Ansatz durch direkte Reflexion von Sonnenlicht (sun glint) und durch Schattenbereiche. Die Einbeziehung des Coastal Blue Bandes bringt einen Mehrwert in Bezug auf die Verteilung der Tiefen des Testgebietes.*

## 1 Einführung

### 1.1 Motivation

Die Vermessung der Erdoberfläche mittels Photogrammetrie ist ein etabliertes Verfahren. Koordinaten von Objektpunkten können bei bekannter Orientierung der Kameras mittels Vorwärtschnitt berechnet werden, wenn der jeweilige Punkt in zwei oder mehr Bildern beobachtet wird. Die Anwendung dieses Verfahrens in Wasserbereichen ist jedoch komplexer (MAAS 2015). Dennoch ist es insbesondere in Flachwassergebieten notwendig, Wassertiefen zu kartieren, zum Beispiel für sichere Schiffsnavigation oder bei der Bestimmung des Volumens eines Sees, welcher für Brandbekämpfung benötigt wird.

Die Komplexität bei Mehrmedienphotogrammetrie besteht darin, dass die Messung identischer Punkte aufgrund der spiegelnden und dynamischen Beschaffenheit der Wasseroberfläche erschwert wird. Darüber hinaus kommt es beim Übergang des Bildstrahls zwischen zwei Medien zur Brechung an der Wasseroberfläche. Für die Erzeugung eines Orthophotos bedeutet dies, dass jedes Pixel in jedem Bild einen individuellen gebrochenen Strahl aufweist, der der Wasseroberfläche entsprechend auch lokale Dynamiken aufweist.

Um die Bodenpunkte jedes Pixels zu detektieren, müssen somit die Strahlen mit der durch die Orientierung des Bildes gegebenen Richtung, unter Berücksichtigung der Brechung an der Wasseroberfläche, verfolgt werden. Des Weiteren gilt, dass selbst wenn die Richtung jedes Strahls

---

<sup>1</sup> Universität Stuttgart, Institut für Photogrammetrie, Geschwister-Scholl Str. 24D, D-70174 Stuttgart, E-Mail: st148625@stud.uni-stuttgart.de, [gottfried.mandlburger, michael.koelle, uwe.soergel]@ifp.uni-stuttgart.de

bekannt ist, genügend identische Punkte gefunden werden müssen, um deren Koordinaten mit Hilfe der sich schneidenden Strahlen aus den Bildern zu berechnen. Dies ist ebenso anspruchsvoll, da Gewässerböden typischerweise homogen sind und darüber hinaus Signale im Wasser gedämpft werden. Dieser Effekt wird zudem durch Reflexion und andere Faktoren verstärkt, da dieselben Punkte unterschiedlich erscheinen, wenn sie aus verschiedenen Perspektiven betrachtet werden. Aufgrund der unterschiedlichen Stärke der Lichtabsorption in der Wassersäule für verschiedene Spektralbänder ist es auch möglich, ein lineares oder höher dimensionales Regressionsmodell aufzustellen, welches das Verhältnis von Radiometrie zu Tiefe approximiert. Sobald jedoch der Bereich verschiedene Arten von Vegetation auf dem Grund des Gewässers enthält, ist ein komplexeres Regressionsmodell erforderlich. Darüber hinaus wird die spektrale Tiefenschätzung üblicherweise auf der Grundlage von Orthophotos oder sehr schmalwinkeligen Satellitenbildern durchgeführt. Die Vernachlässigung der Strahlbrechung an der Wasseroberfläche führt dabei zu geometrischen Fehlern. Zusätzlich beschreiben nur Pixelwerte aus dem Bildmittelpunkt (Nadirrichtung) direkt die Wassertiefe, während Pixel vom Bildrand die schräge Wasserstrecke anzeigen. Jedes Pixel eines Luftbildes wiederum speichert radiometrische Informationen, die sich auf die potenziell schräge Unterwasserdistanz des jeweiligen Bildstrahls beziehen. Insbesondere bei Weitwinkelobjektiven ist es daher sinnvoll, die bathymetrische Schätzung nicht auf dem Orthophoto, sondern auf den (orientierten) Bildern durchzuführen.

Um den linearen Regressionsansatz zu erweitern, kann ein Convolutional Neural Network (CNN) verwendet werden, um Variationen des Bodenreflexionsgrades miteinzubeziehen. Die pixelweise Tiefenschätzung auf Basis der orientierten Luftbilder erfordert die schrägen Wasserdistanzen für die Bildpunkte zum Training. Diese Informationen können zum Beispiel aus bathymetrischen LiDAR (Light Detection And Ranging) Daten abgeleitet werden, insbesondere wenn diese gleichzeitig mit den Bildern erfasst wurden.

Des Weiteren bringen CNNs den Vorteil, dass räumliche Kontextinformationen berücksichtigt werden. Die Zuverlässigkeit des Netzes wird dadurch erhöht, da Nähe oft ähnliche Tiefen impliziert.

In dieser Arbeit wird der Ansatz des Trainings eines CNN zur Prädiktion der schrägen Distanzen von Bildstrahlen in einem Gewässer untersucht. Neben der Qualitätsbewertung wird auch diskutiert, inwieweit der Coastal Blue Kanal Einfluss auf das Netz hat.

Im Abschnitt 2 wird der zugrunde liegende Datensatz und seine Aufbereitung vorgestellt. Die Methoden, einschließlich der Vorverarbeitung der Referenzdaten und des Trainings des Netzes, werden dann im Abschnitt 3 beschrieben. Der darauffolgende Abschnitt 4 präsentiert und diskutiert die Ergebnisse des trainierten Netzes. Abschnitt 5 fasst schließlich die Arbeit zusammen und befasst sich zudem mit den möglichen Themen für weitergehende Forschungen.

## **1.2 Verwandte Arbeiten**

Ein etabliertes Verfahren zur Unterwasserkartierung ist SoNAR (Sound Navigation and Ranging) (MASNADI-SHIRAZI et al. 1992). Da die Instrumente unter einem Schiff im Wasser montiert sind, gibt es im Vergleich zu Sensoren oberhalb der Wasseroberfläche keinen Übergang zwischen unterschiedlichen Medien. Die Dynamik der Wasseroberfläche beeinflusst bei der SoNAR Technik daher lediglich die Sensororientierung. Gegenüber optischen Verfahren stellt SoNAR geringere

Ansprüche an die Klarheit des Wassers und ist für die Vermessung größerer Messtiefen jenseits der Sichttiefen geeignet.

Eine weitere zunehmend angewandte Technik für die Ableitung von Wassertiefen im Flachwasserbereich ist Flugzug-getragene Laserbathymetrie (IRISH & WHITE 1998). Anstelle von Schallwellen wie bei SoNAR werden Laserpulse ausgesendet, deren Reflexion (Echos) gemessen werden. Durch den Übergang des Laserpulses von der Atmosphäre in das Wasser muss die Änderung der Ausbreitungsgeschwindigkeit und der Strahlrichtung berücksichtigt werden (Refraktionskorrektur). Ein Vergleich mit dem SoNAR-Ansatz wurde u.a. von COSTA et al. (2009) durchgeführt. Der Hauptvorteil ist die erhöhte Effizienz vor allem für große Flachwassergebiete. Darüber hinaus ist es auch möglich, Daten in abgelegenen Gebieten sowie in Randgebieten eines Gewässers zu erfassen, die mit SoNAR nicht zugänglich sind.

Für topografische Anwendungen ist die Stereophotogrammetrie eine gängige Strategie. Identische Punkte werden in mehreren Bildern erfasst und ihre Position durch Bündelausgleichung berechnet. Die Herausforderungen, die sich für diesen Prozess in bathymetrischen Anwendungen ergeben, werden von MULSOW et al. (2019) im Vergleich zur Datenerfassung mit einem luftgestützten Laserscanner diskutiert. Die Kosten der Datenerfassung können bei annähernd gleicher Genauigkeit reduziert werden. Grundlage ist jedoch das Auffinden identischer Punkte, welche durch Textur bestimmt werden. Insbesondere in tiefen oder homogenen Bereichen ist dies nicht gegeben und es treten daher größere Abweichungen auf.

Beim Ansatz der spektralen Bathymetrie dagegen, wird ein mathematisches oder physikalisches Modell erstellt, das eine Verbindung zwischen Reflexionsgrad und Wassertiefe herstellt. Somit ist es also möglich, die Wassertiefe für jedes Pixel zu schätzen. Eine häufige Annahme ist, dass sich die Reflexion vom Grund negativ exponentiell verhält, bezogen auf die Wassertiefe (LYZENGA et al. 2006). Darüber hinaus muss ein Offset einbezogen werden, der den Reflexionsgrad für eine unendliche Wassertiefe modelliert.

Spektrale Tiefenbestimmung kann, abhängig von der Lichtdämpfung in der Wassersäule, für seichte Gewässer angewendet werden. Da die Reflexion für verschiedene Bodendeckungen stark variiert, kann es sinnvoll sein, mehrere Spektralbänder (LEGLEITER et al. 2009) einzubeziehen, da die Rückstrahlung für bestimmte Bodentypen in angrenzenden Bändern ähnlich ist. Insbesondere logarithmische Verhältnisse zwischen den Bändern eignen sich für die Tiefenschätzung. Auch eine Kombination aus dem roten Kanal, der am stärksten von der Signalabsorption beeinflusst wird, und einem besser durchdringenden Spektralband weist einen hohen Informationsgehalt auf (MANDLBURGER et al. 2018).

Von WANG et al. (2019) wurde ein bathymetrisches Modell entwickelt, welches sowohl spektrale als auch räumliche Informationen mit Hilfe eines neuronalen Netzwerks einbezieht. Bei diesem Ansatz wird die räumliche Information durch die Koordinaten X und Y gegeben. Zusammen mit vier multispektralen Bandverhältnissen bilden sie einen Merkmalsvektor, welcher den ersten Layer des Netzes repräsentiert. Das neuronale Netz besteht nach dem ersten Layer aus drei versteckten (hidden) Layern und einem Output Layer, die zusammen ein Multi Layer Perceptron (MLP) bilden.

Der hier angewandte Ansatz unterscheidet sich vor allem in zwei Punkten. Erstens werden die räumlichen Informationen mit Hilfe eines CNN aufgenommen, das Gewichte für kleine Faltungskerne lernt, die einen bestimmten Bereich um ein Pixel herum abdecken und somit nicht abhängig

von der tatsächlichen X- und Y-Koordinate eines Punktes sind. Aufgrund dieser allgemeineren Annahme kann das Netz auch auf andere Bereiche übertragen werden, da eine semantische Segmentierung eines gesamten Bildes durchgeführt wird, anstatt isolierte Tiefen von Einzelpunkten anhand bestimmter Merkmale abzuleiten. Zweitens ist es dem Netz durch die größere Anzahl an Layern möglich, komplexere Features zu extrahieren.

## 2 Datensatz

Die Erprobung des Ansatzes erfolgte auf Basis einer Befliegung des Autobahnsees (Augsburg) im April 2018. Dabei wurden zeitgleich Luftbilder und Laserbathymetriedaten erfasst. Der Autobahnsee ist ein 5 m tiefer Baggersee mit einer Ausdehnung von ca. 750 m × 500 m und beinhaltet eine kleinere Insel. Der Untergrund des Sees besteht aus offenen Schotterflächen sowie unterschiedlichen Vegetationsfeldern (Abb. 1).

### 2.1 Bilddaten

Für die Bilddatenerfassung kamen zwei IGI DigiCAM 100 Kameras zum Einsatz, die auf Phase One iXU-RS 1000 Kameras mit je 11608 × 8708 Pixel basieren. Während der Sensor der RGB Kamera mit einem Bayer Pattern ausgestattet ist, wurde bei der Coastal Blue Kamera ein panchromatischen Sensor und vorgesetztem Filter für die Coastal Blue Wellenlänge verwendet (MANDLBURGER et al. 2018).



Abb. 1: Orthophoto Autobahnsee mit verschiedenen Grundbedeckungen

Um die Informationen aus beiden Bildern nutzen zu können, sollten diese von derselben Position mit derselben Orientierung aufgenommen werden. Aus praktischen Gründen mussten die Kameras jedoch nebeneinander montiert werden. Die Transformation des zeitgleich aufgenommenen Coastal Blue Bildes in den Bezugsrahmen des korrespondierenden RGB-Bildes erfolgte mittels Homographie mit der Software MATLAB (2018). Da die Höhenvariation des Geländes im Bezug zur Flughöhe vergleichsweise gering ist, sind die Restfehler vernachlässigbar. Durch die Transformation entstehen im neuen Coastal Blue-Bild leere Pixel im Überlappungsbereich mit dem RGB-Bild. Um sicherzustellen, dass ein Bild in jedem Pixel für alle Bänder Werte besitzt, wird für die zusammengeführten Bilder im RGB-Bilddatum ein kleiner Rahmen an den Grenzen jedes Bildes definiert. Alle Pixel in diesem Randbereich werden später ausmaskiert.

## 2.2 LIDAR

Abgesehen von den beiden Kameras integriert das eingesetzte hybride Sensorsystem auch einen RIEGL VQ-880-G topo-bathymetrischen Laserscanner (RIEGL 2019). Aus der damit synchron aufgenommenen 3D Punktwolke, konnte das Wasserflächenmodell und das Geländemodell extrahiert werden. In Abb. 2 ist das Geländemodell für das beobachtete Gebiet dargestellt. Es wird deutlich, dass es am Boden des Sees komplexe Strukturen gibt, die einerseits auf topographische Details und andererseits auf nicht durchdringbare Unterwasser-Vegetation zurückzuführen sind. Oberflächen- und Geländemodell bilden die Basis für die Extraktion der Referenzdaten, welche in diesem Fall die schrägen Distanzen der Bildstrahlen im Wasser sind. Der Scanner operiert mit grüner Laserstrahlung (Wellenlänge: 532 nm) und ist für die Flachwasserkartierung konzipiert. An Land beträgt die mittlere Punktdichte beträgt etwa 40 Punkte/m<sup>2</sup>. In den tieferen Wasserbereichen ist die Punktdichte aufgrund der Dämpfung des Lasers geringer, allerdings immer noch ausreichend, um ein DGM mit 0.5 m Rasterweite zu berechnen.

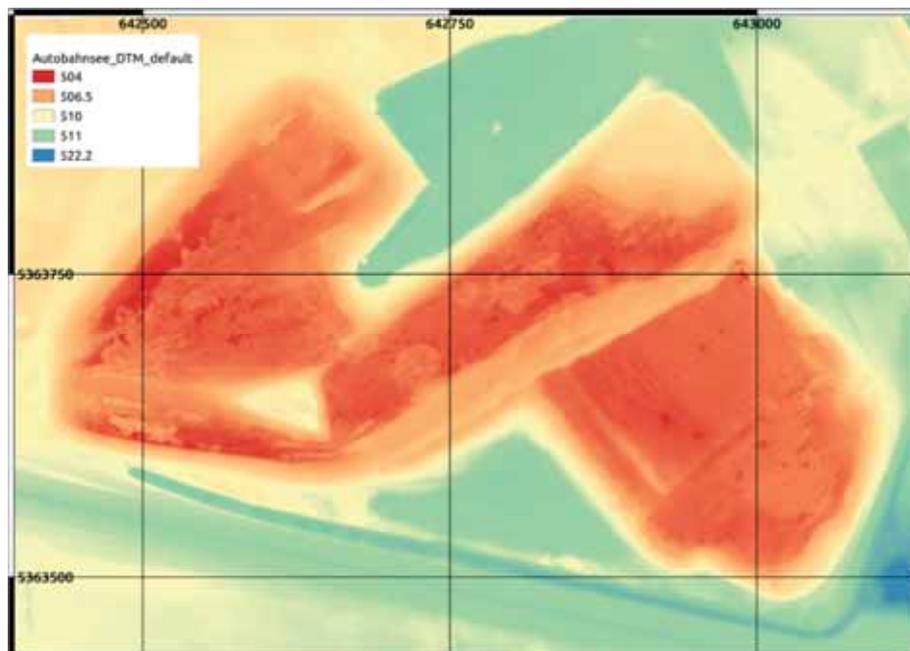


Abb. 2: Digitales Geländemodell des Autobahnsees aus topo-bathymetrischem LiDAR.

### 3 Methodik

#### 3.1 Vorverarbeitung von Referenzdaten

Im folgenden Abschnitt wird die angewandte Methodik zur Ableitung der Referenzdaten diskutiert, mit denen das angewandte CNN trainiert wurde. Als Referenzdaten fungieren dabei die jeweiligen Schrägdistanzen der Bildstrahlen jedes Pixels im Wasser. Um den Strahlengang von der Kamera zum entsprechenden Bodenpunkt unter Berücksichtigung der Brechung an der Wasseroberfläche zu verfolgen, wird die Orientierungen der Kamera, ein digitales Wasseroberflächenmodell (DWM), sowie ein Geländemodell (DGM) benötigt. Das DWM wurde aus den ersten Echos des Laserscanners geschätzt, während die letzten Echos die Grundlage für die Filterung der Bodenpunkte und schließlich der DGM-Berechnung sind. Abb. 3 veranschaulicht die Zusammenhänge in Form einer Prinzipskizze.

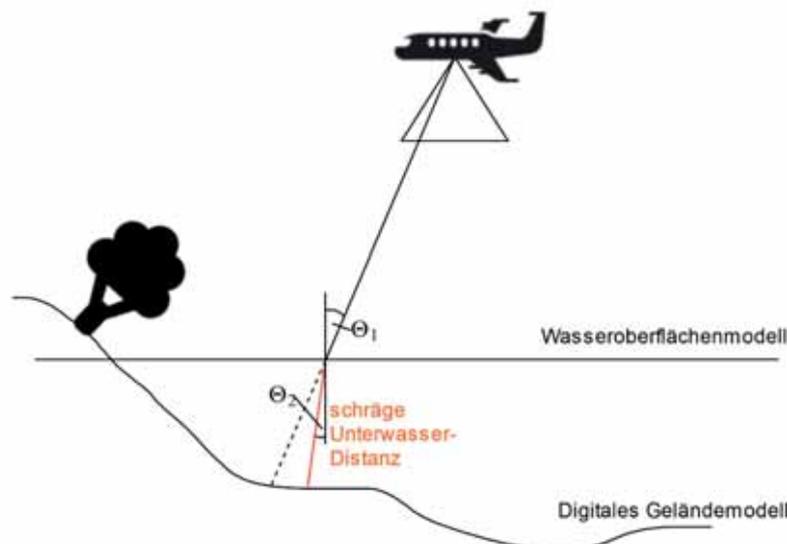


Abb. 3: Refraktion des Bildstrahls an der Wasseroberfläche.

Um die schrägen Distanzen zu erhalten, können die den einzelnen Pixel entsprechenden Bildstrahlen im lokalen Kamerakoordinatensystem aus der inneren Orientierung der Kamera berechnet werden. Über die Elemente der äußeren Orientierung wird der Bildstrahl anschließend in ein globales Koordinatensystem transformiert (KRAUS 1996). Jeder Bildstrahl wird anschließend mit dem DWM verschnitten, der Strahlbrechung gemäß dem Brechungsgesetz von Snellius unterworfen und der gebrochene Bildstrahl anschließend mit dem DGM verschnitten. Die Distanz zwischen den Strahlschnittpunkten mit dem DWM und dem DGM entspricht der im jeweiligen Pixel abgebildeten schrägen Wasserdistanz. Für die Implementierung des beschriebenen Workflows kam die Laserscanning Software OPALS (PFEIFER et al. 2014) zum Einsatz. Die Berechnung der Bildstrahlen im Objektkoordinatensystem erfolgte in Python basierend auf den Pixelkoordinaten und den Kameraorientierungen. Weitere Details zur Methodik sind in NÜBEL (2019) beschrieben.

## 3.2 Deep Learning

### 3.2.1 Convolutional Neural Networks (CNNs)

CNNs sind eine spezifischere Art von neuronalen Netzwerken, mit denen insbesondere Bilder präzise und effizient verarbeitet werden können, indem sie von ihrer räumlichen Struktur profitieren (NIELSEN 2015). Dies wird realisiert, indem man in den einzelnen Layern Faltungen durchführt. Abhängig von der Größe des Kernels liefern kleine Bereiche von z. B.  $3 \times 3$  Pixel aus einem Layer Informationen für ein korrespondierendes Neuron in einem tieferen Layer. Effektive Feature extrahierende Faltungskerne können nicht nur an einer Position, sondern überall im Bild nützlich sein. Zur Reduktion wird deshalb nur ein Kernel für das gesamte Layer trainiert, anstatt verschiedene Kernel zu trainieren, die sich auf bestimmte Regionen im Bild beziehen. Anstatt also für jedes Neuron in jedem Layer Gewichte zu trainieren, müssen nur die Gewichte für ein Kernel trainiert werden. Dadurch lassen sich tiefere Architekturen mit mehr Layern realisieren, um High-Level-Features einzubeziehen, aber dennoch effizient in Bezug auf Trainingsdauer und den erforderlichen Trainingsdaten zu sein.

### 3.2.2 U-Net

Das U-Net (RONNEBERGER et al. 2015), welches als Grundlage für das angewandte Netz dient, ist ein CNN, welches für die biomedizinische Bildsegmentierung konzipiert wurde. Es handelt sich um ein Fully Convolutional Network, was bedeutet, dass die Größe des Ausgabebildes gleich der Eingabegröße ist. Dies ist im vorliegenden Kontext insofern erforderlich, da für jedes Pixel des Wasserbereiches ein Tiefenwert prädiiziert werden soll.

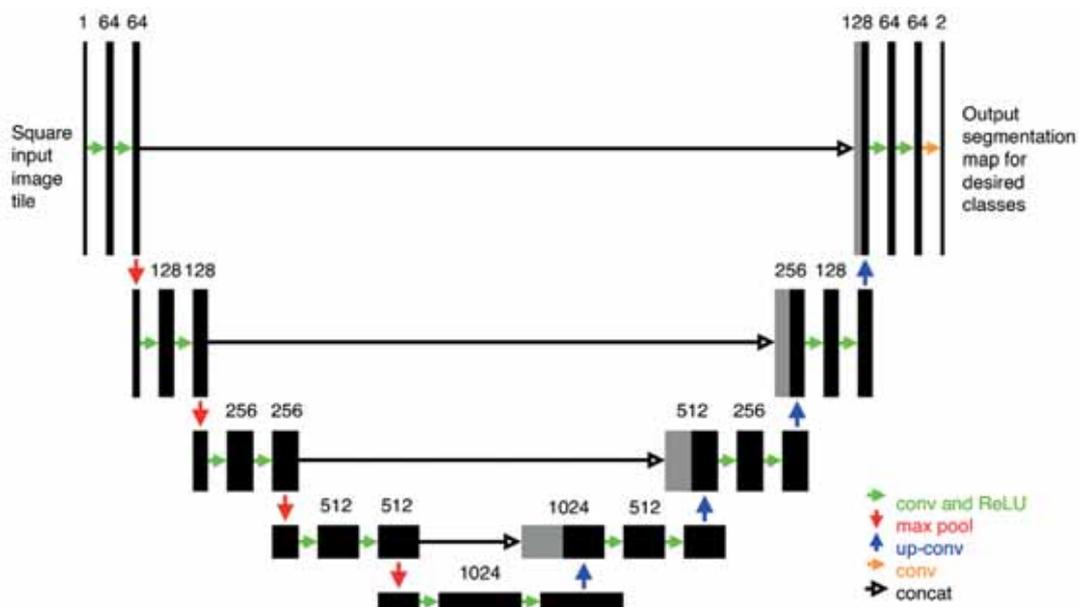


Abb. 4: U-Net-Architektur. Jede schwarze Box entspricht einer Multi-Kanal Feature Map. Die Anzahl der Kanäle ist auf der Oberseite der Box angegeben. Die grauen Boxen stellen kopierte Feature Maps dar. Die Pfeile kennzeichnen die verschiedenen Operationen (RONNEBERGER et al. 2015)

Der Name ergibt sich aus der Architektur des Netzes, da es einen komprimierenden und einen symmetrischen expandierenden Pfad aufweist (siehe Abb. 4). Es besteht aus Convolutional Layers, gefolgt von einer Rectified Linear Unit (ReLU) und Max-Pooling-Layern in den Kompressions- sowie Up-Convolution-Layern in den Expansionspfaden. Die ReLU-Aktivierungsfunktion ist definiert durch  $f(x) = \max(0, x)$  und wird häufig in modernen neuronalen Netzen verwendet. Im Komprimierungspfad wird bei jedem Pooling-Schritt die Dimension der jeweiligen Bildkachel halbiert. Nach jedem Schritt wird die Anzahl der mit den Convolutional Layers berechneten Feature-Maps verdoppelt. Die Verkleinerung der Bildkacheln erfolgt durch Max-Pooling, d.h. für die jeweiligen Werte des verschiebenden Kernels wird der Maximalwert ausgegeben (GOODFELLOW et al. 2016). Um die Eingabedimension wiederherzustellen, muss für die Kacheln anschließend ein Upsampling durchgeführt werden, was durch eine inverse Faltung geschieht (DUMOULIN & VISIN 2016). Anstatt ein interpolierendes Kernel für das Upsampling zu definieren, wird dieses ebenfalls erlernt und optimiert, während das Netz trainiert wird.

Die Ergebnisse im expandierenden Pfad des U-Net werden mit den gegenüberliegenden Layern aus dem Kompressionspfad verknüpft, um die räumlichen Informationen zu erhalten. Schließlich erfolgt im letzten Layer eine  $1 \times 1$  Faltung, die jeden Feature-Vektor auf die benötigte Anzahl von Klassen abbildet.

### 3.2.3 Anpassung der U-Net-Architektur

Die Implementierung des CNN erfolgt in Python unter Verwendung der Deep Learning Bibliothek Keras (CHOLLET 2018) mit Tensorflow Backend (ABADI et al. 2015) und das Netz wird auf einer Graphics Processing Unit (GPU) trainiert, da dies effizienter im Vergleich zu einer Central Processing Unit (CPU) ist.

Wie im Abschnitt 3.2.2 erwähnt, dient das U-Net (RONNEBERGER et al. 2015) als Grundlage für das CNN. Der Hauptunterschied besteht darin, dass das Netz nicht wie bei den meisten Ansätzen zur Segmentierung verwendet wird, sondern zur Anpassung eines Regressionsmodells. Anstatt also auf mehrere Klassen mit einer Ausgabe der jeweiligen Wahrscheinlichkeit für jedes Pixel zu trainieren, wird nur auf die schräge Wasserdistanz als Gleitkommazahl trainiert. Daher ist die Ausgabe des letzten Faltungslayers nur ein Pixel tief und anstelle einer Aktivierungsfunktion wie Sigmoid, mit der die Klassenwerte bestimmt werden können, wird erneut ReLU verwendet. Als Verlustfunktion wird der Root Mean Square Error (RMSE) zwischen der Referenz und den vorhergesagten Werten berechnet.

Um die Größe des Eingabebildes bei der Durchführung einer Faltung beizubehalten, wird Padding verwendet. Folglich hat das Ausgabebild auf der letzten Ebene die gleiche Größe wie die Eingabedaten. Wie in der ursprünglichen Architektur des U-Net werden auch hier Breite und Höhe in jedem Pooling-Schritt durch 2 dividiert, um die doppelte Anzahl an Filtern zu ermöglichen. Um Overfitting zu vermeiden, wird nach den Pooling-Schritten ein zusätzliches Dropout-Layer hinzugefügt, das während des Trainings zufällig einen Prozentsatz der Neuronen ignoriert.

Um einerseits die notwendige Speichermenge bezüglich des Arbeitsspeichers der GPU möglichst gering zu halten, andererseits aber auch nach der letzten Pooling-Schicht eine natürliche Anzahl von Pixeln in Breite und Höhe zu haben, werden Patches von  $480 \times 480$  Pixel genutzt. Diese Wahl wurde getroffen, da das Netz bei kleineren Patches Probleme haben könnte, die Eigenschaften in

Uferbereichen zu erlernen, während größere Patches dazu führen könnten, dass es stärker die Struktur des Sees erlernt, anstatt das Verhalten der Spektralbänder in Bezug auf die Tiefe. Die zufällig in den Bildern gewählten Patches werden angenommen, wenn sie einen bestimmten Prozentsatz an gültigen Pixeln (Wasserpixel) aufweisen. Dadurch wird sichergestellt, dass jeder Ausschnitt genügend Daten enthält und auch Uferbereiche berücksichtigt werden, sodass nicht nur das Zentrum des Sees trainiert wird.

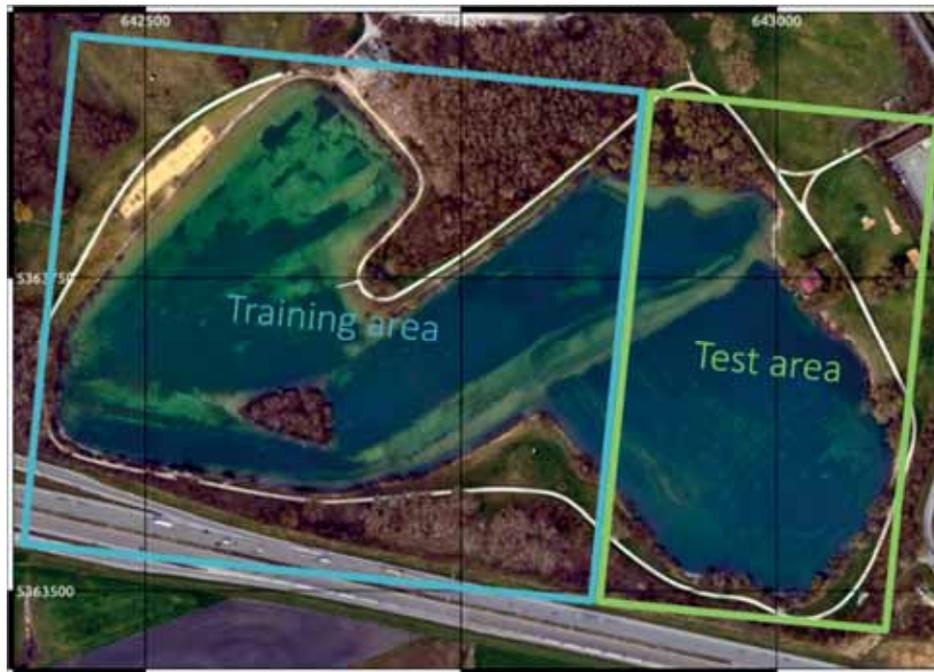


Abb. 5: Verteilung des in den Bildern beobachteten Trainings- und Testbereichs.

Zuletzt müssen die Bilder in Trainings- und Testbilder unterteilt werden. Die Trainingsbilder enthalten dabei auch einen Prozentsatz an Validierungsdaten, wogegen die Testbilder ausschließlich aus Bereichen stammen, die nicht zum Trainieren des Netzes verwendet wurden. Dazu wird die Seefläche in zwei Teile geteilt, die in Abb. 5 markiert sind. Um sicherzustellen, dass keine Teilbereiche bereits im Training enthalten waren, wurden die Bilder im Überlappungsbereich weder im Training noch zum Testen verwendet. Insgesamt wurden 41 Bilder in der Trainingsphase und 24 in der Testphase eingesetzt. Es ist jedoch zu beachten, dass die Datenmenge in jedem Bild von der Anzahl der Pixel abhängig ist, die das Gewässer abbilden. Die Strukturen in den ausgewählten Bereichen unterscheiden sich sehr stark, so dass beurteilt werden kann, ob das Netzwerk für den Trainingsbereich zum Overfitting neigt, oder ob es Eigenschaften lernt, die auch auf andere Gewässer übertragbar sind.

## 4 Ergebnisse und Diskussion

### 4.1 Angewandtes CNN für kombiniertes RGB- und Coastal Blue Band

Ein Indikator für die Qualität beim Training des Netzes ist der Loss-Plot, welcher sich hier auf den RMSE von Prädiktionen mit dem aktuellen Netz für jede Trainingsepoche bezieht (Abb. 6). Für

die Trainingsphase werden die verwendeten Bilder wiederum in zwei Teile gegliedert. Die eigentlichen Trainingsbilder werden dann zum Trainieren und Anpassen der Gewichte verwendet, während ein kleinerer Satz von Bildern zur Validierung dieser Anpassungen dient. Unter Berücksichtigung der Tatsache, dass der Loss-Plot für die Trainingsdaten ständig abnehmen sollte, gefolgt von den Validierungsdaten mit einem kleinen Offset, zeigt der Plot das erwartete Verhalten.

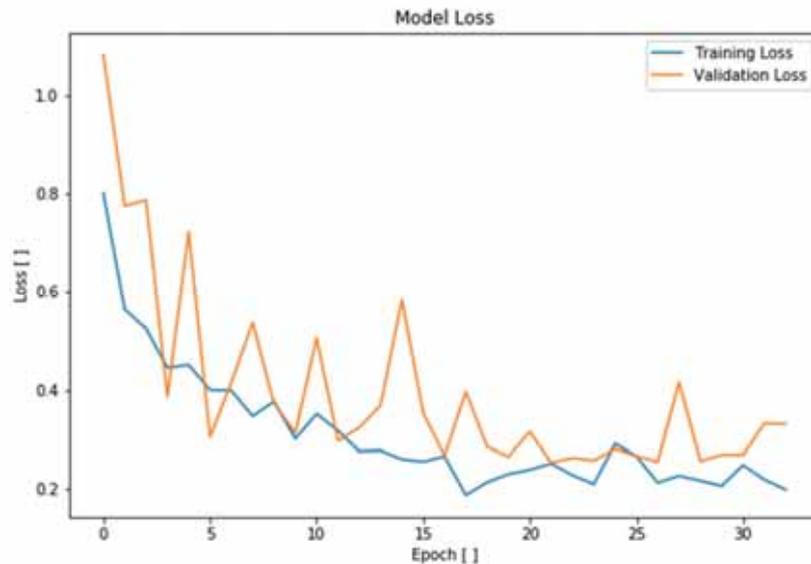


Abb. 6: Loss-Plot mit RMSE während der Epochen des Trainings.

Die Anwendung des trainierten Netzes auf bisher unbekannte Daten ermöglicht einen unabhängigen Leistungstest des Netzes. Diese Daten bestehen aus einer Teilmenge aller Bilder, die als Testbilder markiert sind. So kann überprüft werden, wie gut das Netz wirklich bestimmte Eigenschaften gelernt hat, anstatt sich nur die Trainingsdaten zu merken. Ein Beispiel für die Vorhersage eines Testbildes im Vergleich zu den Referenzdaten ist in Abb. 7 zu sehen. Abgesehen von dem Bereich oben rechts im Bild, in dem die schrägen Unterwasserdistanzen als zu groß prädiziert werden, stimmen die vorhergesagten Werte mit der Referenz generell gut überein. Außerdem gibt es keine größeren Unterschiede in Bezug auf den Trend der Wasserdistanzen. Zu beobachten ist jedoch ein Rauschen in den Prädiktionen, das durch den Kamerasensor oder durch die Dynamik der Wasseroberfläche verursacht sein kann.

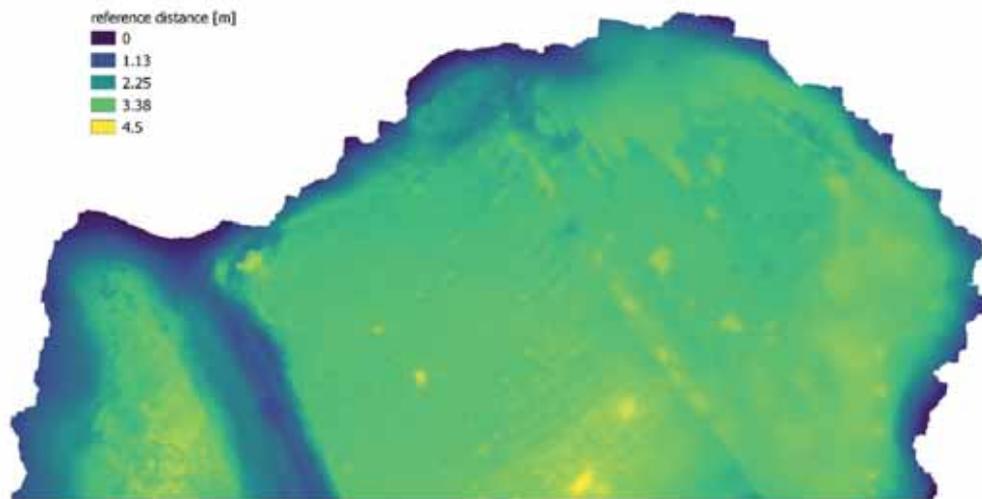


Abb. 7.1: Referenztiefenkarte eines Testbildes.

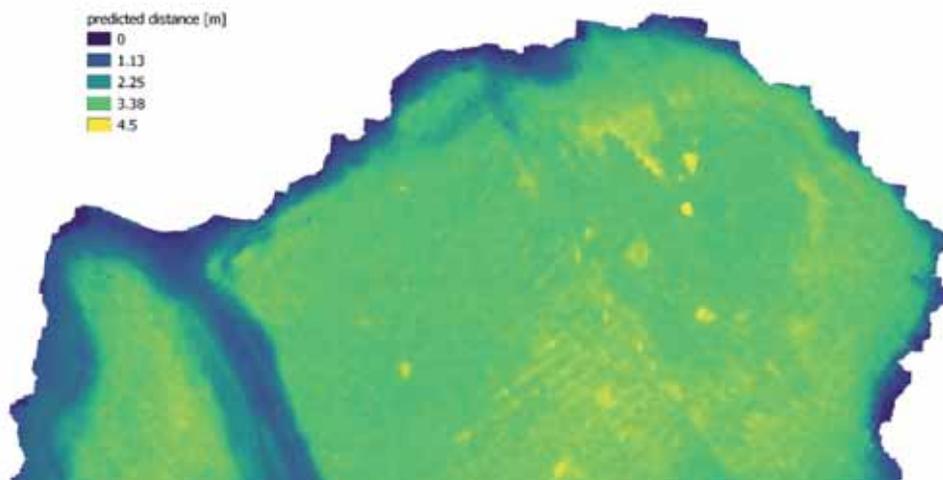


Abb. 7.2: Prädizierte Tiefenkarte eines Testbildes

Nachdem die Distanzen für alle Testbilder prädiziert wurden, kann die Genauigkeit evaluiert werden, indem die prädizierten Distanzen von der Referenz subtrahiert werden. Durch Zusammenführen der Abweichungen für alle Pixel aller Testbilder kann ein Histogramm über alle Tiefenabweichungen (Abb. 8) erhalten werden. Es wird darauf hingewiesen, dass nur Wasserpixel berücksichtigt werden, während alle Pixel in der Vegetation und auf dem Festland maskiert sind. Das Histogramm zeigt eine Verschiebung des Mittelwertes von 4,3 cm in negativer Richtung und weist eine leichte Schiefe in positiver Richtung auf. Es ist nahezu normalverteilt mit einer mittleren absoluten Abweichung von 32,2 cm.

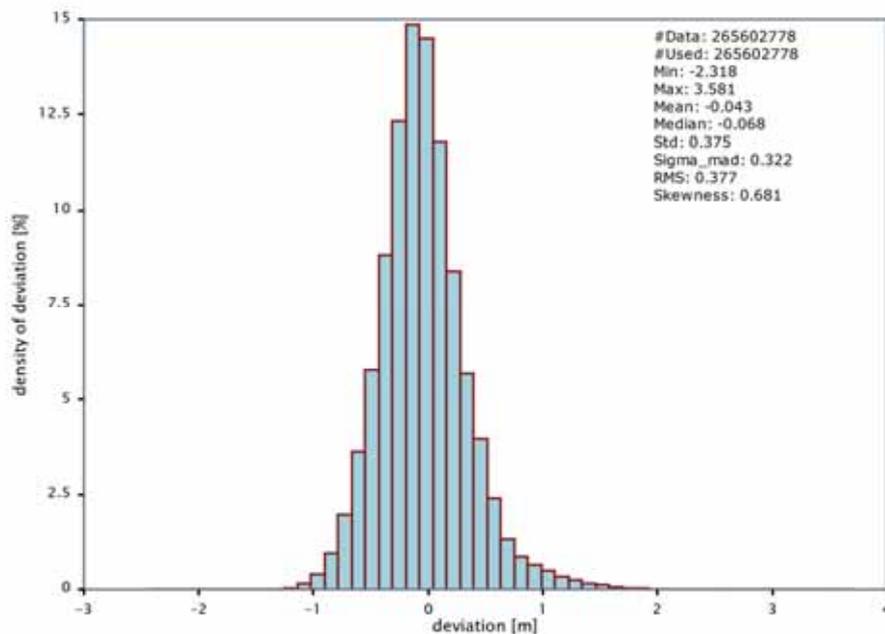


Abb. 8: Histogramm der Abweichungen der prädizierten Unterwasserdistanzen im Vergleich zur den Referenzdistanzen für 24 Testbilder.

Abb. 9 stellt mittels einer Heatmap die Verteilung der vorhergesagten Unterwasserdistanzen bezogen auf die Referenzdistanzen dar. Die Abweichungen nehmen mit zunehmender Tiefe zu. Darüber hinaus neigt das Netz bei Referenzabständen gegen Null dazu, größere Werte zu prädizieren. Dies könnte zum Teil auf schattige Bereiche an der Uferlinie zurückgeführt werden, bei welchen Dunkelheit größere Tiefen impliziert.

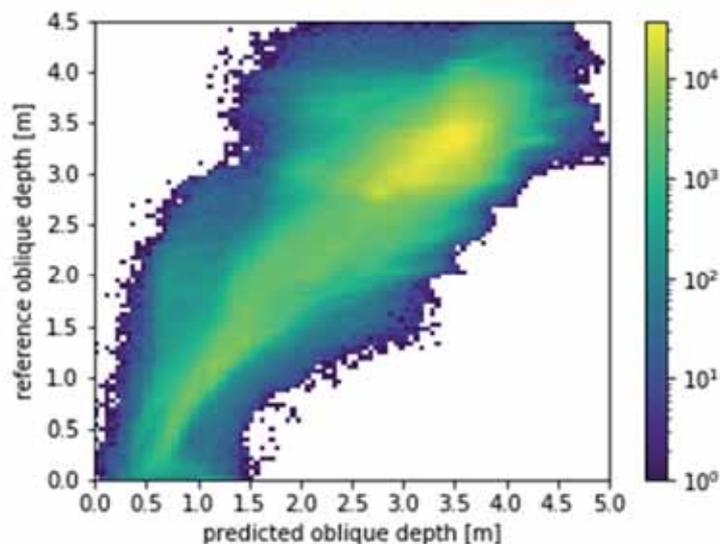


Abb. 9: Heatmap der prädizierten und Referenz-Unterwasserdistanzen.

Ein häufiges Problem bei Luftbildern von Gewässern sind Glanzlichter von der Wasseroberfläche (engl.: sun glint), welche durch direkte Reflexionen des Sonnenlichts in das Sichtfeld des Sensors hervorgerufen werden. Dies führt zu hellen Flecken in den Bildern, in denen es keine Möglichkeit gibt, Merkmale am Grund des Gewässers zu extrahieren. Für Ansätze, die von spektralen Informationen profitieren, kann dies zu großen Fehlern führen (LYZENGA et al. 2006). Ein Ausschnitt aus einem Testbild mit Glanzlichtern, sowie das Differenzbild für die vorhergesagten schrägen Unterwasserdistanzen ist in Abb. 10 dargestellt, in der vor allem in Glanzlicht-Bereichen große Abweichungen bis zu 3 m zu sehen sind. Die schrägen Unterwasserdistanzen in diesem Bereich werden aufgrund der erlernten umgekehrten Proportionalität von Helligkeit und Tiefe als nahezu Null vorhergesagt. Daher sind die Abweichungen von der Referenzdistanz für die jeweiligen Pixel abhängig, wodurch auch die hohe maximale Abweichung und die positive Schiefe in Abb. 10 erklärt werden kann. Dieses Verhalten spiegelt sich auch in der Heatmap als größere Abweichungen von der Diagonale zu kleineren prädizierten Abständen wider.

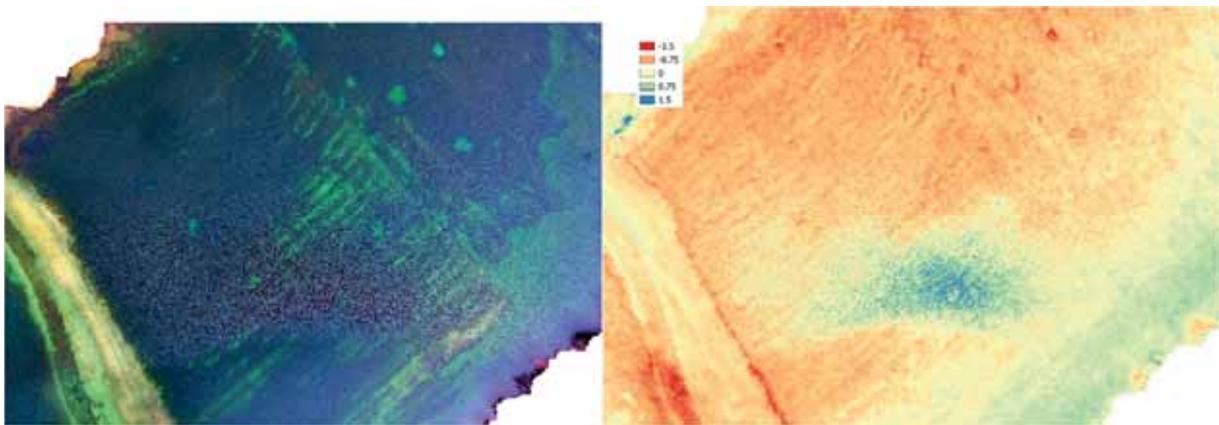


Abb. 20: Glanzlicht-Beispiel in einem Testbild sowie Auswirkung in der Prädiktion.

Darüber hinaus ist unten rechts im Bild in Abb. 10 ein kleiner Bereich mit unmaskierter Vegetation über der Wasseroberfläche, sowie daraus resultierende Schatten zu sehen, die ebenfalls Fehler verursachen. Bei der Untersuchung der kleineren hellen Bereiche oben rechts zeigt sich eine Schwäche der CNNs. Während Faltungskerne Informationen aus der Umgebung berücksichtigen, neigen sie dazu, starke Kanten zu verwischen. So finden sich an den Übergängen von Vegetation (dunkel) zu nacktem Boden (hell) große Abweichungen.

#### 4.2 Angewandtes CNN für RGB ohne Coastal Blue Band

Um den Mehrwert des zusätzlichen Coastal Blue Bandes für die CNN-basierte Tiefenschätzung zu bewerten, wurde das Netz zusätzlich nur auf die RGB-Bilder trainiert. Beim Vergleich der Verteilung der Referenzdistanzen mit beiden Ansätzen in Abb. 11 ist zu erkennen, dass der Ansatz mit Einbeziehung des Coastal Blue Bandes eine wesentlich bessere Übereinstimmung zeigt gegenüber der Variante, welcher lediglich RGB zugrunde liegt.

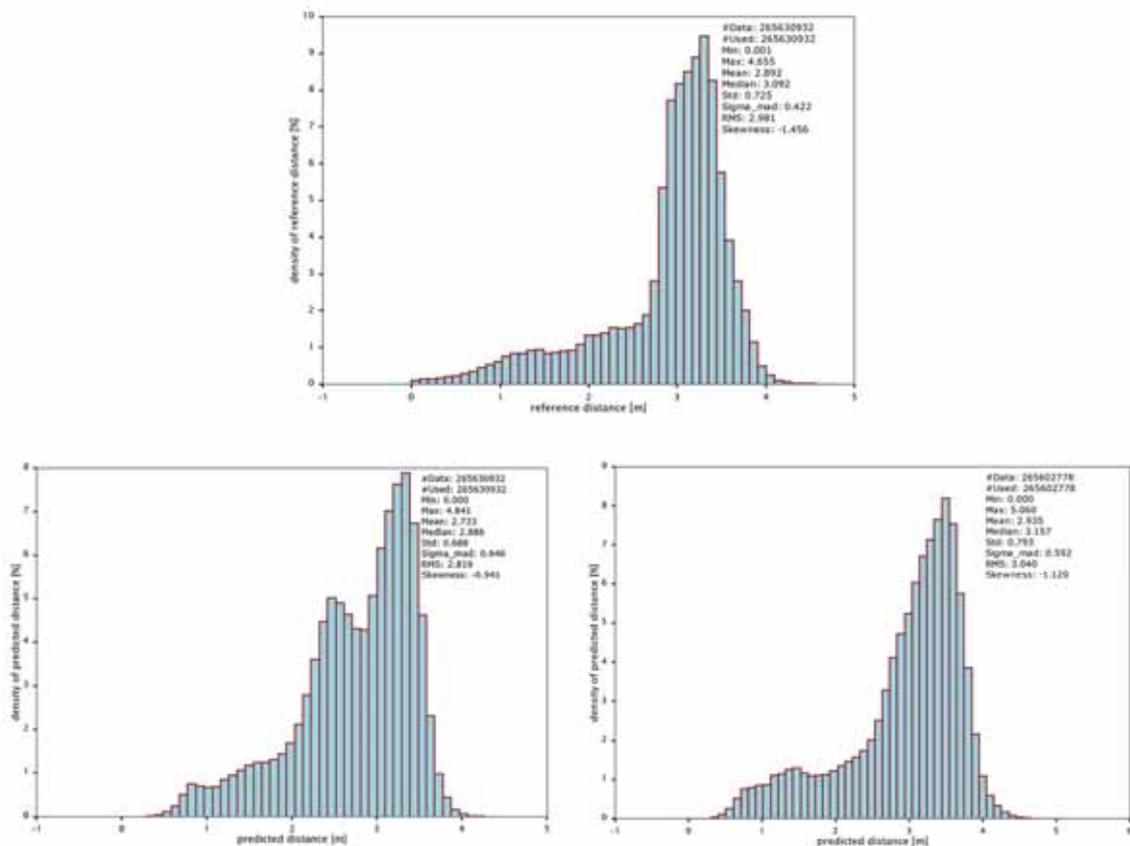


Abb. 31: Histogramme der prädierten Abstände (links: ohne Coastal Blue Kanal, rechts: mit Coastal Blue Kanal) der Testbilder im Vergleich zu dem Histogramm der Referenzabstände (oben).

## 5 Fazit und Ausblick

Im Rahmen des Beitrags wurde ein Ansatz vorgestellt, der die Gewässertiefenschätzung aus Luftbildern mittels faltender neuronaler Netze ermöglicht. Da der Bereich in den Testbildern dem Modell unbekannt ist, sind die Prädiktionen zuverlässig. In Bezug auf die Genauigkeit sind SoNAR oder LiDAR dem beschriebenen Ansatz überlegen. Anderenfalls sind in dieser Arbeit die Vorteile der CNN-basierten Bathymetrie-Schätzung gegenüber dem stereo-photogrammetrischen und linearen Regressionsverfahren aufgezeigt. Aufgrund der unterschiedlichen Bodenbedeckungen des Sees ist ein komplexeres Modell erforderlich als die lineare Regression. Außerdem ist das Ergebnis im Vergleich zu dem photogrammetrischen Ansatz von MULSOW et al. (2019) für homogene Bereiche wesentlich glatter. Darüber hinaus gibt es mehrere Verbesserungsmöglichkeiten. So verursachen beispielsweise Vegetation über der Wasseroberfläche sowie Glanzlichter Abweichungen, die sich sowohl auf das Training als auch auf den Test auswirken. Wenn dies berücksichtigt wird, könnte eine weitere Verbesserung der Ergebnisse erzielt werden.

Ein häufiges Problem bei der Anwendung von CNNs, ist der Mangel an Daten. Der große Vorteil der Verarbeitung im Bildraum, ist die Möglichkeit, den gesamten Datensatz mit allen überlappenden Bereichen zu nutzen, ohne ihn zu reduzieren. Dies ist auch der Grund, warum es möglich war,

die Bilder im Überlappungsbereich von Trainings- und Testbereich abzulehnen, so dass es keine Verbindung gab.

Um die Robustheit des Verfahrens zu evaluieren, kann das Netz auf ähnliche Datensätze (andere Wasserkörper) angewendet werden. Im Idealfall kann das Netz ohne Anpassungen genutzt werden. Andernfalls könnte zur Reduktion der Trainingsdaten „Transfer Learning“ angewendet werden. Der nächste logische Schritt bei diesem Verfahren wäre die Ableitung einer 3D-Punktwolke aus den vorhergesagten Unterwasserdistancen in den Bildern. Zu diesem Zweck ist ein geschätztes Wasserflächenmodell und die Orientierung der Kamera für jedes Bild erforderlich. Dadurch wäre es möglich, die überlappenden Bereiche aufeinanderfolgender Bilder zu analysieren und die unabhängig prädierten Wassertiefen auf Konsistenz zu prüfen. Darüber hinaus könnten Ausreißer in Einzelbildern, z.B. durch Glanzlichter, durch robuste Schätzverfahren für das Endergebnis kompensiert werden.

Eine weitere Überlegung ist die Einführung einer Klassifizierung in „Wasser“ und „Kein Wasser“ während des Trainings, um den derzeit notwendigen Vorverarbeitungsschritt des Maskierens von Trockengebieten zu vermeiden. Dies wurde hier nicht umgesetzt, da die Verfügbarkeit von Referenzdaten für jedes Bild automatisch eine Kein-Wasser-Klassifizierung als Nebenprodukt ermöglichte. Aber gerade bei der Übertragung auf ein anderes Gewässer ohne Referenzdaten sollte dies umgesetzt werden, zumal sonst die Wasserfläche für jedes einzelne Bild segmentiert werden müsste.

Da zumindest für Seen oft eine maximale Tiefe bekannt ist, wäre ein weiterer Ansatz, das vortrainierte Netz zu verwenden, ohne weitere Anpassungen vorzunehmen. Die Abstände könnten dann relativ bestimmt und anschließend auf die maximale Tiefe skaliert werden. Wenn eine gute Genauigkeit erreicht werden kann, wäre es möglich, Tiefenmodelle für Seen abzuleiten, ohne opulente Methoden wie Laserscanning oder Sonar einzusetzen. Der in diesem Beitrag beschriebene Ansatz hat jedenfalls gezeigt, dass moderne CNN-basierte Auswerteverfahren bei transparenten Wasserbedingungen im Flachwasserbereich eine robuste Methode für eine rein bildbasierten Tiefenschätzung darstellen.

## 6 Literaturverzeichnis

- ABADI, M., AGARWAL, A., BARHAM, P., BREVDO, E., CHEN, Z., CITRO, C., CORRADO, G. S., DAVIS, A., DEAN, J., DEVIN, M., GHEMAWAT, S., GOODFELLOW, I., HARP, A., IRVING, G., ISARD, M., JIA, Y., JOZEFOWICZ, R., KAISER, L., KUDLUR, M., LEVENBERG, J., MANE, D., MONGA, R., MOORE, S., MURRAY, D., OLAH, C., SCHUSTER, M., SHLENS, J., STEINER, B., SUTSKEVER, I., TALWAR, K., TUCKER, P., VANHOUCHE, V., VASUDEVAN, V., VIEGAS, F., VINYALS, O., WARDEN, P., WATTENBERG, M., WICKE, M., YU, Y. & GHEMAWAT, S., 2015: TensorFlow: large-scale machine learning on heterogeneous systems. Software available from tensorflow.org. 2015. URL <https://www.tensorflow.org>.
- CHOLLET, F., 2018: Keras: The python deep learning library. Astrophysics Source Code Library.
- COSTA, B.M., BATTISTA, T.A. & PITTMAN, S.J., 2009: Comparative evaluation of airborne LiDAR and ship-based multibeam SoNAR bathymetry and intensity for mapping coral reef ecosystems. Remote Sensing of Environment **113**(5), 1082-1100.

- DUMOULIN, V. & VISIN, F., 2016: A guide to convolution arithmetic for deep learning. arXiv pre-print arXiv:1603.07285.
- GOODFELLOW, I., BENGIO, Y. & COURVILLE, A., 2016: Deep learning. MIT press.
- IRISH, J.L. & WHITE, T.E., 1998: Coastal engineering applications of high-resolution lidar bathymetry. *Coastal engineering* **35**(1-2), 47-71.
- KRAUS, K., 1996: Photogrammetrie Band 2, Verfeinerte Methoden und Anwendungen. Auflage, Dümmler, Bonn.
- LEGLEITER, C.J., ROBERTS, D.A. & LAWRENCE, R. L., 2009: Spectrally based remote sensing of river bathymetry. *Earth Surface Processes and Landforms* **34**(8), 1039-1059.
- LYZENGA, D.R., MALINAS, N.P. & TANIS, F. J., 2006: Multispectral bathymetry using a simple physically based algorithm. *IEEE Transactions on Geoscience and Remote Sensing* **44**(8), 2251-2259.
- MAAS, H.G., 2015: On the accuracy potential in underwater/multimedia photogrammetry. *Sensors* **15**(8), 18140-18152.
- MANDLBURGER, G., KREMER, J., STEINBACHER, F. & BARAN, R., 2018: Investigating the use of Coastal Blue imagery for bathymetric mapping of inland water bodies. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, **42**(1), 275-282, <https://doi.org/10.5194/isprs-archives-XLII-1-275-2018>.
- MASNADI-SHIRAZI, M. A., DE MOUSTIER, C., CERVENKA, P. & ZISK, S. H., 1992: Differential phase estimation with the SeaMARCII bathymetric sidescan sonar system. *IEEE Journal of Oceanic Engineering* **17**(3), 239-251.
- MATLAB, 2018: version 9.5.0 (R2018b). Natick, Massachusetts: The MathWorks Inc.
- MULSOW, C., MANDLBURGER, G., RESSL, C. & MAAS, H.G., 2019: Vergleich von Bathymetriedaten aus luftgestützter Laserscanner- und Kameraerfassung. *Publikationen der Deut. Gesellschaft für Photogrammetrie, Fernerkundung u. Geoinformation e.V., Band 28*, Kersten, T.P. (Hrsg.), 39. Wissenschaftlich-Technische Jahrestagung der DGPF, 20.-22. Februar 2019 in Wien, 318-333.
- NIELSEN, M. A., 2015: *Neural networks and deep learning* (Vol. 25). San Francisco, CA, USA: Determination press.
- NÜBEL, H., 2019: Bathymetry from multispectral aerial images via Convolutional Neural Networks. Bachelorarbeit im Studiengang Geodäsie und Geoinformatik, Universität Stuttgart.
- PFEIFER, N., MANDLBURGER, G., OTEPKA, J. & KAREL, W., 2014: OPALS—A framework for Airborne Laser Scanning data analysis. *Computers, Environment and Urban Systems* **45**, 125-136.
- RIEGL, 2019: RIEGL VQ-880-G: Fully Integrated Topo-Hydrographic Airborne Laser Scanning System. [http://www.riegl.com/uploads/tx\\_pxprigldownloads/Infosheet\\_VQ-880-G\\_2016-05-23.pdf](http://www.riegl.com/uploads/tx_pxprigldownloads/Infosheet_VQ-880-G_2016-05-23.pdf).
- RONNEBERGER, O., FISCHER, P. & BROX, T., 2015: U-Net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, Springer, 234-241.
- WANG, Y., ZHOU, X., LI, C., CHEN, Y. & YANG, L., 2019: Bathymetry Model Based on Spectral and Spatial Multifeatures of Remote Sensing Image. *IEEE Geoscience and Remote Sensing Letters*.

# Remotely sensed information for the protection and management of species-rich grasslands

MICHAEL WACHENDORF<sup>1</sup>, THOMAS ASTOR<sup>1</sup> & JAYAN WIJESINGHA<sup>1</sup>

*Abstract: Knowledge on the forage quality and invasive plant species coverage is important for the management and maintenance of biodiversity in grasslands which are threatened by the invasion of such plants and to evaluate the effect of control activities conducted. Remote sensing (RS) is a promising tool for estimating field data, however, the applicability of RS prediction models depends on the variability of underlying calibration data, which can be brought about by the inclusion of a multitude of grassland types and management practices in the model development. Major aims of this study were (i) to build prediction models for forage quality and invasive species coverage based on unmanned aerial vehicle (UAV)-borne imaging spectroscopy and (ii) to generate maps using the best models obtained. The study examined data from a variety of grasslands which largely differed in terms of vegetation type and cutting regime.*

## 1 Introduction

In Europe, approximately 30 to 35 % of the agricultural area consists of grasslands. Mainly permanent grasslands are incredibly variable in species composition, biodiversity, management practices, as well as in productivity (LESSCHEN et al. 2014). Food provision as forage for ruminants and herbivores and as biomass substrate for energy production are the most comprehensive ecosystem services from grasslands. There exists a multitude of destructive and non-destructive methods to measure or estimate the production and quality of the forage. Usually, farmers use visual criteria to evaluate forage quality, such as the plants' phenological stage, leafiness, or colour. Additionally, lab-based chemical analysis and near-infrared reflectance spectroscopy (NIRS) are utilised by agronomists to evaluate forage quality (HORROCKS & VALLENTINE 1999). Lab-based methods comprise the assessment of chemical forage components that relate to the digestibility of the forage. Acid detergent fibre (ADF), which represents cellulose, lignin and silica, is an important parameter which negatively correlates with forage digestibility (HORROCKS & VALLENTINE 1999). The protein concentration of forages is another parameter essential for the creation of adequate fodder rations. Proteins consists mainly of amino acids, which are fundamental elements of all cells and tissues, and form an essential component of ruminant nutrition and provides nitrogen for ruminants' metabolism and production of milk and meat (FRAME & LAIDLAW 2011).

Remote sensing (RS) is a non-destructive method for estimating grassland biomass and forage quality (NUMATA 2011). From satellite RS to field spectroscopy and from optical RS to synthetic aperture radar were employed for forage quality monitoring and mapping (WACHENDORF 2018). A literature review on RS based forage quality estimation studies published between 2004 and

---

<sup>1</sup> Universität Kassel, Grasland Science and Renewable Plant Resources, Steinstraße 19, D-37213 Witzenhausen, E-Mail: [mwach, thastor, jayan.wijesingha]@uni-kassel.de

2018 shows that more than 60 % of studies (21 out of a total 31 reviewed studies) have utilised field spectroscopy. Often point level spectral reflectance data from visible to short-wave infrared regions of the electromagnetic spectrum is collected in field spectroscopy (MILTON 1987). Crude protein (CP), nitrogen (N), neutral detergent fibre (NDF), and ADF are the most common forage quality parameters that have been estimated using field spectroscopy data for different grasslands with significant accuracy (MUTANGA et al. 2004). Typically, the models developed to estimate forage quality using spectral data are restricted to single grassland types. Consequently, the transferability of the models to other grassland types is limited due to the low variability of the underlying training data. Therefore, a general model that can estimate forage quality parameters, irrespective of the grassland type, would be a preferable.

Biological invasion is threatening to biodiversity in many ecosystems in the world. The invasion by alien plant species is considered as one of the significant drivers for loss of biodiversity and ecosystem functionality. The leading cause for the introduction of alien plant species is human activities. Classification of remotely sensed images to map invasive species is a well-adopted technology for many invasive species in different parts of the world and various ecosystems (ROYIMANI et al. 2018). In grasslands, invasive woody (MIRIK et al. 2013) and shrubby (LALIBERTE et al. 2004) species mapping has shown excellent results with satellite and airborne image classification. Meanwhile, using unmanned aerial vehicle (UAV) borne remotely sensed (RS) imaging was applied for invasive species mapping in the recent years in flood plains and coastal regions (DE SÁ et al. 2018). Apart from the spectral or thermal information, UAV-borne RS data can also provide 3D point cloud data, which can be employed to derive canopy height models (CHM) for grassland canopies (WIJESINGHA et al. 2019). The knowledge of the spatial distribution of lupine in species-rich grasslands is vital to conduct control activities and to monitor their efficacy. Therefore, a repeatable, transferable methodology is needed, that produces lupine distribution maps at different spatial and temporal scales to monitor the lupine distribution and to assess the benefit of control activities.

## 2 Material and Methods

### 2.1 Forage quality

A study was carried out on eight grasslands with different management practice and species composition. Six grasslands were in nature protection areas, where no fertilisation was applied, and they were mowed only once per year. Two of them were mountain hay meadows (MHM), and another two were *Nardus stricta* grasslands (NSG). MHML and NSGL were MHM and NSG grasslands, which were substantially invaded by the neophyte *Lupinus polyphyllus*. The lowland hay meadow (LHM) is an extensively utilised grassland located on the Werra riverbank in Northern Hesse, and it was mowed two times per year. The intensively managed grassland (IMG) was fertilised and was harvested three times per year. MHM1, NSG1, LHM, and IMG grasslands were in Werra-Meißner district in Hesse (9.9 °N, 51.3 °E) (Table 1). Further four grasslands (MHM2, NSG2, MHML, and NSGL) were located in UNSECO Biosphere reserve Rhön, Bavaria and Hesse (10.0 °N, 50.5 °E).

A Cubert Hyperspectral Fireflye S185 SE (Cubert GmbH, [www.cubert-gmbh.de](http://www.cubert-gmbh.de)) snapshot camera (12 mm focal length) was used which acquires spectral images between 450 and 998 nm.

The spectral resolution of the sensor (full width at half maximum) is 4.8 nm at 450 nm and 25.6 nm at 850 nm. The camera records a total of 138 spectral bands with 4 nm sampling interval. Each spectral band image is 50 by 50 pixel, and radiometric resolution is 12 bit (0 - 4096 digital numbers). In addition to the spectral bands, a grayscale panchromatic image is also recorded with 1000 by 990 pixels in size (AASEN et al. 2015). The spatial resolution of the spectral image is ~20 cm at 20 m flying height. The Cubert camera was attached to the UAV (RTK X8 Hyperspectral Mapping, Cubert GmbH, www.cubert-gmbh.de). It is a co-axial multi-rotor UAV equipped with real-time kinematic (RTK) global navigation satellite system (GNSS).

Tab. 1: Details on samplings of the grasslands investigated in the study.

Field ID	Harvest	Sampling date	No. of quality samples
MHM1	First cut	13th July	20†
NSG1	First cut	14th July	19
MHM2, NSG2, MHML, NSGL	First cut	13th June	20 (5, 5, 5, 5)
MHM2, NSG2, MHML, NSGL	First cut	27th June	20 (5, 5, 5, 5)
MHM2, NSG2, MHML, NSGL	First cut	11th July	20 (5, 5, 5, 5)
LHM	First cut	28th May	20
	Second cut	24th September	15
IHM	First cut	10th May	20
	Second cut	06th June	20
	Third cut	01st August	20

Spectral images were acquired between 10:00 and 14:00 under clear-sky conditions. Before the UAV flight, four corners of the study area were staked out using the Leica RTK GNSS (Leica Geosystem, www.leica-geosystems.com) and six black-white 1 m<sup>2</sup> ground control points (GCP) were distributed around the study plot.

After spectral data collection, grass biomass was clipped on the identified 1 m<sup>2</sup> subplots at a stubble height of 5 cm. The fresh biomass was weighted in the field, and total biomass in each plot was divided into two separate samples for dry biomass and forage quality analysis. The samples for the quality analysis were dried at 65°C for 48 hours. Afterwards, the dried samples were ground for 1 mm uniform particles with a Foss CT 193 Cyclotec mill (FOSS, www.fossanalytics.com). Subsequently, the dry matter and the ash content of the ground samples were determined, and the N and ADF were determined. The ADF was determined using the ANKOM 200 Fibre Analyser (ANKOM Technology, www.ankom.com).

## 2.2 Invasive species coverage

A study was carried out in two grassland fields in the UNESCO biosphere reserve Röhn in Germany which were invaded by lupine. One field was classified as a former mountain hay meadow (hereafter referred to as G1), and the other was an old *Nardus stricta* grassland (hereafter referred to as G2). In both fields, rectangle plots of 1500 m<sup>2</sup> (50 m by 30 m) were chosen as study areas, and 15 small plots of 64m<sup>2</sup> (8 m by 8 m) were established within a grid. Three cutting dates (12th June, 26th June, 09th July, hereafter referred to as D1, D2, and D3, respectively) were randomly assigned to 5 replicated plots. At each date, plots were mowed at a stubble height of 5 cm, and biomass was removed from the field.

At each sampling date in each grassland field, UAV-borne images were acquired. A DJI-Phantom IV quadcopter (DJI, China) with an inbuilt off-the-shelf camera (FC330) was employed to obtain UAV-borne RGB images. The camera (FC330) captures a 12-megapixel image in red (R), green (G), and blue (B) bands. The UAV was flown at 20 m flying height, and it resulted in 0.09 m ground sampling distance. The UAV flight mission was designed using Pix4D capture app for Android (App version 4.4.0, Pix4D, Switzerland). The UAV was flown as double grid mission (two perpendicular missions), and the camera was triggered automatically to capture nadir looking images based on the image overlap configuration (80 % both forward and side overlap). All flight sessions were conducted between 12:00 and 14:00. Before each flight session, nine black and white 1 m<sup>2</sup> ground control points were distributed over the study area. Just after the UAV flights, the position of each ground control point was measured using a Leica RTK GNSS (Leica Geosystems GmbH, Germany) with 2 cm 3D coordinate precision. Additional UAV-borne RGB image was taken on 16th August 2019, when the whole fields were mowed.

A FLIR Vue Pro R (FLIR Systems Incorporation, USA) thermal camera was attached to the UAV parallel to the RGB camera. The camera has a 19 mm lens, and it has a spectral sensitivity between 7.500 – 13.500 nm. With a single UAV flight, both thermal and RGB images were captured simultaneously. The thermal camera captures images as a radiometric JPEG which contains radiometrically calibrated temperature data. The thermal image has 640 by 512 pixels (FLIR 2016). The thermal camera was triggered every second throughout the whole UAV mission. Before each thermal data collection, metadata related to the thermal camera was collected using the FLIR UAS 2 app (App version 2.2.4, FLIR Systems Incorporation, USA), such as distance to the target (20 m), relative humidity, air temperature, and emissivity (0.98). All the metadata was saved in each captured image's EXIF data.

A total of six UAV-borne RGB and six thermal datasets were collected. Hereafter, each dataset is labelled according to cutting date and grassland type (DiGj: where  $i = 1, 2, 3$  and  $j=1, 2$ ). In each dataset, maturity stages of grasslands were different due to mowing activities in 64 m<sup>2</sup> small plots. Maturity stage was lowest (V0) in the D1 dataset and was the same for all 30 small plots. At the 2nd cutting date (D2), 20 small plots out of 30 were covered by two weeks older vegetation (V2weeks), while 10 small plots (which were cut at D1) had vegetation which was regrown for two weeks (VR2weeks). The D3 dataset was composed of 10 plots with undisturbed vegetation (V4weeks) which was four weeks older than V0, 10 plots of (VR2weeks), and further 10 plots with vegetation regrown for four weeks (VR4weeks) after D1. The UAV-borne RGB images and coordinates of ground control points were processed with the Agisoft PhotoScan Professional version 1.4.4 software (Agisoft LLC, Russia). The software applied structure from motion (SFM) technique to align multi-view overlapping images and to build a dense 3D cloud point. The procedure of point cloud generation and canopy height computation was adopted from WIJESINGHA et al. (2019), and further details of the process can be found there.

RGB ortho-mosaic was obtained after further processing of the dense point cloud in PhotoScan software. The output RGB ortho-mosaic was geo-referenced with a 1 cm spatial resolution. The RGB ortho-mosaic was converted into hue (H), intensity (I), and saturation (S) colour model using GRASS GIS and hereafter it was considered as HIS ortho-mosaic (GONZALEZ & WOODS 2008). Like with RGB, thermal ortho-mosaics with 2 cm spatial resolution were generated using calibrated thermal images. Using an object-based image analysis (OBIA) each segmented object

was created as a polygon. Four geometric attributes (area (A), perimeter (P), fractional dimension (FD), and circle compactness (CC) for the segmented objects were calculated. Based on all raster data (RGB image, HIS image, CHM raster, PD raster, thermal image, SSI image, and texture raster), the mean and standard deviation values for each polygon was computed as image-based attributes. Ten percent of the segmented objects were manually labelled as either lupine (L) or non-lupine (NL) based on visual observation using the RGB ortho-mosaics. The labelled objects with attributes were utilised to develop a supervised classification model.

Classification model training and testing were conducted using R statistical software (R CORE TEAM 2019). The random forest (RF) machine learning classification algorithm was employed to build a classification model using the `mlr` package in R software (BISCHL et al. 2016). The model was trained with repeated spatial cross-validation resampling. According to predicted labels and actual labels, the model performance was evaluated by calculating overall accuracy (OA), true-positive-rate (TPR), false-positive-rate (FPR), and Kappa (K) values. A single RF classification model ( $M_{all}$ ) was trained using all labelled objects from the six datasets. Based on predicted labels from  $M_{all}$ , a lupine coverage map was generated (hereafter referred to as classification-based lupine coverage map).

### 3 Results

#### 3.1 Forage quality

Crude protein concentration varied between 5.1 and 23.3 %DM, while ADF varied between 22.5 and 38.5 %DM. Forage from the intensively managed grassland (IGM) had the highest average CP and the lowest average ADF. However, forage from the two grasslands invaded by *Lupinus polyphyllus* (MHML, NSGL) contained higher CP than non-invaded grasslands. Normalised mean reflectance values in the visible region obtained for every 1 m<sup>2</sup> sampling plot were lower for samples with higher CP values along with higher values in the near-infrared region. A similar pattern was found for ADF data.

Based on 100 different model runs with random train and test data sets the model with the highest accuracy (lowest median  $RMSE_p$  and lowest median  $nRMSE_p$ ), the highest precision (highest  $R^2_p$ ) and the highest stability (lowest standard deviation of  $RMSE_p$ ) was identified as the best-performing model (Table 2). Accordingly, the SVR model (median  $RMSE_p = 1.9$  %DM; median  $nRMSE_p = 10.6$  %; median  $R^2_p = 0.79$ ; SD  $RMSE_p = 0.29$  %DM) was the best model for CP estimation, whereas for ADF the CBR model (median  $RMSE_p = 2.2$  %DM; median  $nRMSE_p = 13.4$  %; median  $R^2_p = 0.56$ ; SD  $RMSE_p = 0.23$  %DM) was the optimal model. PLSR was the least performing model type among all the predictive algorithm models. For CP estimation,  $nRMSE_p$  of the SVR model varied between 7.0 % to 14.5 % and  $nRMSE_p$  varied from 6.5 to 16.4 % of the CBR model. Comparably,  $nRMSE_p$  of ADF models varied from 10.7 to 18.3 % and from 10.5 to 16.7 % for SVR and CBR models respectively. Further, the precision of (median  $R^2_p$ ) CP models was larger than 0.73, except for the PLSR model. However, the precision of (median  $R^2_p$ ) the ADF models was lower than for CP models.

Tab. 2: Summary of the predictive algorithm models (from 100 different model runs) for CP and ADF estimation of different grasslands. (SD: standard deviation, CBR: cubist regression, GPR: gaussian processing regression, PLSR: partial least squares regression, RFR: random forest regression, and SVR: support vector regression)

Forage quality parameter	Algorithm	Median $R^2_p$	Median $RMSE_p$ (%DM)	SD $RMSE_p$ (%DM)	Median $nRMSE_p$
CP	PLSR	0.48	3.0	0.36	16.5 %
	GPR	0.73	2.3	0.33	12.4 %
	RFR	0.74	2.1	0.38	11.5 %
	SVR	0.79	1.9	0.29	10.6 %
	CBR	0.77	1.9	0.45	10.4 %
ADF	PLSR	0.39	2.6	0.31	16.4 %
	GPR	0.51	2.3	0.25	14.5 %
	RFR	0.52	2.3	0.24	14.5 %
	SVR	0.50	2.3	0.26	14.5 %
	CBR	0.56	2.2	0.23	13.4 %

The plots of fit for the best-performing models show the model fit across all grasslands (Fig. 1). Overall, prediction accuracy tended to be lower at higher levels of CP, whereas for ADF accuracy was consistent across the whole range of values observed.

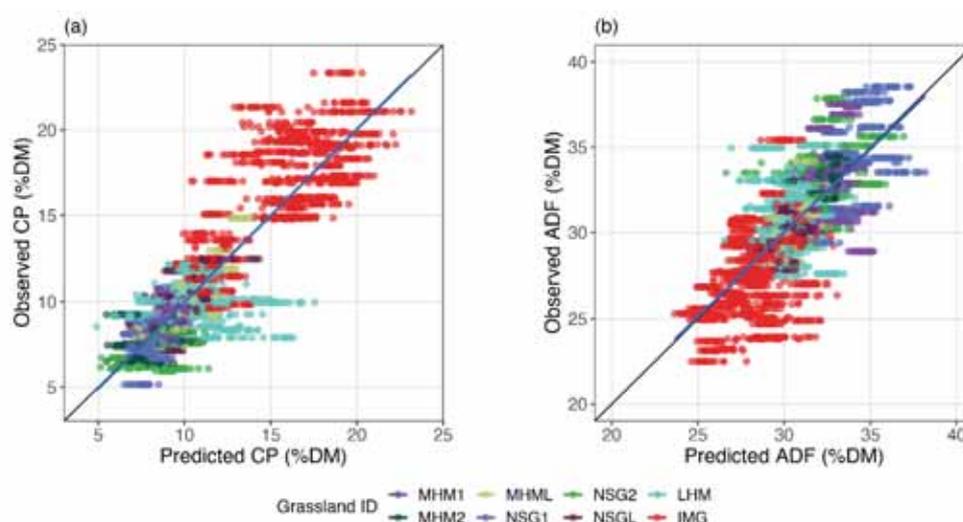


Fig. 1: Prediction vs. observation scatter plots from the SVR model for CP (a) and the CBR model for ADF (b) concentrations in different grasslands. Colours represent different grasslands. The black line is the 1:1 line, and the blue line represents the linear regression line between observed and predicted values.

With SVR and the CBR predictive algorithms being identified as the best algorithms to estimate CP and ADF, repeated cross-validations were performed using the complete data set. The SVR model resulted in a  $nRMSE_{cv}$  of 9.6 % with  $R^2_{cv} = 0.81$  for CP estimation, while for ADF estimation the CBR model had a  $nRMSE_{cv}$  of 13.0 % and a  $R^2_{cv}$  of 0.60. The errors of the final models were found between the errors from 100 different models in the model training and testing phase.

### 3.2 Invasive species coverage

Six classification models were trained while holding out one dataset at each time. The model results are summarised in Table 3. Based on the all performance measures in model testing phase, model M12 (model tested with D1G2 data) obtained the lowest performances (OA = 78.2 %, K = 0.55) and model M32 (model tested with D3G2 data) achieved the highest values (OA = 97.2 %, K = 0.94). Although model M12 accurately classified all lupine objects (100 % TPR), it also categorised nearly half of the non-lupine objects as lupine objects (47.3 % FPR). Additionally, models that tested with D1 data (M11 and M12) obtained slightly lower performances compared to other models.

Tab. 3: Classification model results. L: lupine, NL: non-lupine, TA: training accuracy, OA: overall accuracy, K: Kappa statistics, TPR: true positive rate, FNR: false negative rate

Model	Training	Testing				
	No. of objects	No. of objects	OA (%)	TPR (%)	FPR (%)	K
M11	L = 1581; NL = 1545	L = 311; NL = 261	83.2	76.8	9.2	0.67
M12	L = 1394; NL = 1381	L = 498; NL = 425	78.2	100.0	47.3	0.55
M21	L = 1578; NL = 1429	L = 314; NL = 377	90.6	84.1	4.0	0.81
M22	L = 1701; NL = 1638	L = 191; NL = 168	96.4	95.8	3.0	0.93
M31	L = 1530; NL = 1405	L = 362; NL = 401	90.1	88.4	6.7	0.82
M32	L = 1676; NL = 1632	L = 216; NL = 174	97.2	96.7	2.3	0.94

After testing six classification models with the different spatial-temporal dataset, the complete classification model (M<sub>all</sub>) was trained using all available data (3698 objects) with spatial cross-validation. The M<sub>all</sub> model achieved 94.2 % training accuracy. Based on visual observation between digitised lupine map and classified lupine map, both maps showed similar visual representation. Figure 2 illustrates lupine coverage maps from both digitising and classification for three sampling dates (D1, D2, D3) in G1 field. Relationship between the relative LA and MA indicated a negative exponential trend (Fig. 3). The correlation coefficient between relative LA and MA was -0.88, and trend line had 0.80 goodness of fit. Regardless of the vegetation maturity, the explained relationship was valid. Until LA reached 25 %, it showed a strong relationship with MA, but over 25 % LA the MA values were scattered around the regression curve.

## 4 Discussion

### 4.1 Forage quality

Several predictive modelling algorithms were tested to identify the best algorithms to estimate CP and ADF from the full spectral data as no single consistent algorithm was shown to surpass all the given circumstances every time in a study by YUAN et al. (2017). Moreover, model consistency was evaluated by training and testing 100 models using 100 different random train and test data sets, which allowed to disclose the model performance irrespective of the calibration data set. Except for the PLSR, other tested predictive algorithms (GPR, RFR, SVR and CBR) proved promising for the underlying data. The SVR and CBR models for CP and ADF estimation showed the maximum precision (highest R<sup>2</sup><sub>p</sub>) and prediction accuracy (lowest nRMSE<sub>p</sub>) followed by RFR, GPR, and PLSR models, respectively. SVR and CBR predictive modelling algorithms were

previously utilised to estimate water quality parameters based on spectral data (HAFEEZ et al. 2019), however, to our knowledge such algorithms have not been employed so far to estimate forage quality parameters.

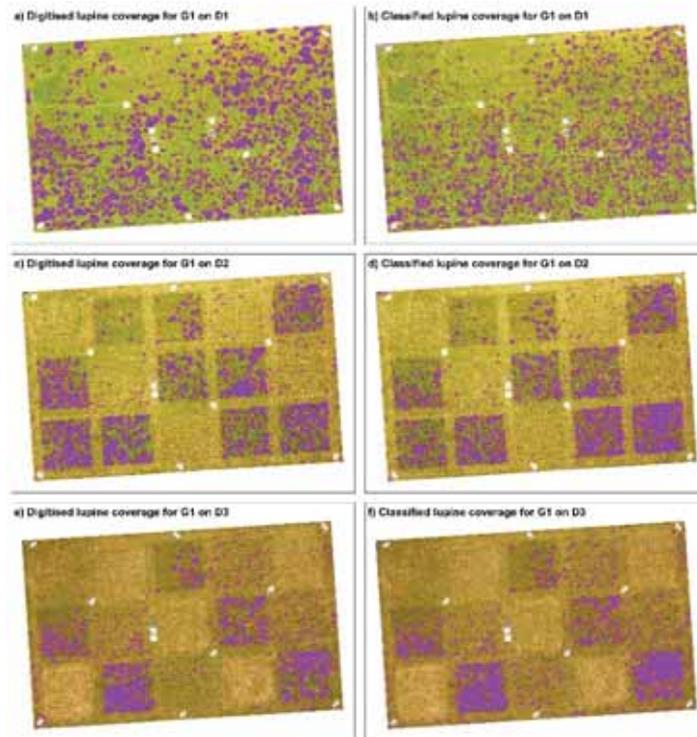


Fig. 2: Lupine coverage map of the G1 field with a, c, e, showing manually digitised lupine cover (purple) at D1 (12th June), D2 (26th June), and D3 (9th July) and b, d, f, showing lupine cover classified by UAV-borne RS data and OBIA

According to the literature, PLSR and RFR were the most prominent predictive modelling algorithms in forage quality parameter estimation using spectral data. For example, SAFARI et al. (2016) report on PLSR models, which obtained nRMSE values of 8.5 % and 7.3 % for CP and ADF respectively. However, PLSR models resulted in the lowest accuracies both for CP and ADF in this study. Moreover, PULLANAGARI et al. (2018) achieved an nRMSE of 11.2 % for CP with the RFR model, and SINGH et al. (2017) obtained an nRMSE of 21.7 % for ADF with the same modelling algorithm. It is noteworthy that the studies mentioned above only tested one predictive modelling algorithm. Thus, no conclusions are possible considering the comparison with other algorithms.

Both CP and ADF estimation models resulted in less than 15 % relative prediction error. But CP estimation had slightly lower relative error (median nRMSE<sub>p</sub> = 10.6 %) than ADF estimation (median nRMSE<sub>p</sub> = 13.4 %). Similar relative error pattern for CP and ADF estimation models was obtained in previous studies that utilised field spectroscopy data (KAWAMURA et al. 2008; SAFARI et al. 2016). Moreover, the data points from the IMG were clustered out and acted as the driver of the CP model according to the plot of fit (**Fehler! Verweisquelle konnte nicht gefunden werden.1a**). Though, ADF data points in the observed against predicted plot did not highlight a similar pattern. The high variation of CP between the cuts in IMG due to intensive management

practice might be the reason caused for the mentioned pattern. However, the data points from other grasslands were mostly grouped in both plots because they almost experienced similar management practice.

To summarise, predictive modelling algorithms allow an adequate forage quality estimation regardless of the grassland type, and cutting regimes applied. Ultimately, fine-tuning of the calibrated models with data from further diverse grassland types could increase the robustness of the models to generate forage quality maps for grasslands with different vegetation types and management practices.

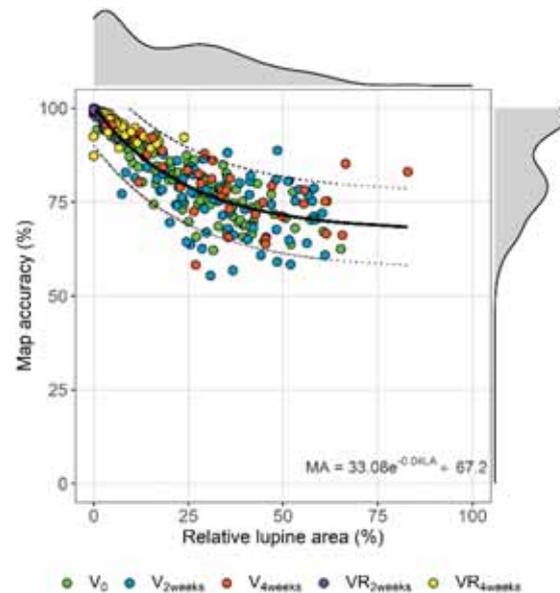


Fig. 3: The relationship between relative lupine area (LA) from manual digitising and map accuracy based on the generalised model, comprising undisturbed/not mowed vegetation (V0, V2weeks, V4weeks), and regrown vegetation after mowing (VR2weeks, VR4weeks). Grey areas indicate the data density along the x- and y-axis. The black line represents the fitted exponential curve, and dotted lines show the 95th confidence interval of the fitted curve.

## 4.2 Invasive species coverage

Invasion by lupine endangers biodiversity and ecosystem functionality (KLINGER et al. 2019). The spatial and temporal distribution of lupine is essential to understand the invasive pattern, to plan appropriate management strategies and to monitor the impact of control actions. OBIA has shown its effectiveness to work with very high-resolution (< 1 m spatial resolution) images, where several pixels represent one object rather than classifying each pixel separately. The first step of the proposed workflow was to segment UAV-borne images into image objects that represent either lupine or non-lupine plants. USPO based area-specific threshold values benefitted for obtaining good object delineation. Though USPO based area-specific threshold values provided good object delineation, it may lead to increased computational time for a multitude of image areas, depending on the size of the areas and the spatial resolution of the images. Several attributes related to plant structure or architecture as well as colour were essential predictors in the  $M_{all}$  model. The height difference between lupine plants and grass vegetation contributed to the classification of segmented objects. Segment's area and perimeter were further vital geometric features in the final

classification model, whereas fractional dimensions and circle compactness were not useful. A closer look at the segmented objects shows that the area and perimeter of lupine objects were substantially smaller compared to non-lupine objects, irrespective of the lupine coverage.

Lupine containing areas showed lower temperatures in thermal images as surrounding grass areas, which may be due to higher water contents compared to grasses (HENSGEN & WACHENDORF 2016) and increased shades around the bush-like growth of the lupine. However, no temperature-related attributes emerged as significant predictors in the classification models which can be evaluated positively, as it leads to reductions in costs for sensors and platforms as well as in model complexity and computing time. Mapping accuracy declined with increasing lupine coverage both for undisturbed and regrown grassland vegetation of different maturities. In general, early detection of invasive plant species and rapid action is critical to control invasive species (COCK & WITTENBERG 2001). Similarly, lupine control in the biosphere reserve is mainly conducted in regions with low lupine coverage, as this stage of invasion facilitates a fast eradication and containment compared to regions with lupine dominance. Thus, the use of lupine coverage maps can help to identify regions with relatively small lupine coverage and precisely locate single lupine plants for eradication.

## 5 Conclusions

The present study aimed to estimate forage quality of a multitude of grasslands with different vegetation composition and cutting regimes applied based on UAV-borne imaging spectroscopy data. It was demonstrated that the resulting models could accurately estimate CP and ADF irrespective of the grassland type and that model accuracies are in the same range as those obtained with the use of field spectroscopy. Further a workflow was developed that can accurately map lupine coverage in a grassland using UAV-borne RS and OBIA. A robust RF classification model allowed the classification of lupine and non-lupine image objects. Such classification models can be transferred to other regions, and thereby overcome limitations of the standard way of lupine mapping. Finally, the developed procedures can be adopted for mapping forage quality and invasive species which may provide benefit for practical grassland management and maintenance.

## 6 References

- AASEN, H., BURKART, A., BOLTEN, A. & BARETH, G., 2015: Generating 3D Hyperspectral Information with Lightweight UAV Snapshot Cameras for Vegetation Monitoring: From Camera Calibration to Quality Assurance. *ISPRS Journal of Photogrammetry and Remote Sensing* **108**, 245-259, doi:10.1016/j.isprsjprs.2015.08.002.
- BISCHL, B., LANG, M., KOTTHOFF, L., SCHIFFNER, J., RICHTER, J., STUDERUS, E., CASALICCHIO, G. & M. JONES, Z. M., 2016: Mlr: Machine Learning in R. *Journal of Machine Learning Research* **17**, 1-5.
- COCK, M. J. W. & WITTENBERG, R., 2001: Early Detection. In *Invasive Alien Species: A Toolkit of Best Prevention and Management Practices*, edited by Matthew J W Cock and Rüdiger Wittenberg, 101-123. Wallingford: CAB International. doi:10.1079/9780851995694.0000.

- DE SÁ, N. C., CASTRO, P., CARVALHO, S., MARCHANTE, E., LÓPEZ-NÚÑEZ, F. A. & MARCHANTE, H., 2018: Mapping the Flowering of an Invasive Plant Using Unmanned Aerial Vehicles: Is There Potential for Biocontrol Monitoring? *Frontiers in Plant Science* **9**, 1-13, doi:10.3389/fpls.2018.00293.
- FRAME, J. & LAIDLAW, A.S., 2011: Feeding Value of Grass. In *Improved Grassland Management*, edited by John Frame and A.S. Laidlaw, 167-180. Wiltshire: Corwood Press.
- GONZALEZ, R. C. & WOODS, R. E., 2008: Color Image Processing. In *Digital Image Processing*, edited by Rafael C Gonzalez and Richard E Woods, Third Edit, 401-414. New Jersey: Pearson Education.
- HAFEEZ, S., WONG, M. S., HO, H. C., NAZEER, M., NICHOL, J., ABBAS, S., TANG, D., LEE, K. & PUN, L., 2019: Comparison of Machine Learning Algorithms for Retrieval of Water Quality Indicators in Case-II Waters: A Case Study of Hong Kong. *Remote Sensing* **11**(6), 617, doi:10.3390/rs11060617.
- HENSGEN, F. & WACHENDORF, M., 2016: The Effect of the Invasive Plant Species *Lupinus Polyphyllus* Lindl. on Energy Recovery Parameters of Semi-Natural Grassland Biomass Sustainability (Switzerland) **8**(10), doi:10.3390/su8100998.
- HORROCKS, R. D. & VALLENTINE, J. F., 1999: Forage Quality - the Basics. In *Harvested Forage*, edited by R Dwain Horrocks and Jhon F Vallentine, 17-47. San Diego: Academic Press, doi:10.1016/B978-0-12-356255-5.50024-9.
- HUYGHE, C., DE VliegHER, A., VAN GILS, B. & PEETERS, A., 2014: Grasslands and Herbivore Production in Europe and Effects of Common Policies. Edited by Christian Huyghe, Alex De Vlieghe, Bert van Gils, and Alain Peeters. Quae Editor, Versailles, doi:10.35690/978-2-7592-2157-8.
- KAWAMURA, K., WATANABE, N., SAKANOUÉ, S., & INOUE, Y., 2008: Estimating Forage Biomass and Quality in a Mixed Sown Pasture Based on Partial Least Squares Regression with Waveband Selection. *Grassland Science* **54**(3), 131-145, doi:10.1111/j.1744-697X.2008.00116.x.
- KLINGER, Y. P., HARVOLK-SCHÖNING, S., ECKSTEIN, R. L., HANSEN, W., OTTE, A., & LUDEWIG, K., 2019: Applying Landscape Structure Analysis to Assess the Spatio-Temporal Distribution of an Invasive Legume in the Rho UNESCO Biosphere Reserve. *Biological Invasions* **21**(8), 2735-2749, doi:10.1007/s10530-019-02012-x.
- LALIBERTE, A. S., RANGO, A., HAVSTAD, K. M., PARIS, J. F., RELDON, F., BECK, ROB MCNEELY, AND AMALIA L. GONZALEZ: 2004: Object-Oriented Image Analysis for Mapping Shrub Encroachment from 1937 to 2003 in Southern New Mexico. *Remote Sensing of Environment* **93**(1-2), 198-210, doi:10.1016/j.rse.2004.07.011.
- LESSCHEN, J. P., ELBERSEN, B. S., HAZEU, G W., VAN DOORN, A., MUCHER, C. A. & VELTHOF, G. L., 2014: Defining and Classifying Grasslands in Europe - Task 1. Wageningen.
- MILTON, E. J., 1987: Principles of Field Spectroscopy. *International Journal of Remote Sensing* **8**(12), 1807-1827, doi:10.1080/01431168708954818.
- MIRIK, M., CHAUDHURI, S., SURBER, B., ALE, S. & R. ANSLEY, R. J., 2013: Detection of Two Intermixed Invasive Woody Species Using Color Infrared Aerial Imagery and the Support Vector Machine Classifier. *Journal of Applied Remote Sensing* **7**(1), 073588. doi:10.1117/1.jrs.7.073588.

- MUTANGA, O., SKIDMORE, A. K. & PRINS, H. H.T., 2004: Predicting in Situ Pasture Quality in the Kruger National Park, South Africa, Using Continuum-Removed Absorption Features. *Remote Sensing of Environment* **89**(3), 393-408, doi:10.1016/j.rse.2003.11.001.
- NUMATA, I., 2011: Characterization on Pastures Using Field and Imaging Spectrometers. In *Hyperspectral Remote Sensing of Vegetation*, edited by Thekabail Prasad S, Lyon John G, and Huete Alfredo, 207-225. Florida, USA: CRC Press.
- PULLANAGARI, R., KERESZTURI, G. & YULE, I., 2018: Integrating Airborne Hyperspectral, Topographic, and Soil Data for Estimating Pasture Quality Using Recursive Feature Elimination with Random Forest Regression. *Remote Sensing* **10** (7), doi:10.3390/rs10071117.
- R CORE TEAM, 2019: R: A Language and Environment for Statistical Computing. Vienna, Austria. <https://www.r-project.org/>.
- ROYIMANI, L., MUTANGA, O., ODINDI, J., DUBE, T. & MATONGERA, T. N., 2018: Advancements in Satellite Remote Sensing for Mapping and Monitoring of Alien Invasive Plant Species (AIPs). *Physics and Chemistry of the Earth, Parts A/B/C* **112**, 237-245, doi:10.1016/j.pce.2018.12.004.
- SAFARI, H., FRICKE, T. & WACHENDORF, M., 2016: Determination of Fibre and Protein Content in Heterogeneous Pastures Using Field Spectroscopy and Ultrasonic Sward Height Measurements. *Computers and Electronics in Agriculture* **123**, Elsevier B.V., 256-263. doi:10.1016/j.compag.2016.03.002.
- SINGH, L., MUTANGA, O., MAFONGOYA, P. & PEERBHAY, K., 2017: Remote Sensing of Key Grassland Nutrients Using Hyperspectral Techniques in KwaZulu-Natal, South Africa. *Journal of Applied Remote Sensing* **11**(3), 036005, doi:10.1117/1.JRS.11.036005.
- WACHENDORF, M., 2018: Advances in Remote Sensing for Monitoring Grassland and Forage Production. In *Improving Grassland and Pasture Management in Temperate Agriculture*, edited by Athole Marshal and Rosemary Collins. Cambridge: Burleigh Dodds Science Publishing.
- WIJESINGHA, J., MOECKEL, T., HENSGEN, F., & WACHENDORF, M., 2019: Evaluation of 3D Point Cloud-Based Models for the Prediction of Grassland Biomass. *International Journal of Applied Earth Observation and Geoinformation* **78**, Elsevier: 352-359, doi:10.1016/J.JAG.2018.10.006.
- YUAN, H., YANG, G., LI, C., WANG, Y., LIU, J., YU, H., FENG, H., XU, B., ZHAO, X. & YANG, X., 2017: Retrieving Soybean Leaf Area Index from Unmanned Aerial Vehicle Hyperspectral Remote Sensing: Analysis of RF, ANN, and SVM Regression Models. *Remote Sensing* **9**(4), doi:10.3390/rs9040309.

# SpectroAnalyst – Management von Hyperspektraldaten für die Grünlandforschung

ANDREAS SCHAUMBERGER<sup>1</sup> & MANUEL ADELWÖHRER<sup>1</sup>

*Zusammenfassung: Die Messung der Entwicklung von Pflanzenbeständen setzt nicht destruktive Beobachtungen voraus. Dabei stellt die Reflexionsspektroskopie eine Methode dar, mit der viele Vegetationscharakteristiken aus dem elektromagnetischen Spektrum des sichtbaren und nahen Infrarotbereiches ausgewertet werden. Wirtschaftsgrünland entwickelt sich in mehreren, durch Schnitte voneinander getrennten, relativ kurzen Wuchsperioden, in denen die vertretenen Arten den Pflanzenbestand unterschiedlich stark dominieren. Für Grünland braucht es deshalb dichte Zeitreihen mit vielen Messwiederholungen. Schnell fallen dabei mehrere tausend Spektralsignaturen an. Das hier vorgestellte Werkzeug SpectroAnalyst unterstützt den gesamten Arbeitsprozess von der Speicherung über Rohdatenverarbeitung und Korrektur bis hin zur Berechnung von Vegetationsindizes.*

## 1 Einleitung

Die Pflanzenbestände einer Grünlandfläche entwickeln sich in Abhängigkeit der Nutzungsintensität und der Standortbedingungen sehr unterschiedlich. Anders als im Ackerbau besteht das Grünland aus mehreren Arten mit verschiedenen Ansprüchen und Entwicklungsphasen und wird meist mehrmals im Jahr geerntet. Sowohl die Art der Bewirtschaftung als auch der Standort mit Klima, Boden und Topographie wirken auf die Zusammensetzung des Pflanzenbestandes und in weiterer Folge auf den Ertrag und die Futterqualität (PÖTSCH 2009). Ein zentrales Thema der Grünlandforschung ist die Wirkung dieser Einflussfaktoren auf die Leistungsfähigkeit eines Grünlandbestandes. Versuchsernten im Rahmen wissenschaftlicher Feldexperimente werden meist in Abstimmung mit der zu untersuchenden Schnitthäufigkeit, also wenige Male pro Jahr vorgenommen und das so gewonnene Pflanzenmaterial im Labor untersucht. Für die Grünlandwirtschaft können auf diese Weise Aussagen über den Zusammenhang zwischen Höhenstufen und Futterertrag (BUCHGRABER & PÖTSCH 1999), Schnittzeitpunkt und Futterinhaltsstoffe (BUCHGRABER 1997), Bewirtschaftungsintensität und Stickstoffkreislauf, Bewirtschaftungssysteme und Futterqualität (PÖTSCH et al. 2006) sowie Düngung und Futterinhaltsstoffe (PÖTSCH 1995) getroffen werden, um einige Beispiele zu nennen.

Eine kontinuierliche, nicht destruktive Beobachtung von Pflanzenbeständen während eines Aufwuchses ermöglicht im Gegensatz zur Ernte die Erfassung von Entwicklungsparametern, da hier die oberirdische Biomasseentwicklung nicht beeinflusst wird. Neben traditionellen Erhebungen in Grünlandexperimenten wie zum Beispiel die Schätzung der Artengruppenverhältnisse, phänologische Beobachtungen oder Höhenmessungen des Pflanzenbestandes, bietet die Reflexionsspektroskopie unter Einsatz eines Feldspektrometers eine weit darüber hinaus gehende Fülle an strukturellen Informationen. Der hohe Informationsgehalt von Spektralsignaturen wird unter

---

<sup>1</sup> Höhere Bundeslehr- und Forschungsanstalt für Landwirtschaft Raumberg-Gumpenstein, Altdrning 11, A-8952 Irdning-Donnersbachtal, E-Mail: andreas.schaumberger@raumberg-gumpenstein.at

anderem dazu genutzt, Modelle zur Ertrags- und Qualitätsschätzung zu entwickeln bzw. Vegetationskennwerte abzuleiten (vgl. BIEWER et al. 2009; DORIGO 2008; KLINGLER et al. 2019; PSOMAS et al. 2011; PUNALEKAR et al. 2018; THENKABAIL et al. 2014). Mit der Umformung von Hyperspektraldaten in Sentinel-2 entsprechende Multispektraldaten mit Hilfe von Spectral Response Functions (S2-SRF) (ESA 2018) besteht zudem die Möglichkeit, Modelle zu entwickeln, die durch Einbindung von Sentinel-2-Szenen regional anwendbar sind. Auf diese Weise können Grünlandexperimente als Datengrundlage verwendet werden, deren Parzellengröße in der Regel nicht ausreicht, um darauf Sentinel-2-Daten sinnvoll auszuwerten und für die Modellierung zu verwenden.

Um die einzelnen Entwicklungsschritte eines Pflanzenbestandes differenzieren und sie nach Möglichkeit mit Umwelteffekten, wie z. B. der Witterung, in Beziehung setzen zu können, ist eine Beobachtung in kurzen Zeitabständen notwendig. Zudem dauert einer von meist mehreren Grünlandaufwüchsen nur wenige Wochen, in denen eine relativ rasche Entwicklung erfolgt. Wird auch die räumliche Variabilität eines Pflanzenbestandes erfasst, müssen innerhalb der beobachteten Flächen Messungen an möglichst vielen Positionen durchgeführt werden. Die hohe zeitliche und räumliche Beobachtungsdichte mit einem Feldspektrometer führt zwangsläufig und sehr rasch zu einer erheblichen Datenmenge. Die Verarbeitung der Rohspektren, deren Verwaltung und Auswertung erfordert ab einem gewissen Umfang zwangsläufig technische Unterstützung, ohne die eine effektive und effiziente Arbeit mit Hyperspektralsignaturen nicht mehr möglich ist. Der hier als Beispiel vorgestellte SpectroAnalyst ist ein Managementwerkzeug, das dem Benutzer erlaubt, die gesamte Spektraldatenverarbeitung mit möglichst geringem Aufwand zu bewerkstelligen.

## 2 Problemstellung

Mit dem Feldspektrometer können innerhalb kürzester Zeit viele Beobachtungen durchgeführt werden, da der Messaufwand bei günstigen Witterungsbedingungen vergleichsweise gering ist. Vor allem bei komplexen Grünlandversuchen, in denen die Entwicklung von Pflanzenbeständen innerhalb kurzer Zeitabstände beobachtet werden soll, steigt die Zahl der erfassten Spektren kontinuierlich an. Eine manuelle Verarbeitung der Daten würde dabei viel Zeit und Energie erfordern. Ausgehend von den Rohdaten bis hin zur Anwendung der hyperspektralen Daten sind viele Arbeitsschritte notwendig, die ohne softwaretechnische Unterstützung und konsistentem Datenmanagement kaum zu bewältigen sind. Vor allem dann, wenn eine allgemeine Spektraldatenbank als Infrastruktur für beliebige Versuche und verschiedene Projekte entstehen soll.

Ein wichtiger Aspekt für die langfristige Verwertbarkeit von Daten ist deren Verknüpfung mit Metadaten. Sie spielen eine entscheidende Rolle bei der Beurteilung der Qualität von Spektren, indem sie Einblicke in die Entstehungsgeschichte geben (RASAIAH et al. 2015). Ohne Metadaten ist weder eine systematische Speicherung der Spektren noch eine Kriterien basierte Selektion und Auswertung möglich. Neben der Bereitstellung von allgemeinen Informationen über Messzeitpunkt, Wetter, Art des Pflanzenbestandes, seine phänologische Entwicklungsphase sowie Messpersonal muss jede Spektralsignatur einem Experiment und Messgerät zugeordnet werden können. In der Regel bestehen Grünlandversuche aus Varianten mit mehrfachen Wiederholungen. Auch hier ist eine entsprechende Zuordnung der einzelnen Spektren notwendig, um Auswertungen nach Faktoren und Faktorstufen effizient durchführen zu können. Zusätzlich zur Beschreibung

empfehlenswert, fotografische Aufnahmen der gemessenen Oberfläche in die Metadaten einzugliedern; sie erleichtern bei der späteren Datenanalyse die Identifikation unerklärbarer Abweichungen und Ausreißer in den Spektren (MACARTHUR 2011).

Um eine vollständige Verknüpfung zwischen Metadaten, Rohdaten, weiter verarbeiteten Daten und den daraus abgeleiteten Statistiken und Indizes zu realisieren, ist der Einsatz eines Datenbankmanagementsystems sinnvoll und notwendig. Eine wichtige Anforderung an Datenbanken im Allgemeinen und einer Spektraldatenbank im Besonderen ist die Skalierbarkeit, um die mit der Zeit wachsende Datenmenge weiterhin effizient bewältigen zu können. Ebenso wichtig ist eine umfassende Erweiterbarkeit des Systems, ohne Änderungen an den Strukturen vornehmen zu müssen. Bei der Implementierung eines relationalen Datenschemas ist in diesem Zusammenhang auf die vollständige Einhaltung der Normalisierungsregeln zu achten. Eine Anlehnung an Richtlinien und Anforderungsdefinitionen für wissenschaftliche Datenbanken, wie sie unter anderem PORTER (2000), VAN EVERT et al. (1999a), VAN EVERT et al. (1999b) oder WIGGINS et al. (2013) vorschlagen, ist hilfreich und vermindert die Wahrscheinlichkeit, nachträgliche und mit hohem Aufwand verbundene Änderungen des Datenschemas vornehmen zu müssen.

Um Spektralsignaturen evaluieren zu können, ist deren Visualisierung als Graph von großem Vorteil. Neben einem geeigneten Datenmodell ist daher die Benutzerschnittstelle, welche sowohl die grafische Ausgabe der Spektren als auch die Eingabe von Daten unterstützt, eine wichtige Voraussetzung für die effiziente Verwaltung und Bearbeitung von Spektren. Im SpectroAnalyst wurde der Programmoberfläche große Aufmerksamkeit geschenkt, da nur mit ihrer Hilfe ein gemeinsamer Zugriff auf Spektraldaten und all ihren Verknüpfungen (Metadaten, Sensoren, Varianten, usw.) möglich ist. In den folgenden Abschnitten wird das Konzept dieses Werkzeuges vorgestellt. Es soll beispielhaft zeigen, wie die Ansprüche an eine wissenschaftliche Datenbank mit den Anforderungen an das Management von Hyperspektraldaten kombiniert werden.

### **3 Material und Methoden**

#### **3.1 Datenmodell**

Bei der Entwicklung einer Software für das Management von Hyperspektraldaten muss im Vorfeld entschieden werden, ob die Spektren dateibasiert oder in einer Datenbank gespeichert werden. Ein dateibasiertes System muss mit Funktionen ausgestattet sein, die eine Zusammenstellung und dateiübergreifende Auswertung unterstützen. Beispiele für diese Art der Verwaltung sind HyperCor von AASEN et al. (2014) oder 3S von MERONI & COLOMBO (2009). Die Software GRAMS von ThermoFisher Scientific geht einen Schritt weiter und verknüpft die Oberfläche für Management und Auswertung mit einer Datenbank, in der die Spektren vorgehalten werden. Hier wird bereits eine direkte Beziehung von Metadaten und einzelnen Spektren realisiert und zudem eine breite Palette an Auswertungsmöglichkeiten angeboten. Die Entscheidung, dennoch ein eigenes System wie den SpectroAnalyst zu entwickeln, entstand aus der Notwendigkeit heraus, den gesamten Workflow von der Rohdatenverarbeitung bis hin zur Auswertung in ein System zu integrieren und dabei jeden Arbeitsschritt zusammen mit den Rohdaten im Sinne einer umfassenden Nachvollziehbarkeit zu speichern. Das Erreichen eines Höchstmaßes an Flexibilität

und die Möglichkeit einer beliebigen Erweiterbarkeit von Funktionen waren dabei die wichtigsten Motive. Anforderungen an das System und speziell an das Datenmodell sind dabei folgende:

- Der Import verschiedener proprietärer Formate (z. B. ASD FieldSpec, tec5 HandySpec, OceanOptics, Carl Zeiss Corona, usw.) mit unterschiedlichen Wellenlängenbereichen (VNIR, SWIR1, SWIR2) in eine einheitliche, homogene Datenstruktur wird unterstützt.
- Die Rohdaten werden im Originalzustand gespeichert und können jederzeit für eine neuerliche Datenverarbeitung bereitgestellt werden.
- Zu jeder Spektralsignatur werden Metainformationen gespeichert.
- Spektren werden Experimenten und in weiterer Folge einzelnen Varianten zugeordnet.
- Glättungen, Interpolationen und Korrekturen (Detector Offset) werden getrennt gespeichert und können jederzeit entfernt bzw. geändert werden.
- Spektralsignaturen können geglättet und zur Erstellung von Mittelwertspektren beliebig zusammengefasst werden.
- Die Gleichungen für die Berechnung von Vegetationskennwerte werden in der Datenbank abgelegt und können über die Benutzerschnittstelle auf beliebige Spektren angewendet werden.

Um die hier angeführte Funktionalität zu unterstützen, wurde das folgende relationale Datenmodell (Abb. 1) entwickelt und im Datenbankmanagementsystem MS SQL Server implementiert:

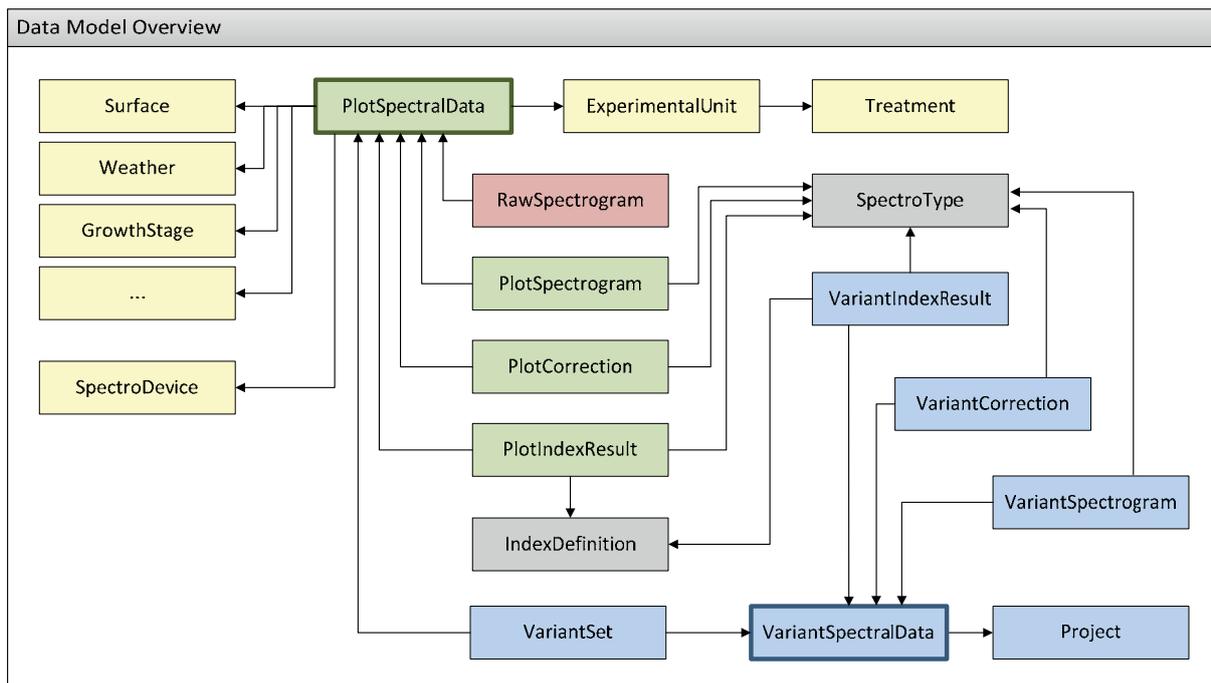


Abb. 1: Vereinfachtes Datenmodell des SpectroAnalyst auf Relationenebene

Die mit den verschiedenen Feldspektrometern erfassten Daten werden in einer schlanken Tabelle *RawSpectrogram* mit nur vier Feldern gespeichert. Diese vier Felder beinhalten ein Indikator-Attribut (ID), das auf die Tabelle *PlotSpectralData* verweist, eine weitere ID, welche alle Reflexionswerte einer einzelnen Spektralsignatur als zusammengehörend kennzeichnet, die

Wellenlänge sowie den Reflexionswert selbst. Indem für diese Tabelle ein dreifach zusammengesetzter Primärschlüssel mit der *PlotSpectralData*-ID als Fremdschlüssel, der Rohsignatur-ID sowie dem Wellenlängenwert definiert ist, können sämtliche gespeicherte Reflexionswerte exakt zugeordnet werden. Auf diese Weise ist es möglich, Spektren aus verschiedenen Quellen und Wellenlängenbereichen in einer gemeinsamen Struktur zu speichern.

Werden beispielsweise auf einer Versuchsparzelle Spektren in vierfacher Wiederholung gemessen, ergeben sich für die Eingabe in die Datenbank folgende Arbeitsschritte: Zunächst wird ein neuer Datensatz in der Tabelle *PlotSpectralData* mit einer laufenden ID angelegt. Dieser Datensatz enthält eine Reihe von Fremdschlüsselattributen, die auf Tabelleneinträge verweisen, in denen Metadaten definiert sind (Wetter, Oberfläche, phänologische Phase, usw.). Ebenso wird auf die Parzellendefinition (*ExperimentalUnit*) mit der jeweiligen dort stattfindenden Behandlung (*Treatment*) verwiesen. Ein Tupel in *PlotSpectralData* bildet sozusagen die Klammer über Spektral- und Metainformationen. Die Tabelle *RawSpectrogram* enthält die einzelnen Reflexionswerte in einer jeweils eigenen Zeile. Alle vier erfassten Spektralsignaturen verweisen auf das gleiche Tupel in *PlotSpectralData*, damit werden sie als zusammengehörende Wiederholungen gekennzeichnet. Die vier Spektren selbst unterscheiden sich durch die in einem weiteren Feld gesetzte *RawSpectrogram*-ID, indem alle Reflexionswerte einer Spektralsignatur die gleiche ID erhalten. Innerhalb des Spektrums sind die Einzelwerte der jeweiligen Wellenlänge zugeordnet.

Um aus Messwiederholungen eine für die erfasste Versuchsfäche repräsentative Spektralsignatur zu generieren, werden die vier Spektren gemittelt und als Mittelwertsignatur in die Tabelle *PlotSpectrogram* gespeichert. Sie ist ebenso wie *RawSpectrogram* mit nur wenigen Feldern ausgestattet: Ein Fremdschlüsselattribut verweist auch hier auf *PlotSpectralData*, mit dem Effekt, dass die zunächst gesetzte Klammer um Rohspektren und Metainformationen auf alle davon abgeleiteten Daten erweitert wird. In *PlotSpectrogram* werden nicht nur Mittelwertspektren aus den Wiederholungen, sondern auch weitere Ableitungen gespeichert. Aus diesem Grund wird der Primärschlüssel dieser Tabelle ebenfalls aus mehreren Attributen zusammengesetzt und enthält neben der Wellenlänge auch eine Identifikation des Typs, indem mit einem ID-Feld auf die Tabelle *SpectroType* verwiesen wird. Auf diese Weise können Mittelwertspektren, geglättete Spektren, trendbereinigte Spektren, Spektren aus statistischen Kennzahlen, usw. voneinander unterschieden und in einer gemeinsamen Tabelle mit einheitlicher Struktur abgelegt werden. Über die *PlotSpectralData*-ID bleiben sie mit den dazugehörigen Basisdaten verknüpft.

Im Übergangsbereich verschiedener Sensoren kommt es häufig zu kleineren und größeren Verschiebungen (Detector Offset). Der SpectroAnalyst bietet die Möglichkeit, diesen Versatz zu eliminieren und die Spektralsignatur in den betroffenen Wellenlängenbereichen zu korrigieren. In der Tabelle *PlotCorrection* werden die berechneten Korrekturfaktoren gespeichert und den jeweiligen Spektren zugeordnet. Gespeichert werden keine korrigierten Signaturen, sondern lediglich Faktoren, die über definierte Wellenlängenbereiche additiv oder multiplikativ bei jedem Zugriff auf die Originaldaten dynamisch angebracht werden.

In der SpectroAnalyst-Datenbank werden neben Spektren und Metadaten auch Index-Definitionen in der Tabelle *IndexDefinition* gespeichert. Sie enthält eine Spalte mit allen Formeln der erfassten Indizes, die in der Benutzeroberfläche angezeigt und bei entsprechender Auswahl auf beliebige Spektren angewendet werden können. Der Vorteil dieser Vorgehensweise besteht darin, dass die Indexliste über die Datenbank beliebig erweitert werden kann und über die Benutzeroberfläche für

die Anwendung sofort zur Verfügung steht. In der Tabelle *PlotIndexResult* werden die Berechnungsergebnisse abgelegt und auch hier besteht über die *PlotSpectralData*-Tabelle eine direkte Verknüpfung mit allen dazugehörigen Spektral- und Metainformationen.

Die in Abb. 1 in blauer Farbe markierten Relationen weisen die gleiche Funktionalität wie die grün markierte Plot-Struktur auf, beziehen sich jedoch auf Spektren, die zu Varianten zusammengefasst wurden. In der Tabelle *VariantSet* sind beliebig viele *PlotSpectralData* miteinander verbunden und können über *VariantSpectralData* auf die gleiche Weise wie die Spektren auf Parzellenebene verarbeitet und ausgewertet werden. Über die Beziehung zu *PlotSpectralData* bleiben die Varianten weiterhin mit den Rohdaten in *RawSpectrogram* verbunden, wobei die Tabelle *VariantSpectrogram* die davon abgeleiteten Spektren enthält.

### 3.2 Korrektur der Rohspektren

Bevor Hyperspektraldaten ausgewertet und verwendet werden, müssen die Rohdaten einer Kontrolle und gegebenenfalls einer Bearbeitung unterzogen werden. Im SpectroAnalyst werden zu diesem Zweck alle neu importierten Spektren in ihrem Originalzustand grafisch dargestellt. Weichen einzelne Messwiederholungen innerhalb einer beobachteten Versuchsparzelle stark voneinander ab, wird in der Regel das vom Messzeitpunkt verfügbare Bild des Pflanzenbestandes herangezogen, um mögliche Ausreißer innerhalb einer bestimmten Parzelle zu identifizieren und diese vor der Bildung eines Mittelwertspektrums auszuschließen.

In einem nächsten Schritt wird die Korrektur eines möglicherweise aufgetretenen Detector Offsets durchgeführt. Eine Verschiebung der Signatur im Übergangsbereich verschiedener im Feldspektrometer verbauter Sensoren, wie beispielsweise bei etwa 1000 nm von VNIR zu SWIR1, ist ein immer wieder auftauchendes Problem und sollte bei der Rohdatenbearbeitung unbedingt berücksichtigt werden (AASEN et al. 2014; DORIGO 2008; DORIGO et al. 2006; GARFAGNOLI et al. 2013; MACARTHUR et al. 2012; MILTON et al. 2009). Die Ursachen dafür sind nach eigenen Erfahrungen mit den HandySpec von tec5 nicht exakt zu identifizieren, dürften nach HUENI & BIALEK (2017) aber ihren Grund im unterschiedlich starken Einfluss von Umweltbedingungen (Temperatur und Luftfeuchte) auf die verschiedenen Sensoren haben. Aufgrund des häufigen Auftretens dieser Situation wird mit dem SpectroAnalyst jeder Detector Offsets vor der Bildung des Mittelwertspektrums korrigiert. Die Herausforderung dabei ist, eine vom Messrauschen in diesem Übergangsbereich unabhängige Differenz zwischen den beiden Spektralbereichen zu ermitteln. Dabei wird folgender Algorithmus angewendet: Zunächst erfolgt eine Savitzky-Golay-Glättung in einer variablen Breite (smoothing region) und Polynom 2. Grades (SAVITZKY & GOLAY 1964) für jenen Bereich des Spektrums, dessen Niveau an das des benachbarten angeglichen werden soll. Auf dem letzten Abschnitt der geglätteten Spektralsignatur, also unmittelbar vor dem Sensorübergang, wird eine Forecast- bzw. Trend-Funktion angewendet, um den Basiswert für die anschließende Korrektur zu ermitteln. Dieser Basiswert wird entweder absolut oder relativ in Beziehung zum unmittelbar benachbarten Wert des nicht zu korrigierenden Spektralbereiches gesetzt und so ein Korrekturwert ermittelt, der in Abhängigkeit der Benutzerangaben additiv oder multiplikativ auf den zu korrigierenden Spektralbereich angewendet wird. Es besteht die Möglichkeit, VNIR an SWIR1 oder SWIR1 an VNIR anzupassen und, als dritte Option, beide Bereiche zu gleichen Teilen aufeinander hin zu korrigieren. Bevorzugt erfolgt die

Korrektur multiplikativ, um sie im Verhältnis zur Höhe des Reflexionswertes anzubringen. Beim tec5-HandySpec sollte laut Herstellerangabe der SWIR1-Bereich unverändert gehalten und andere mit dem Detector Offset abgesetzte Spektralbereiche darauf hin korrigiert werden.

### 3.3 Glättung der Rohspektren

Nachdem die Rohspektren korrigiert und um den Detector Offset bereinigt wurden, wird die Spektralsignatur geglättet, um Messrauschen weitgehend zu eliminieren (KING et al. 1999; RUFFIN & KING 1999). Damit bei dieser Prozedur spektrale Informationen bestmöglich erhalten bleiben, wurde eine dynamische Savitzky-Golay-Glättung nach BARAK (1995) implementiert, bei der abschnittsweise Polynome verschiedenen Grades in Abhängigkeit der Kurvencharakteristik mit mehrfacher Iteration angewendet werden.

Bei den mit dem HandySpec (tec5) gemessenen Spektralsignaturen von 400 bis 1690 nm wird der Bereich der Wasserabsorption zwischen 1350 bis 1450 nm sowie jener am Ende des Messbereiches ab 1650 nm bei der dynamischen Glättung nicht berücksichtigt. Hier kann der SpectroAnalyst-Benutzer optional eine kubische Spline-Interpolation aus der Programmbibliothek ALGLIB anwenden, um die am Ende des Messbereichs vermehrt auftretenden Schwingungen bestmöglich zu glätten und im Wasserabsorptionsband einen ebenfalls geglätteten Übergang zu schaffen. Dieser Bereich zwischen 1350 und 1450 nm ist für eine Auswertung zwar nicht relevant, schafft jedoch durch die Interpolation aus optischen Gründen und alternativ zum Herausschneiden dieses Kurvenabschnittes einen glatten Übergang zwischen den beiden relevanten SWIR1-Bereichen 1000 bis 1350 und 1450 bis 1690.

### 3.4 Werkzeuge zur Analyse und Auswertung

Für sämtliche gespeicherte Spektren können Vegetationskennwerte berechnet werden, deren Definitionen beim Start des SpectroAnalyst aus der Datenbank eingelesen werden und dem Benutzer über die Programmoberfläche zur Verfügung stehen. Die Berechnungen können auf die originalen bzw. korrigierten und/oder auf die noch zusätzlich geglätteten Spektralsignaturen angewendet werden.

Im Zuge der Speicherung der korrigierten Spektren (vgl. Abschnitt 3.2) werden automatisch Sentinel-2-Bänder (S2-SRF) (ESA 2018) generiert und in einer eigenen Tabelle abgelegt. Auf diese Weise entsteht neben den Hyperspektraldaten auch ein multispektraler Datensatz, der mit einem Bestimmtheitsmaß von 99 % (KLINGLER et al. 2019) jenen von Sentinel-2-Bändern entspricht und daher in satellitenbasierten, regionalen Modellen Anwendung finden kann. So fließt zum Beispiel dieser Datensatz in ein neuronales Netz zur Berechnung des LAI nach BARET et al. (2010) ein, das mit den Strahlungstransfermodellen PROSPECT und SAIL (JACQUEMOUD et al. 2009) speziell auf Sentinel-2 abgestimmt wurde. Ausgeführt wird das externe Programm zur Berechnung des LAI über eine vom SpectroAnalyst aufgerufene MATLAB-Runtime. Dabei werden acht S2-SRF-Bänder (B3, B4, B5, B6, B7, B8A, B11 und B12) sowie Sonnenwinkelparameter, die in Abhängigkeit des Zeitstempels der Spektren in R (Package solarpos) berechnet wurden, mitgegeben.

Für ein beliebig zusammenstellbares Set von Spektren besteht die Möglichkeit, statistische Auswertungen zu rechnen. Realisiert wird dies über die im SpectroAnalyst implementierte Schnittstelle zum Statistikprogramm R und der Anwendung des Packages `prospectr` (STEVENS & LOPEZ 2014). Auf diese Weise können gegebenenfalls weitere R-Packages integriert und der Funktionsumfang des SpectroAnalyst angepasst und erweitert werden.

## 4 Implementierung

### 4.1 Systemkomponenten

Das Softwaresystem des SpectroAnalyst entspricht dem Konzept einer Dreischichtenarchitektur. Auf Ebene der Datenhaltung wurde ein relationales Datenmodell im Datenbankmanagementsystem MS SQL Server implementiert und auf die im Abschnitt 3.1 dargestellte Weise strukturiert. Die Art und Weise der Speicherung von Spektren und der damit verbundenen und daraus abgeleiteten Daten ist die Voraussetzung für die im SpectroAnalyst implementierte Funktionalität. Die Anwendungslogik befindet sich in MS C# erstellten Klassen des .NET-Frameworks und ist über ADO.NET Data Provider Objects mit der Datenhaltungsebene verbunden. Neben den eigenen Klassen wird weitere Funktionalität über externe Prozesse eingebunden. Das für die Auswertung von Spektraldaten entwickelte R-Package `prospectr` wird beispielsweise über einen R-Prozess ausgeführt, welcher mit der Bibliothek R.NET über NuGet in den SpectroAnalyst eingebunden wurde. Auf diese Weise können auch weitere Packages verwendet und der SpectroAnalyst effizient erweitert werden. Die vom Benutzer in der Anwendungsoberfläche ausgewählten Spektren werden in einer zweidimensionalen Array-Struktur an den R-Prozess übergeben, dort verrechnet und die Ergebnisse mit gleicher Struktur wieder entgegengenommen und gespeichert. Ebenso wie R können auch MATLAB-Algorithmen über Einbindung der MATLAB-Runtime und deren Aufruf mittels `PROCESS`-Objekt verwendet werden. Im SpectroAnalyst wird die Berechnung des LAI aus den Sentinel-2-Multispektraldaten auf diese Weise durchgeführt.

Die Präsentationsschicht als Benutzerschnittstelle wurde mit C#.NET GUI-Klassen erstellt und bietet dem Benutzer die Möglichkeit, geladene Spektren dynamisch zu visualisieren, Kurvenausschnitte zu vergrößern und zu verkleinern sowie Werte abzufragen. Neben der grafischen Darstellung kann im linken Bereich (siehe Abb. 2) navigiert und zwischen den Aggregationsebenen (Plot und Variant) gewechselt werden. Die Oberfläche bietet die Möglichkeit, Spektraldaten zusammen mit Metainformationen zu importieren, aus der Datenbank nach Meta-Kriterien aufzurufen und zu exportieren. Ebenso ist der Export der mittels SRF berechneten Sentinel-2-Bänder sowie der Indizes als Tabellen vorgesehen. Die Benutzeroberfläche erlaubt die Auswahl einzelner Spektren, um sie beispielsweise aus einer Zusammenstellung für die Mittelwertbildung herauszunehmen bzw. überhaupt aus der Datenbank zu löschen. Mit „Show Picture“ kann für die Erhebung auf einer Versuchsparzelle das zugehörige Foto eingeblendet werden, um eine Evaluierung der Spektralsignaturen zu erleichtern. Für die Metadateneingabe und -änderung, die Optionen für die Rohdatenverarbeitung (Korrektur, Mittelwertbildung, Glättung) sowie die Optionen für die Index-Berechnung und für die statistische Analyse (`prospectr`) können über Subfenster ein- und ausgeblendet werden. Abb. 2 zeigt die wichtigsten Elemente der Oberfläche.

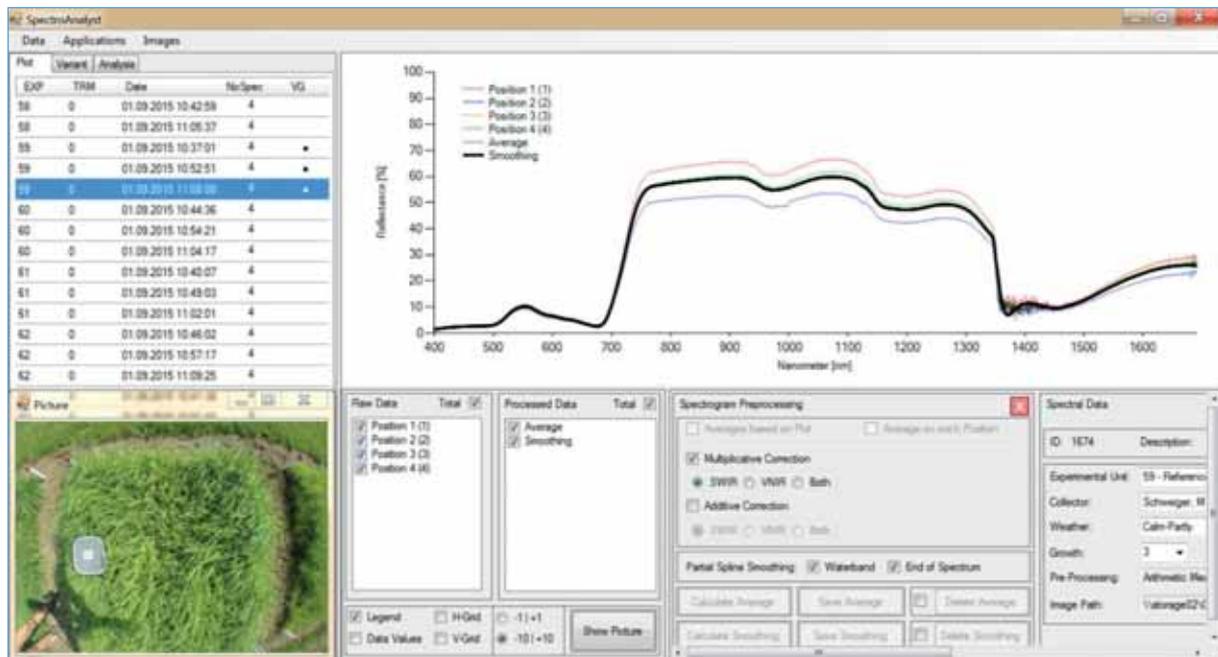


Abb. 2: Benutzeroberfläche des SpectroAnalyst

## 4.2 Erfassung und Abfrage von Daten

Beim Import neuer Spektren wird über Zeitstempel geprüft, ob diese bereits in der Datenbank vorhanden sind. Ist dies nicht der Fall, werden alle sich innerhalb eines ausgewählten Verzeichnisses befindlichen Spektren importiert und in die Datenbank gespeichert. Beim Import werden auch die Metainformationen vom Benutzer eingegeben, die als Template für alle im aktuellen Import-Set enthaltenen Spektren verwendet, bei Änderungsbedarf für einzelne Spektren aber nachträglich jederzeit angepasst werden können. In den SpectroAnalyst-Settings können einige, meist gleichbleibende Informationen wie Pfadangaben für Import, Feldspektrometer, Versuchsdetails usw. bereits vorgegeben werden. Ebenso enthalten die Settings änderbare Parameter für fixe und dynamische Savitzky-Golay-Glättungen, für Spline-Interpolationen und für die Sensorkorrektur (beispielsweise die Position des Sensorübergangs, welcher gerätespezifisch unterschiedlich sein kann). Wenn eine größere Menge an Spektren importiert wird, können Korrektur und Glättung in einem Batch-Prozess für alle markierten Spektren durchgeführt und damit Zeit gespart werden. Dies ist allerdings nur dann sinnvoll, wenn kaum Ausreißer in den Daten vorhanden sind.

Der Benutzer hat die Möglichkeit, Spektren nach Kriterien wie Datum, Versuch, Feldspektrometer und anderen Metadaten auszuwählen und aus der Datenbank in die Anwendung zu laden. Hier können sämtliche bereits durchgeführte und gespeicherte Verarbeitungsschritte und Dateneingaben wieder rückgängig gemacht und/oder ergänzt werden.

### 4.3 Datenaufbereitung für Auswertungen

Im Navigationsbereich der in Abb. 2 angeführten Spektren sind drei Zeilen in der Spalte VG (= VariantGroup) mit einem Punkt markiert. Sie zeigen an, dass sie Bestandteil einer Variante sind. Die Zusammensetzung von Varianten kann in einem eigenen Fenster verwaltet werden, in dem auch Metadaten auf Variantenebene eingegeben werden. Wird der Reiter im Navigationsfenster auf Variante umgestellt, werden alle zu einer Variante gehörenden Spektren angezeigt und der Benutzer kann sämtliche Funktionen (Mittelwert, Glättung, Indexberechnung, Statistik, usw.) in gleicher Weise wie auf Parzellenebene anwenden und die Ergebnisse in die Datenbank ablegen (vgl. blau markierte Relationen in Abb. 1). Beim Export ist es dem Benutzer möglich, sämtliche Spektren und davon abgeleitete Daten sowohl auf Parzellen- als auch auf Variantenebene auszugeben. Das Format der Exportdateien ist CSV; sie können damit problemlos in Tabellenkalkulations- oder Statistikprogramme für weitere Analysen übernommen werden.

## 5 Fazit & Ausblick

Der SpectroAnalyst wurde mit dem Anspruch entwickelt, ein offenes System zur Verwaltung von Hyperspektraldaten aus verschiedenen Quellen bereitzustellen. In den letzten vier Jahren wurden knapp 28.000 Spektren für mehrere Grünlandexperimente erhoben, verarbeitet und gespeichert. Die systematische Bearbeitung dieser Menge an Spektren erfordert ein Werkzeug, das den Benutzer bei allen Arbeitsschritten bestmöglich unterstützt und den dafür erforderlichen Zeitaufwand auf ein Mindestmaß beschränken kann. Für etwa 200 Spektren (400 bis 1690 nm in 1-nm-Auflösung), das entspricht in etwa jener in der Vegetationsperiode wöchentlich anfallenden Datenmenge, können innerhalb von maximal 30 Minuten alle Metadateneingaben, Korrekturen, die Mittelwertbildung für Parzellen und Varianten sowie die gesamte statistische Auswertung inklusive der Berechnung von ca. 40 verschiedenen Indizes bewerkstelligt und vollständig nachvollziehbar gespeichert werden.

Damit der SpectroAnalyst auch für künftige Forschungsprojekte und Anforderungen gerüstet bleibt, wurde zum einen die Erweiterbarkeit des Datenmodells durch die streng relationale Struktur sichergestellt und zum anderen Schnittstellen zu extern entwickelten Analysewerkzeuge, wie beispielsweise R-Packages oder MATLAB, implementiert.

Der SpectroAnalyst wurde für den Eigenbedarf entwickelt und wird im Rahmen dieser Arbeit als Implementierungsbeispiel vorgestellt. Die Kombination von eigener C#.NET-Implementierung und externen Komponenten wie R, MATLAB, Klassenbibliotheken sowie die Verwendung einer kommerziellen Datenbankmanagementsoftware erfüllt zwar die Anforderungen an ein offenes System und deckt den Bedarf der Entwickler bestmöglich ab, macht aber eine Distribution der Software schwierig. Das primäre Ziel der künftigen Weiterentwicklung des SpectroAnalyst ist die Bereitstellung der Software zur allgemeinen Benutzung.

## 6 Literaturverzeichnis

- AASEN, H., GNYP, M.L., MIAO, Y. & BARETH, G., 2014: Automated Hyperspectral Vegetation Index Retrieval from Multiple Correlation Matrices with HyperCor. *Photogrammetric Engineering & Remote Sensing* **80**(8), 785-795.
- BARAK, P., 1995: Smoothing and Differentiation by an Adaptive-Degree Polynomial Filter. *Analytical Chemistry* **67**(17), 2758-2762.
- BARET, F., WEISS, M., BICHERON, P. & BERTHELOT, B., 2010: Sentinel-2 MSI Products WP1152 Algorithm Theoretical Basis Document for Product Group B, INRA-EMMAH, Avignon, France.
- BIEWER, S., FRICKE, T. & WACHENDORF, M., 2009: Determination of Dry Matter Yield from Legume-Grass Swards by Field Spectroscopy. *Crop Science* **49**(5), 1927-1936.
- BUCHGRABER, K., 1997: Der Einfluss des Schnittzeitpunktes auf den Mineralstoffgehalt des Grünlandfutters. Conference on nutrition of domestic animals „Zadravec-Erjavec-Days“, Radenci, Zivinorejsko-Veterinarski zavod za Pomurje Murska Sobota, 27.-28. Oktober 1997, 57-62.
- BUCHGRABER, K. & PÖTSCH, E.M., 1999: Einfluss der Höhenstufen auf Futterertrag, Futterqualität, Tierbesatz und Milchleistung: Landkalender 2000, 63. Jahrgang, *Leopold Stocker Verlag*, Graz, 133-138.
- DORIGO, W.A., 2008: Retrieving canopy variables by radiative transfer model inversion - a regional approach for imaging spectrometer data. Dissertation, Technische Universität München, Wissenschaftszentrum Weihenstephan, 230 S.
- DORIGO, W.A., BACHMANN, M. & HELDENS, W., 2006: AS Toolbox & Processing of field spectra, User's manual, German Aerospace Center (DLR), Oberpfaffenhofen, 31 S.
- ESA, 2018: Sentinel-2 Spectral Response Functions (S2-SRF), 5 S., [https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/document-library/-/asset\\_publisher/Wk0TKajiISaR/content/sentinel-2a-spectral-responses](https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/document-library/-/asset_publisher/Wk0TKajiISaR/content/sentinel-2a-spectral-responses), (15.01.2020).
- GARFAGNOLI, F., MARTELLONI, G., CIAMPALINI, A., INNOCENTI, L. & MORETTI, S., 2013: Two GUIs-based analysis tool for spectroradiometer data pre-processing. *Earth Science Informatics* **6** (4), 227-240.
- HUENI, A. & BIALEK, A., 2017: Cause, Effect, and Correction of Field Spectroradiometer Interchannel Radiometric Steps. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **10**(4), 1542-1551.
- JACQUEMOUD, S., VERHOEF, W., BARET, F., BACOUR, C., ZARCO-TEJADA, P.J., ASNER, G.P., FRANÇOIS, C. & USTIN, S.L., 2009: PROSPECT + SAIL models: A review of use for vegetation characterization. *Remote Sensing of Environment* **113**, 56-66.
- KING, R.L., RUFFIN, C., LAMASTUS, F.E. & SHAW, D.R., 1999: The analysis of hyperspectral data using Savitzky-Golay filtering-practical issues (Part 2). *Geoscience and Remote Sensing Symposium, IGARSS '99 Proceedings, Hamburg, IEEE*, 1, June 28 - July 2, 1999, 398-400.
- KLINGLER, A., SCHAUMBERGER, A., VUOLO, F. & POETSCH, E.M., 2019: Vergleich unterschiedlicher Methoden zur Schätzung von Blattflächenindex in Dauergrünlandbeständen. Dreiländertagung der DGPF, der OVG und der SGPF, Vienna, DGPF, 28, Februar 20 - 22, 301 - 309.

- MACARTHUR, A., 2011: Introduction to Field Spectroscopy, NERC Field Spectroscopy Facility, Monte Bondone, Italy, Course Handbook - COST Action ES0903, 63 S.
- MACARTHUR, A., MACLELLAN, C.J. & MALTHUS, T., 2012: The Fields of View and Directional Response Functions of Two Field Spectroradiometers. *IEEE Transactions on Geoscience and Remote Sensing* **50**(10), 3892-3907.
- MERONI, M. & COLOMBO, R., 2009: 3S: A novel program for field spectroscopy. *Computers & Geosciences* **35**(7), 1491-1496.
- MILTON, E.J., SCHAEPMAN, M.E., ANDERSON, K., KNEUBÜHLER, M. & FOX, N., 2009: Progress in field spectroscopy. *Remote Sensing of Environment* **113**, S92-S109.
- PORTER, J.R., 2000: Scientific Databases. In MICHENER & BRUNT (Eds.): *Ecological Data: Design, Management and Processing, Methods in Ecology*, Blackwell Science Ltd, 48-69.
- PÖTSCH, E.M., 1995: Einfluss der Düngung und Nutzung des Grünlandes auf den Gehalt an Futterinhaltsstoffen. Conference on nutrition of domestic animals „Zadravec-Erjavec-Days“, Radenci, Zivinorejsko-Veterinarski zavod za Pomurje Murska Sobota, 26.-27. Oktober 1995, 85-95.
- PÖTSCH, E.M., 2009: Einflussfaktoren auf Ertrag und Qualität von Grünlandfutter. Fortbildungsveranstaltung "Tierärztliche Bestandsbetreuung von Milchviehbetrieben in Österreich", Gumpenstein, LFZ Raumberg-Gumpenstein, 4.-6. Juni 2009, 5-13.
- PÖTSCH, E.M., RESCH, R., SCHAUMBERGER, A. & BUCHGRABER, K., 2006: Effect of different management systems on quality parameters of forage from mountainous grassland. 21st EGF General Meeting “Sustainable Grassland productivity”, Badajoz, Spain, *Grassland Science in Europe*, 11, April 3-6, 2006, 484-486.
- PSOMAS, A., KNEUBÜHLER, M., HUBER, S., ITTEN, K.I. & ZIMMERMANN, N.E., 2011: Hyperspectral remote sensing for estimating aboveground biomass and for exploring species richness patterns of grassland habitats. *International Journal of Remote Sensing* **32**(24), 9007-9031.
- PUNALEKAR, S.M., VERHOEF, A., QUAIFFE, T.L., HUMPHRIES, D., BERMINGHAM, L. & REYNOLDS, C.K., 2018: Application of Sentinel-2A data for pasture biomass monitoring using a physically based radiative transfer model. *Remote Sensing of Environment* **218**, 207-220.
- RASAIHAH, B., JONES, S., BELLMAN, C., MALTHUS, T. & HUENI, A., 2015: Assessing Field Spectroscopy Metadata Quality. *Remote Sensing* **7**(4), 4499-4526.
- RUFFIN, C. & KING, R.L., 1999: The analysis of hyperspectral data using Savitzky-Golay filtering-theoretical basis (Part 1). *Geoscience and Remote Sensing Symposium, IGARSS '99 Proceedings, Hamburg, IEEE*, 2, June 28 - July 2, 1999, 756-758.
- SAVITZKY, A. & GOLAY, M.J.E., 1964: Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry* **36**(8), 1627-1639.
- STEVENS, A. & LOPEZ, L.R., 2014: An introduction to the prospectr package, 22 S., <http://antoinestevens.github.io/prospectr/>, (15.01.2020).
- THENKABAIL, P.S., GUMMA, M.K., TELUGUNTLA, P. & MOHAMMED, I.A., 2014: Hyperspectral Remote Sensing of Vegetation and Agricultural Crops. *Photogrammetric Engineering & Remote Sensing* **80**(8), 697-709.
- VAN EVERT, F.K., SPAANS, E.J.A., KRIEGER, S.D., CARLIS, J.V. & BAKER, J.M., 1999a: A Database for Agroecological Research Data: I. Data Model. *Agronomy Journal* **91**(1), 54-62.

- VAN EVERT, F.K., SPAANS, E.J.A., KRIEGER, S.D., CARLIS, J.V. & BAKER, J.M., 1999b: A Database for Agroecological Research Data: II. A Relational Implementation. *Agronomy Journal* **91**(1), 62-71.
- WIGGINS, A., BONNEY, R., GRAHAM, E., HENDERSON, S., KELLING, S., LEBUHN, G., LITTAUER, R., LOTT, K., MICHENER, W., NEWMAN, G., RUSSELL, E., STEVENSON, R. & WELTZIN, J., 2013: *Data Management Guide for Public Participation in Scientific Research*. GROUP, University of New Mexico, Albuquerque, USA, 15 S.

# Neue ISO-Norm für Langzeitarchivierung von Fernerkundungsdaten

WOLFGANG KRESSE<sup>1</sup>

*Zusammenfassung: Vor allem auf Betreiben der NASA ist eine Norm für die Langzeitarchivierung von Fernerkundungsdaten entstanden. Die Norm hat den langen Titel „Preservation of digital data and metadata — Part 2: Content specifications for Earth observation data and derived digital products“. Diese Norm baut auf zwei Basisnormen auf, nämlich der „ISO 19165-1:2018, Geographic information — Preservation of digital data and metadata — Part 1: Fundamentals“ und der „ISO 14721:2012, Space data and information transfer systems — Open archival information system (OAIS) — Reference model“. Derzeit läuft die Abstimmung zum Draft International Standard, die zum Zeitpunkt der Jahrestagung abgeschlossen sein wird. Inhaltlich folgt die Norm weitgehend einem offenbar bei amerikanischen Stellen üblichen Arbeitsfluss: Konzeptuelle Vorbereitung, Modellierung, Implementierung, Durchführung und Nachbereitung.*

## 1 Einleitung

Wohl jeder hat inzwischen schon einmal eine Situation erlebt, in der Daten verloren gegangen sind, weil der Datenträger nicht mehr lesbar war und gleichzeitig eine vorherige Sicherung versäumt worden war. Diese unerwünschte Situation kann nicht nur im privaten Bereich entstehen, sondern in viel größerem Maßstab auch im professionellen Umfeld. Geodaten haben den wahrscheinlich größten Umfang im Vergleich zu Datensätzen anderer Fachgebiete. Daher ist eine weitere Frage zu beantworten, die sonst seltener gestellt wird: Da die Datenmenge schneller als der verfügbare Speicherplatz wächst, muss ausgewählt werden, welche Daten langfristig erhalten werden können und welche anderen Daten wann wieder gelöscht werden dürfen.

Bei Gesprächen mit amtlichen Archivverwaltungen zeigt sich, dass dort die Geodaten mit ihren besonderen Eigenschaften in Umrissen bekannt sind, aber aufgrund der etwas aufwändigeren Archivierungslösung derzeit noch weitgehend ausgeklammert werden. Daher besteht dringender Handlungsbedarf, um für Geodaten den notwendigen langfristigen Schutz zu gewährleisten.

Seitens der ISO, der International Organization for Standardization, wurden in den letzten zehn Jahren eine Reihe von Normen veröffentlicht, die Grundsätze zur Langzeitarchivierung von Geodaten festschreiben. Diese Normen werden weiter unten vorgestellt. Daneben sind eine Reihe von Data Stores für große Geodatenmengen entstanden. Als Beispiel sei hier der „Copernicus Open Access Hub“ für Sentinel-Szenen der ESA genannt, in dem Ende Januar 2020 nach knapp fünf Jahren des Betriebs schon 23.200.000 Datenpakete mit in Summe etwa 20 PBytes frei zum Download zur Verfügung gestellt werden. Die europäischen Sentinel-Satelliten stellen derzeit weniger als zehn von insgesamt etwa 100 Satelliten mit bildgebenden Sensoren im Orbit dar.

---

<sup>1</sup> Hochschule Neubrandenburg, Brodaer Straße 2, D-17033 Neubrandenburg, E-Mail: kresse@hs-nb.de

## 2 Besonderheiten von Geodaten

Wie oben schon erwähnt erreichen die aktuellen Fragestellungen der Archivwelt nicht die drängenden Probleme der Geodaten-Community. Für die Archivare steht die Festlegung von stabilen Dokumentformaten im Vordergrund, allen voran das pdf-Format. Geodaten hingegen besitzen eine Reihe von Besonderheiten, die in der Archivwelt sonst eher selten eine Rolle spielen und daher unbekannt sind:

- Erfordernis einer verlustfreien Datenkomprimierung
- Zusammenfassung vieler Einzeldateien zu Kartenwerken
- Georeferenzierung und parallele Verarbeitung von Vektor- und Rasterdaten
- Beschreibung von Nachbarschaften über die Topologie
- Kartographische Ausgestaltung durch Signaturierung
- Ergänzungen durch Metadaten und Quicklooks

## 3 ISO-Normen

### 3.1 Überblick

Die ISO hat bisher drei Normen für die Langzeitarchivierung von Geodaten veröffentlicht, zwei davon speziell ausgerichtet auf Fernerkundungsdaten. Diese sind die

- ISO 14721:2012 „Space data and information transfer systems — Open archival information system (OAIS) — Reference model“, die
- ISO 19165-1:2018 „Geographic information — Preservation of digital data and metadata — Part 1: Fundamentals“ und die
- ISO/DIS 19165-2:2019-09 „Geographic information — Preservation of digital data and metadata — Part 2: Content specifications for Earth observation data and derived digital products“ (DIS = Draft International Standard).

### 3.2 ISO 14721:2012

Die Entwicklung der ISO 14721:2012 wurde hauptsächlich von Organisationen initiiert, die im Bereich der Fernerkundung tätig sind, unter anderem von der NASA und der ESA. Mit dieser Norm wurde eine Grundstruktur für Langzeitarchivierung festgeschrieben, auf der die weiteren Normen aufbauen können. Vor allem definiert die ISO 14721:2012 eine Kapselung der zu archivierenden Daten in einzureichende Datenpakete (SIP, Submission Information Package), zu archivierende Datenpakete (AIP, Archival Information Package) und wieder zu verteilende Datenpakete (DIP, Dissemination Information Package). Die Abb. 1 zeigt einen Überblick über dieses Konzept.

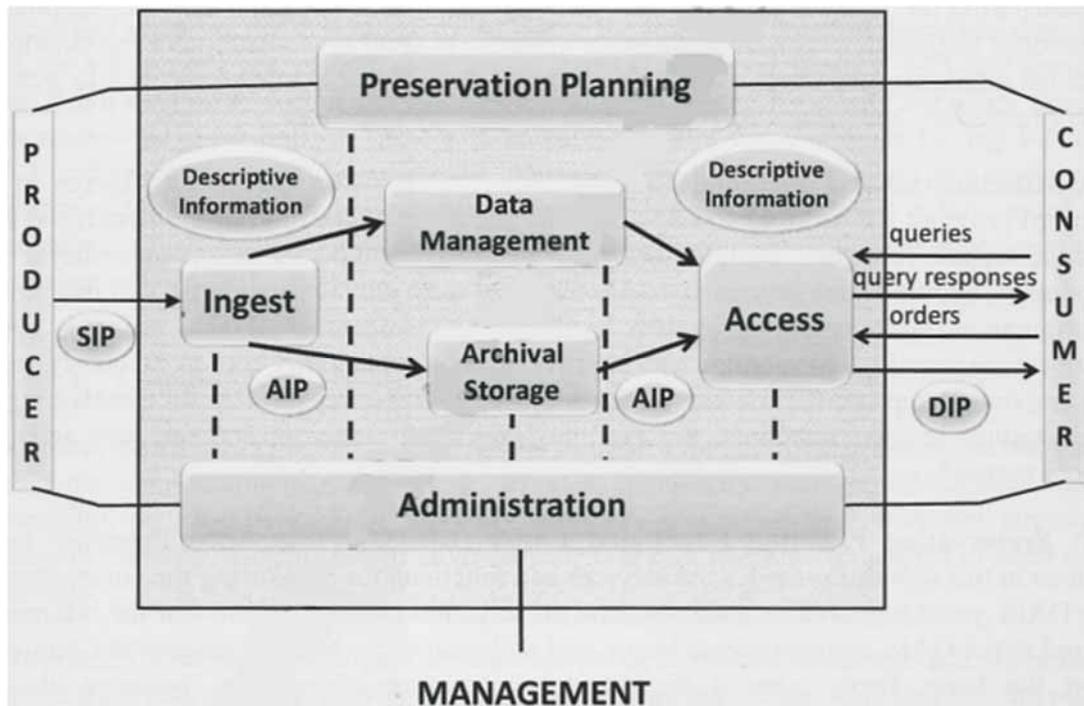


Abb. 1: Struktur und Arbeitsablauf für die Langzeitarchivierung nach ISO 14721:2012 (SIP = Submission Information Package, AIP = Archival Information Package, DIP = Dissimination Information Package)

Bei den Arbeiten zur Norm ISO 19165-1:2018 fiel auf, dass die ISO 14721:2012 ohne Abstimmung mit den Normen des ISO/TC 211 „Geographic information / Geomatics“ entwickelt worden war. Beispielsweise waren und sind viele Definitionen nicht mit denen der Normen der ISO 19100-Serie vereinbar. Bei weiteren Arbeiten fiel auch auf, dass die Autoren der ISO 14721:2012 ihre zweifellos gute, aber eben etwas eigenständige Norm, verbissen verteidigten. Als Folge davon war der Entwicklungsspielraum der ISO 19165-1 stärker eingeschränkt als vorher angenommen.

### 3.3 ISO 19165-1:2018

Für den Autor begann die Arbeit an der Norm Ende 2013. Damals wurde er am Rande einer ISO/TC 211-Plenarsitzung vom Chairman des ISO/TC 211, damals Olaf Østensen, und dem Normungsverantwortlichen der AdV, Markus Seifert, bedrängt, die Projektleitung zu übernehmen. Glücklicherweise stellte sich bald heraus, dass die Aufgabe von mehreren Seiten erkannt und aktiv angegangen wurde, so dass eine ausreichend breite Expertise für die Entwicklung der Norm zur Verfügung stand. Insbesondere ist dem jungen Nachwuchswissenschaftler Joan Masó aus Spanien (Katalonien) zu verdanken, dass durch seine Vorarbeiten für das Open Geospatial Consortium (OGC) und deren Übernahme in die ISO-Norm nach kurzer Zeit ein stabiles Datenmodell zur Verfügung stand. Ein Beispiel für ein Klassendiagramm zeigt die Abb. 2.



Die Klassen der ISO 19165-1:2018 sind als Spezialisierung der Metadatennorm ISO 19115-1:2014 modelliert ist (Abb. 3).

Sehr empfindlich auf die Neuentwicklung vom ISO/TC 211 reagierten die Väter der ISO 14721:2012. Nach einer etwa ein-Jahr-langen Diskussionen beinhaltete die Kompromisslinie eine Weiterentwicklung der SIP, AIP und DIP zu neuen spezialisierten so genannten GeoSIP, GeoAIP und GeoDIP, und den expliziten Hinweis darauf, dass die neue ISO 19165-1 die Festlegungen der ISO 14721:2012 nicht infrage stelle. Diese Abstimmung war der schwierigste Teil der Entwicklung der neuen ISO 19165-1:2018.

Die ISO 19165-1:2018 äußert sich auch zur Frage der physischen Datenspeicherung. Hierbei sollen Daten und Metadaten getrennt abgelegt werden. Maßgeblich ist die ISO/IEC 29500-2:2012 „Open Packaging Convention“. Diese Norm verlangt die zip-Kompression in Verbindung mit xml. Das Ergebnis ist dem tar-Format ähnlich.

### 3.4 ISO/DIS 19165-2

Die ISO/DIS 19165-2 ist, wie der Name mitteilt, noch nicht veröffentlicht, aber fast fertig. „DIS“ bedeutet „Draft International Standard“. Die Projektleitung obliegt Hampapuram Ramapriyan, einem US-Amerikaner mit indischer Herkunft, der die Interessen vor allem der NASA im ISO/TC 211 vertritt.

Diese Norm ist eine Implementierungsnorm der ISO 19165-1 für Attribute, die in der Erdbeobachtung benötigt werden. Erdbeobachtung kann von Satelliten, Flugzeugen oder Vor-Ort-Sensoren (in situ) geleistet werden. Die Norm unterteilt die gelieferten Daten in fünf Kategorien: Konzeptuelle Vorbereitung (Mission Concept Stage), Modellierung (Mission Definition Stage), Implementierung (Mission Implementation Stage), Durchführung (Mission Operations Stage) und Nachbereitung (Post Mission Stage). Die Inhalte lassen sich unter anderem mit den Stichworten Planung, Georeferenzierung, Kalibrierung, Datenqualität, Visualisierung und Datenhistorie beschreiben.

Die Tabellen 1 – 5 zeigen alle in dieser Norm definierten Attribute, die in den meisten Fällen eine Protokollierung der Abläufe darstellen. Um die Übersetzung als Fehlerquelle auszuschließen, werden in der Tabelle die originalen englischen Begriffe aus der Norm verwendet.

Tab. 1: Konzeptuelle Vorbereitung (Mission Concept Stage)

Datenelement	Datentyp	Plattform
Preservation metadata	Documentation	All
Scientific / applications scenario, data producer and user communities	Documentation	All
Mission requirements document	Documentation	Satellite mission
Mission operation plan	Documentation	All
Mission cost and schedule	Documentation	All

Tab. 2: Modellierung (Mission Definition Stage)

<b>Datenelement</b>	<b>Datentyp</b>	<b>Plattform</b>
Preservation metadata	Documentation	All
Mission requirements specifications	Documentation	All
Space or aircraft to ground segment ICDs	Documentation	Primarily satellite missions
Sensor / instrument / platform requirements	Documentation	All
Sensor / instrument processing characteristics	Documentation, data record	All
Sensor / instrument qualification process	Documentation, data record	All
Pre-launch / preoperational calibration and characterization plan	Documentation, data record	Satellite missions
Ground / ocean calibration reference and scientific base	Documentation, data record	Satellite missions
Processing algorithms and data format specification	Documentation	Satellite missions, others, if available
Data product specifications	Documentation	All
Data management plan (DMP)	Documentation	All
Mission cost and schedule	Documentation	All

Tab. 3: Implementierung (Mission Implementation Stage)

<b>Datenelement</b>	<b>Datentyp</b>	<b>Plattform</b>
Preservation metadata	Documentation	All
Mission design	Documentation	All
Detailed space or aircraft to ground segment operations concept and implementation	Documentation	Satellite and aircraft missions
Updated DMP	Documentation	All
On board processing	Documentation	Satellite missions
Sensor / instrument design and Implementation	Documentation	Mostly satellite missions
Calibration and validation	Documentation	Satellite and aircraft missions
Ground processor design, algorithm implementation and supporting information for data processing.	Documentation	Satellite missions
Technical notes scientific papers	Documentation	All
Data format specifications	Documentation	All
Supporting information for processing	Documentation, data records	All
Qualification process	Documentation, data records	All

Tab. 4: Durchführung (Mission Operations Stage)

<b>Datenelement</b>	<b>Datentyp</b>	<b>Plattform</b>
Preservation metadata	Documentation	All
Mission data access and service Requirements document and user handbook	Documentation	Satellite missions
Sensor ground segment operations plan	Documentation	Satellite missions
Mission operations acquisition plans and reports	Documentation, data records	Satellite missions
Raw / level 0	Data records	All
Level 1	Data records	All
Level 2+	Data records	All
Browse images	Data records	All
Ancillary data	Data records	Mostly satellite and aircraft missions
Auxiliary data	Data records	Mostly satellite and aircraft missions
Calibration and validation data	Documentation, data records	All
Quality parameters	Documentation, data records	All
Metadata	Documentation, data records	All
Level 0 consolidation	Software code	Satellite missions
Data processing software	Documentation, software code	Satellite and aircraft missions
Quality control software	Software code	Satellite missions
Science data tools	Software code	Satellite missions
Visualization tools	Software code	Satellite missions
Value-added software	Software code	Satellite missions
Product qualification and quality assurance monitoring reports	Software code	Satellite missions
Sensor / instrument evolution and history records	Documentation	All
Referred publications and papers	Documentation	All
Tandem and / or combined campaigns, comparisons	Documentation	All
Cross-campaign, cross-comparisons and cross-calibration activities documentation and data	Documentation, data records	All
Data access policy	Documentation	All

Tab. 5: Nachbereitung (Post Mission Stage)

Datenelement	Datentyp	Plattform
Preservation metadata	Documentation	All
Data consolidation & reprocessing strategy, implementation plans, and consolidated / reprocessed data: Processing	Documentation	Satellite missions
Data consolidation & reprocessing strategy, implementation plans, and consolidated / reprocessed data: Ancillary, Auxiliary	Documentation, data records	Satellite missions
Data consolidation & reprocessing strategy, implementation plans, and consolidated / reprocessed data: QA	Documentation, data records	All
Data consolidation & reprocessing strategy, implementation plans, and consolidated / reprocessed data: L0, L1, L2+	Data records	All
Data consolidation & reprocessing strategy, implementation plans and consolidated / reprocessed data: Metadata	Data	All
Referred publications and papers	Documentation	All
Historical data access policy	Documentation	Satellite missions
Historical mission user handbook	Documentation	Satellite missions

## 4 Schlussbetrachtung

Die drei Normen zur Langzeitarchivierung zeigen, dass in Archivierungsfragen zu Geodaten die ISO noch nicht zu einer ganz einheitlichen Linie gefunden hat. Während die ISO 19165-1:2018 generisch aufgebaut ist und dem gemeinsamen Datenmodell des ISO/TC 211 folgt (Abb. 3), sind die ISO 14721:2012 und die ISO 19165-2 eher auf die Bedürfnisse von Weltraumorganisationen wie NASA und ESA zugeschnitten. Die ISO 14721:2012 ist dabei gänzlich ohne Bezug zu ISO/TC 211 entstanden. Es bleibt zu hoffen, dass weitere Implementierungen auf Basis der ISO 19165-1:2018 entwickelt werden und dass nach dem Sammeln von Erfahrungen vielleicht eine weiter verbesserte neue Version der ISO 19165-1 entstehen kann.

## 5 Literaturverzeichnis

- ISO 14721:2012, 2012: „Space data and information transfer systems — Open archival information system (OAIS) — Reference model“. iso.org
- ISO 19165-1:2018, 2018: „Geographic information — Preservation of digital data and metadata — Part 1: Fundamentals“. ISO/TC 211, iso.org
- ISO/DIS 19165-2:2019-09, 2019: „Geographic information — Preservation of digital data and metadata — Part 2: Content specifications for Earth observation data and derived digital products“. ISO/TC 211, iso.org

# The Importance of Radiometric Feature Quality for Semantic Mesh Segmentation

DOMINIK LAUPHEIMER<sup>1</sup>, MOHAMAD HAKAM SHAMS EDDIN<sup>1</sup> & NORBERT HAALA<sup>1</sup>

*Abstract: We propose a pipeline for the semantic segmentation of textured meshes in urban scenes as generated from imagery and LiDAR data. Key idea is to represent the mesh as a set of face centroids (COG cloud). This enables the comparison of various point-based classifiers of varying learning abilities. Fine-tuned PointNet++ showed the best results due to hierarchical feature learning. One of the main differences between meshes and point clouds is the availability of high-resolution texture. Hence, we evaluate the importance of radiometric feature quality as a proxy for texture importance. Color information increases performance by at least 5 % (mIoU) for the used data. We achieved to double the performance gain by improving the radiometric feature quality, i.e. utilizing color information of the entire face. Our study shows that texture is beneficial for non-uniform dense and non-balanced data sets. However, it also shows the inherent limitations of textural features like occlusions, absence of imagery, and the quality of the geometric reconstruction.*

## 1 Introduction

The semantic segmentation of 3D data is an everyday issue in the domain of photogrammetry and remote sensing. Common 3D data representations are point clouds, meshes, volumetric representations, and projected views (i.e. RGB-D images or renderings). Amongst them, the semantic segmentation of point clouds may currently be the most popular topic.

The unstructured nature of 3D point sets prevents to apply well-established Deep Learning (DL) methods of the image space directly to point clouds. Accordingly, common approaches structure data into grid-like representations by voxelization (3D) or multi-view rendering (2D) (GRAHAM et al. 2018; BOULCH et al. 2017). Both approaches enable the use of Convolutional Neural Networks (CNNs) but come along with information loss (discretization, occlusions, projection, etc.). Therefore, DL approaches that directly operate on 3D point clouds have been investigated recently (QI et al. 2017a; QI et al. 2017b; BOULCH 2019). The huge number of points and the related matter of subsampling are still big issues.

We claim that meshes are well-suited to tackle these issues due to their lightweight geometric representation. Generally, points on planar surfaces do not provide extra information and hence, do not have to be stored. During the meshing process, such points are eliminated resulting in a smaller memory footprint. For instance, the mesh footprint of the used high-resolution data set (cf. section 2.1) covers ~30 % of the respective LiDAR point cloud (considering XYZ only).

Compared to voxelization or multi-view renderings, structuring a point cloud in the mesh representation, ideally, does not come along with information loss since neither rasterization nor projection is involved. Quite the contrary, for closed surfaces, meshing may increase the

---

<sup>1</sup> Universität Stuttgart, Institut für Photogrammetrie, Geschwister-Scholl-Straße 24D, D-70174 Stuttgart, E-Mail: [Dominik.Laupheimer, Norbert.Haala]@ifp.uni-stuttgart.de, hakam.shams@hotmail.com

information content due to an explicit topology and an explicit surface description (i.e. unambiguous normal vectors).

Moreover, textured meshes carry high-resolution image information stored in so-called texture atlases. In summary, textured meshes provide geometric and textural information in a lightweight fashion and inherently enable data fusion from LiDAR and image data acquisition. According to our observation, meshes currently replace unstructured point clouds as a final user product.

For these reasons, we investigate the semantic segmentation of textured meshes in urban areas as generated from LiDAR data and oblique imagery. Despite their advantages, meshes are a mostly overlooked topic in the domain of photogrammetry and remote sensing. So far, only a few works deal with semantic mesh segmentation in urban scenes (ROUHANI et al. 2017; TUTZAUER et al. 2019). In comparison, in the domain of computer vision, meshes are a default data representation. However, that community deals with small-scale (indoor) data sets (KALOGERAKIS et al. 2017; GEORGE et al. 2017).

We establish a pipeline to segment meshes semantically with three different classifiers: Random Forest (RF), PointNet and PointNet++ (cf. section 2.3). Section 2 describes data preprocessing and feature calculation. The used data is described in detail in (CRAMER et al. 2018; TUTZAUER et al. 2019) and visualized in Fig. 1. Section 3 describes the implementation of the used classifiers and compares their performance. In this study, we focus on the texture importance for the semantic segmentation of meshes utilizing a PointNet++ classifier. We assume that color information in the form of high-resolution texture is very important for semantic segmentation. The color information may have more impact than per-point color information in case of point clouds. The results are discussed in section 4.

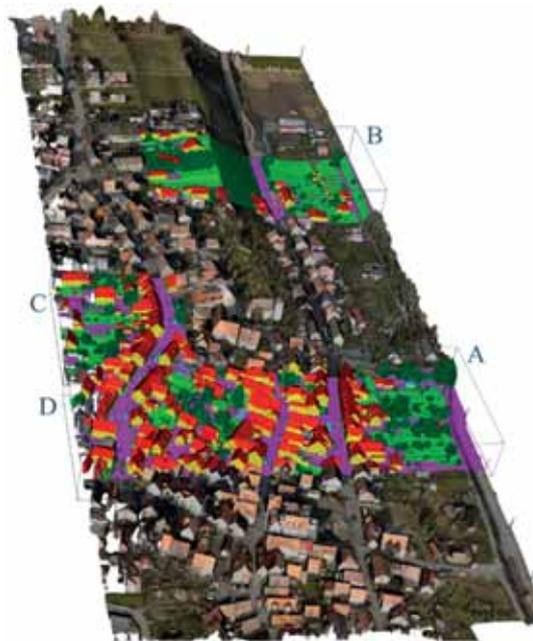


Fig. 1: Used data for training and evaluation ( $800\text{ m} \times 300\text{ m}$ ). Tiles A, B, C, D (depicted in labeled fashion) are used as validation or test set (mutually exclusive). The remaining part (colored by median RGB per face) is used as training set. Classes: *building mass/facade* (yellow), *roof* (red), *impervious surface* (violet), *green space* (light green), *mid and high vegetation* (dark green), *vehicle* (light blue), *chimney/antenna* (orange), and *clutter* (gray).

## 2 Methodology

Compared to point clouds, meshes provide high-resolution texture instead of per-point color only. Hence, we want to investigate the importance of available color information per face. Our approach represents the mesh as a set of face centroids and leverages point-based classifiers. We evaluate radiometric feature importance at various qualities as a proxy for texture importance.

### 2.1 Data Set and Data Preparation

Our study uses the high-resolution data described in (CRAMER et al. 2018). The Airborne Laser Scanning (ALS) data consists of up to 800 points/m<sup>2</sup> for the entire area. The GSD of the oblique images is  $\sim 2.5$  cm. TUTZAUER et al. (2019) obtained a textured 2.5D mesh by fusing the simultaneously acquired ALS data and aerial oblique imagery with software SURE 2 from nFrames (ROTHERMEL et al. 2012). They manually labeled the mesh and split the data into training, validation and test set (cf. Fig. 1). The considered classes are (relative frequency is given in parentheses): *building mass/facade* (9.28 %), *roof* (6.34 %), *impervious surface* (5.67 %), *green space* (5.97 %), *mid and high vegetation* (63.38 %), *vehicle* (0.83 %), *chimney/antenna* (0.31 %), and *clutter* (8.22 %). In regards to the *closed world assumption*, class *clutter* contains all faces that do not match the other classes.

Similar to TUTZAUER et al. (2019), we represent each face by its center of gravity (COG) associated with features (cf. section 2.2). To this end, the mesh is represented by a set of COGs (COG cloud). While still benefiting from mesh-based features like the availability of high-resolution texture, we can apply classifiers that have been designed for point clouds originally.

In our experiments, we focus on PointNet++ since it outperforms the other tested classifiers (cf. section 3.2). In accordance with experiments of (QI et al. 2017a), our experiments showed increased performance with an increased number of input points (i.e. number of sampled COGs in the first level of PointNet++). However, performance gain saturates at a certain input density. Furthermore, the number of input COGs is limited due to GPU memory. For this reason, we partition the data into spatially overlapping tiles with a fixed dimensionality of  $50\text{ m} \times 50\text{ m}$  and an 80 % overlap. Based on these, we generate mini-batches in two different flavors to train PointNet/PointNet++: squared mini-batches or circular mini-batches (QI et al. 2017b; WINIWARTER et al. 2019). Squared mini-batches subsample the generated tiles using random sampling or Farthest Point Sampling (FPS) on the fly (cf. section 3.1). The circular-shaped mini-batches (viewed from above) are generated around the tile centers using kNN, i.e. no subsampling is involved. At the edge of acquired data, the mini-batches can have arbitrary shape since the tile center may already be at the edge.

Whereas memory limits the upper bound of mini-batch sizes, the scene and considered classes limit the lower bounds of mini-batch dimensions. For instance, mini-batches should cover the whole shape of a building, car, etc. and incorporate sufficient spatial context.

For evaluation, we aggregate predictions of overlapping areas for PointNet/PointNet++. We assume the results will be more stable due to redundant predictions (cf. Fig. 6).

Semantic mesh segmentation differs from semantic point cloud segmentation in several aspects. Generally, face density is significantly sparser than point density. Hence, the COG cloud is sparse and has a comparatively small memory footprint. For instance, tile A consists of  $\sim 39.6$  million

LiDAR points or  $\sim 0.3$  million faces/COGs respectively. The COG cloud is  $\sim 133$  times smaller. On the other hand, meshing increases non-uniform density and class imbalance. Non-uniform density is a mesh-inherent issue. Meshes tend to have larger faces in planar areas and many small faces in vivid areas. Thereby, the meshing implicitly shifts class imbalance towards non-planar classes (e.g. vegetation classes). Therefore, handling the non-uniformity and class imbalance is of particular importance. We tackle non-uniform density and class imbalance with FPS, Multi-Scale Grouping (MSG), random dropout on the fly and class weighting (cf. section 3.1). Fig. 2 compares (non-uniform) densities of the dense LiDAR point cloud the respective COG cloud. It also shows the label distributions before (i.e. LiDAR point cloud) and after meshing (i.e. COG cloud) where both data representations are labeled. The LiDAR point cloud was manually labeled (KÖLLE et al. 2019). The labels have been transformed to fit our label scheme. Since the label scheme of the LiDAR point cloud does not consider *chimney/antenna* we manually labeled chimneys and antennas for the comparison. The standard deviation of the class frequencies can be interpreted as a measure for the deviation of equal distribution / class balance (23.6 % for the mesh, 14.6 % for the point cloud).

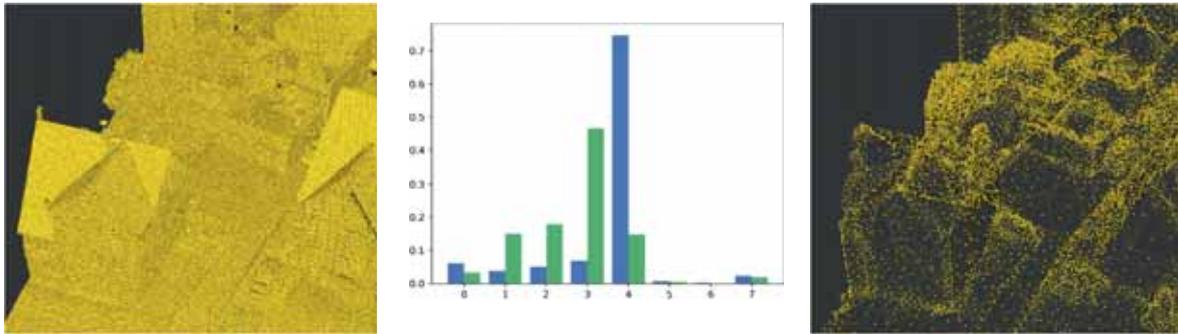


Fig. 2: Non-uniform density of the very dense LiDAR point cloud (*left*) and the mesh represented as COG cloud (*right*). The histogram (*center*) shows the class imbalance of the manually labeled point cloud (*green*) and manually labeled mesh (*blue*) where both representations overlap. Classes: *building mass/facade* (0), *roof* (1), *impervious surface* (2), *green space* (3), *mid and high vegetation* (4), *vehicle* (5), *chimney/antenna* (6), and *clutter* (7).

## 2.2 Feature Calculation

TUTZAUER et al. (2019) attach handcrafted contextual features to the COG cloud. They calculate various geometric and radiometric features based on several scales for each face. Similar to TUTZAUER et al. (2019), we associate per-face features with COGs. However, we do not use contextual features and adapt only the normal vector and the median HSV per face. The normal vector is computed by the cross product of per-face edges. The median HSV per face is extracted from the texture atlas. First, texture coordinates of vertices are transformed into image coordinates. Subsequently, we calculate the median HSV based on the entire face projected onto the texture atlas utilizing previously calculated image coordinates. Our approach can be extended easily by other handcrafted features. However, we limit ourselves to mesh-inherent features in order to reduce computation overhead.

We assume that color information in the form of high-resolution texture is very important for semantic mesh segmentation. The color information may have more impact than per-point color information in case of point clouds. Therefore, our objective is to evaluate the importance of

radiometric feature quality as a proxy for texture importance. Please note, the proxy is needed since the considered classifiers rely on 1D feature vectors. To the best of our knowledge, no classifier is able to combine 1D features with 2D texture information per face. For this reason, we calculate scalar radiometric features with varying granularity. That is why we additionally extract HSV features, which carry less color information. Besides the median HSV per face (*median HSV (face)*), we calculate the mean and median HSV based on the three vertex colors only (*mean HSV (vertices)* and *median HSV (vertices)* respectively). The latter two mimic a colored meshed point cloud, i.e. color information is only available for mesh vertices. We choose HSV space to be independent of illumination variations. Fig. 3 depicts the different qualities of color features transformed into RGB space. The median versions (robust against outliers) provide crisper colors whereas the mean version smears colors the most. *median HSV (face)* uses color information of the entire face and therefore represents faces at border areas (e.g. the transition of the roof to facade) better than *median HSV (vertices)*. In particular, for large faces, this may ensure correct colors (cf. schematic drawing in Fig. 3) and stabilize the semantic segmentation.



Fig. 3: Comparison of radiometric features at different qualities in RGB color space. The faces are colored based on mean/median RGB values. Quality of radiometric features increases from left to right. *Left*: Mean RGB of vertices, *center*: median RGB of vertices, *right*: median RGB of all pixels per face. A schematic drawing is attached to each subfigure in the upper left.

### 2.3 Classifiers Applied

Representing the mesh as COG cloud enables using point-based classifiers. We compare several classifiers with different capabilities of context mapping or hierarchical learning respectively. In this study, we compare the pre-deep-learning era classifier RF with the first neural networks applicable to unordered point sets (PointNet and its extension PointNet++).

The gist of PointNet is to use a symmetric function that is independent of set permutation. The entire COG cloud is encoded in a global feature vector, which is attached to each encoded per-face feature vector. The prediction bases on this concatenated feature vector. A downside of PointNet is the weak local context as it only operates on a global scale. Its extension PointNet++ hierarchically applies PointNets to the iteratively subsampled COG cloud and thus operates on several scales. This procedure enables hierarchical feature learning with increased contextual information similar to CNNs in image space. On the contrary, PointNet and RF cannot operate on

several scales. Moreover, vanilla RF is a pure per-point classifier and does not support feature learning at all. The provided feature set limits its performance. Therefore, RF relies more on expert knowledge/feature design than PointNet/PointNet++. PointNet++ needs the least expert knowledge. Both, PointNet and PointNet++ mainly rely on coordinates as features. Optionally, additional (handcrafted) features can be provided. Best classifier configurations are given in section 3.1.

### 3 Configuration and Comparison of Classifiers

In section 3.1, we describe configurations for each classifier, which we found to perform best for the used data set. In this respect, we focus on PointNet++ and the respective data preparation as we achieve the best results with this classifier (cf. section 3.2). We use recall, precision, and Intersection over Union (IoU) as per-class evaluation metrics for the COG cloud. Furthermore, we report overall accuracy (OA), mean recall (mR), mean precision (mP), and mean IoU (mIoU) at a global scale.

As we work with an imbalanced data set, OA suffers from the accuracy paradox. On the contrary, mIoU can be understood as an average per-class accuracy. The reported values might look low. However, this is due to adapting point cloud metrics to the COG cloud while not considering that each face covers a differently sized area. Thereby, each face has the same impact on the evaluation metrics although face areas vary significantly. As can be seen in Fig. 4 and Fig. 5, the major area is predicted correctly. For tile A, roughly 89 % of the surface area has been predicted correctly. Approximately 83 % of the entire test surface is predicted correctly.

PointNet/PointNet++ are implemented with the DL framework *TensorFlow* (QI et al. 2017a; QI et al. 2017b). RF relies on the *scikit-learn* implementation (python package). For all classifiers, both training and testing are performed on a machine with an NVIDIA GeForce GTX 1080 Ti GPU, 64 GB RAM, and a 12-core CPU.

#### 3.1 Classifier Configurations

Like described in section 2.1, a major part for PointNet/PointNet++ is preparing mini-batches. We run experiments regarding different numbers of input COGs for our two mini-batch modes (squared and circular mini-batches). For squared mini-batches, FPS performed best for a small number of COGs since it provides a better distribution compared to random sampling. Circular mini-batches as generated by kNN do not give a sufficient representation of the neighborhood in that case. However, with an increased number of COGs, random sampling and FPS are almost on par (squared mini-batches) but worse than kNN (circular mini-batches). Random sampling is the fastest; FPS is the slowest method. In our case, circular mini-batches generated with kNN ( $k = 18000$ ) provide reasonable sizes of mini-batches and give the best results at feasible computation times. The circular mini-batches perform better than the squared mini-batches since no subsampling is involved for the mini-batch creation. Therefore, the circular mini-batches provide higher information content.

Regardless of the mini-batch creation, mini-batches are subsampled with FPS in PointNet++ to define the centers of local regions. FPS is used since it is not agnostic to the data distribution. The

local regions are the input to the next level of PointNet++. Limited by GPU memory, 5000 local regions can be processed in the first level of PointNet++.

The local regions (demarcated around FPS-sampled COGs) can be defined in two variants: fixed radius (ball neighborhood) or fixed number of points (kNN). The definition of local neighborhoods with kNN did not perform well. We assume this may be due to non-uniform density limiting the capability of generalization (QI et al. 2017b). In case of fixed radius, MSG proved to be best performing with the following radii: [0.5 m – 1.0 m – 2.0 m] for the first level, [2.0 m – 4.0 m – 8.0 m] for the second level and [4.0 m – 8.0 m – 16.0 m] for the third level. The radii have been chosen heuristically according to the scale of the data set and dimensions of desired objects. MSG improves robustness against density variations but comes along with a great memory footprint and training/inferencing time. These quantities depend mainly on the size of the local regions (defined by number of points or radii), the number of scales, and the number of abstraction levels. To tackle non-uniform density further on, we apply random dropout on the fly. PointNet is less affected by the non-uniform density because of its global abstraction and weak local descriptors.

Both PointNet and PointNet++ use weighted sparse categorical cross-entropy as loss function. We tested several class weighting methods applied to the loss function during training. The best performance is achieved with the inverse square roots of the relative class frequencies in the training data as class weights (WINIWARTER et al. 2019; SCHMOHL & SOERGEL 2019). In order to avoid training on mini-batches that cover mainly a single class, we filter mini-batches whose standard deviation of relative class frequencies exceeds a predefined threshold, similar to (WINIWARTER et al. 2019). We found the best results with a threshold between 30-40%.

We found that PointNet and PointNet++ need distinct treatment for COGs. PointNet++ internally shifts COGs for each local region into a local frame centered at the FPS-sampled COGs. By this means, we achieve local learning based on relative coordinates while keeping the real scale. Counterintuitively, for PointNet, we achieved the best results when we scale the COGs to the unit sphere per tile. Thereby, we follow the original implementation (QI et al. 2017a).

Parameters of the default RF (acts on CPU) have been defined using a grid search with two parameters (*number of trees* and *depth of trees*). We trained an RF without *XYZ* and another with *XYZ* features. We achieved the best results with 30 trees with depth equals 13 (no *XYZ* features) and 46 trees and depth of 7 using *XYZ* features (OA = 70.80 % / mIoU = 23.49 % and OA = 69.54 % / mIoU = 22.46 % respectively). We report both versions since RF without *XYZ* is faster (~25 %) and performs better (~1 %). Nevertheless, for the sake of fair comparison, in section 3.2, we report only the version that uses *XYZ* features. However, we are aware of the fact that the used *XYZ* features for the RF cannot be compared directly with the used *XYZ* features of PointNet and PointNet++ due to the used scaling/normalization in the DL approaches.

### 3.2 Comparison of Classifiers

In section 2.3, we described the key differences of the considered classifiers. Fig. 4 shows ground truth and the predictions of all considered classifiers for tile C. Unavoidably, we suffer from label noise. For instance, in the upper left, a face on the roof is mislabeled. Some faces on the street are associated with wrong labels. Not correctly reconstructed cars are labeled as *clutter*, but are consistently predicted as *vehicle* for PointNet/PointNet++ (cf. oval in Fig. 4). Such faces will

decrease performance metrics. In particular, precision for class *vehicle* and recall for class *clutter* is decreased.

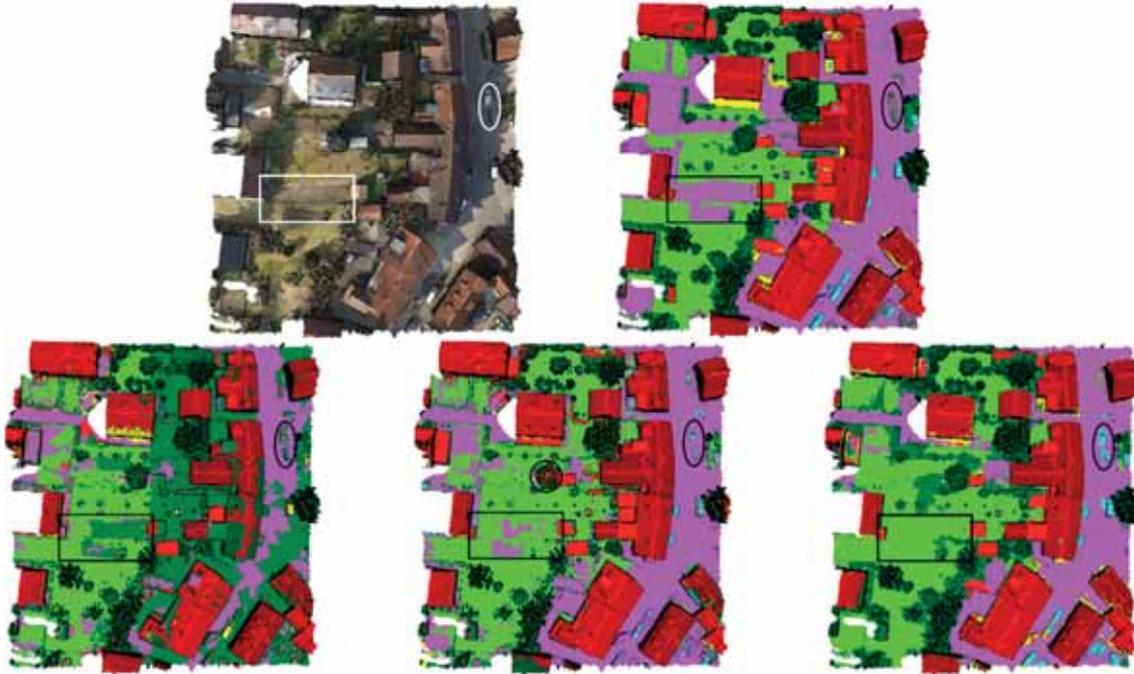


Fig. 4: Ground truth and predictions of several classifiers on tile C. The feature vector consists of  $XYZ$ , *normal vector* and *median HSV (face)*. *Top row*. Tile C colored with median RGB per face (*left*) and ground truth (*right*). The oval shows cars labeled as *clutter* (label noise) that are consistently predicted as *vehicle* for PointNet/PointNet++. *Bottom row (from left to right)*: RF (OA = 69.54 % and mIoU = 22.46 %), PointNet (OA = 70.89 %; mIoU = 29.44 %), PointNet++ (OA = 80.62 %; mIoU = 43.54 %). The circle shows a false prediction (*roof, chimney/antenna*) for a tree due to feature encoding at global scale (PointNet). The rectangle shows that all classifiers have issues separating *green space* from *impervious surface*. Best viewed digitally.

Section 3.1 denotes the best configurations for the used classifiers. For better comparability, the same feature vector configuration has been presented to all classifiers. The feature vector contains radiometric (*median HSV (face)*) and geometric ( $XYZ$ , *normal vector*) features. As expected, RF performs worst (OA = 69.54 %; mIoU = 22.46 %) and PointNet++ performs best (OA = 80.62 %; mIoU = 43.54 %). PointNet (OA = 70.89 %; mIoU = 29.44 %) performs  $\sim 14$  % (mIoU) worse than PointNet++. This is due to the hierarchical feature learning of PointNet++ that causes spatially smoothed predictions at the same time. Furthermore, for PointNet/PointNet++, we aggregate predictions of overlapping tiles (which has a smoothing effect, too). However, even without aggregating, PointNet++ outperforms RF and PointNet. As can be seen in Fig. 4, PointNet++ outperforms the other classifiers in predicting classes *mid and high vegetation*, *vehicle* and *chimney/antenna* (i.e. classes with high and low support). For instance, the recall of *mid/high vegetation* is 88 % and 98 % for PointNet and PointNet++ respectively. A common issue of PointNet is to predict roofs and chimneys within trees due to its global feature encoding (cf. circle in Fig. 4). PointNet++ captures small details and therefore, detects classes with low support better (e.g. *vehicle* and *chimney/antenna*). All classifiers have issues separating *green space* from

*impervious surface* (cf. rectangle in Fig. 4). In particular, PointNet++ consistently labels dirt roads as *green space*. This indicates that even color information is not enough to separate those two classes to the full extent. The reason might be that dirt roads are brownish and texture of *green space* is greenish/brownish since images are captured under leaf-off canopy conditions. Hence, the separation of the respective features might be too difficult, particularly, since the information is smeared in the interpolation layers of PointNet++.

TUTZAUER et al. (2019) feed a multi-branch 1D CNN with many handcrafted contextual features for three different scales. The network is able to learn new features per scale. However, no hierarchical feature learning is possible. They report an OA of 74.76 % for tile C. This is better than PointNet (70.89 %) but worse than PointNet++ (80.62 %). This comparison is not entirely fair since PointNet/PointNet++ uses a smaller feature vector consisting only of *XYZ*, *normal vector* and *median HSV (face)*. Nevertheless, the achieved result emphasizes the superiority of PointNet++ that achieved significantly better results with significantly less handcrafted features. However, when comparing training times we see that multi-branch 1D CNN is much faster than PointNet++ (~15 min vs. 15 h respectively) although using a multiple of features. The reason is that PointNet++ enables hierarchical learning, which includes feature encoding for iteratively increasing local regions, interpolation of encoded features, and MSG. MSG causes the main computational burden. For the 1D CNN, the heavy lifting is already done before training (i.e. calculation of contextual features). PointNet trains ~22 min and RF trains ~2 min.

TUTZAUER et al. (2019) also trained an RF on their feature vector. We achieve ~4% less OA with our RF. The worse performance is expected since we use fewer features and no contextual features at all.

## 4 Investigations on Radiometric Feature Quality

Section 3.2 shows that PointNet++ outperforms the other tested approaches. For this reason, we choose PointNet++ to analyze the importance of radiometric feature quality as a proxy for texture importance. Tab. 1 lists performance metrics for several feature vector configurations for tile A using PointNet++. Fig. 5 shows the respective predicted results for tile A.

Tab. 1 verifies our assumption that texture matters and increases performance on a global scale. Utilizing color information improves performance with respect to OA, mP, mR, and mIoU. Comparing the results of row 1 (geometry only) and row 2 (lowest quality of radiometric features) shows a performance gain at the global scale by approximately 5 % (mP, mR, mIoU). Furthermore, increased textural feature quality improves performance. The model trained on the feature vector with *median HSV (face)* (i.e. best radiometric feature quality) outperforms the model using only geometric features by ~10 % for mIoU and mR; mP is increased by roughly 5 %. By that, the configuration using the best radiometric feature quality (row 4) approximately doubles the achieved performance gain (for mIoU and mR) of the configuration using the lowest radiometric feature quality (row 2) with respect to the purely geometric configuration (row 1). This is quite astonishing since the improvement entirely depends on a substituted scalar radiometric feature. Thereby, the textural content of the entire face is used. However, high-resolution information is not used explicitly since we rely on one-dimensional feature vectors.

Tab. 1: Evaluation metrics for PointNet++ for Tile A using various feature vector configurations (*from top to bottom*: precision, recall, IoU [in %]). The last column lists mP, mR, and mIoU per feature vector configuration (the respective OAs (OAs with respect to surface area) from top to bottom: 84.13 % (85.98 %), 85.06 % (87.69 %), 85.61 % (88.23 %), and 85.68 % (88.56 %)). Geometric features consist of XYZ and *normal vector*. Radiometric feature quality increases from top to bottom.

Feature Vector Configuration	Building Mass/ Facade	Roof	Impervious Surface	Green Space	Mid and High Vegetation	Vehicle	Chimney/ Antenna	Clutter	mP mR mIoU
geometry only	70.31 <b>79.59</b> 59.57	<b>86.26</b> 73.71 65.97	66.64 41.12 34.16	41.03 42.21 26.27	<b>90.49</b> 98.10 88.93	73.49 47.87 40.82	51.67 11.05 10.02	56.57 0.98 0.97	67.06 49.33 40.84
geometry, mean HSV (vertices)	70.44 78.34 58.96	84.70 75.02 66.06	<b>80.08</b> 47.53 42.50	<b>59.00</b> 31.85 26.08	89.02 <b>98.94</b> 88.18	71.60 <b>66.79</b> 52.80	56.96 31.37 25.36	65.43 3.72 3.65	72.15 54.20 45.45
geometry, median HSV (vertices)	73.10 74.27 58.33	81.16 <b>78.30</b> 66.26	76.07 <b>52.76</b> <b>45.25</b>	57.76 <b>45.06</b> <b>33.89</b>	90.39 98.56 <b>89.20</b>	72.81 63.43 51.28	<b>62.64</b> 30.48 25.79	<b>83.85</b> 4.28 4.24	<b>74.72</b> 55.89 46.78
geometry, median HSV (face)	<b>73.67</b> 74.96 59.12	86.00 76.16 <b>67.76</b>	76.41 52.41 45.11	56.52 40.66 30.97	89.86 98.81 88.90	<b>76.33</b> 64.11 <b>53.48</b>	49.56 <b>60.61</b> <b>37.49</b>	69.13 <b>11.16</b> <b>10.63</b>	72.19 <b>59.86</b> <b>49.18</b>

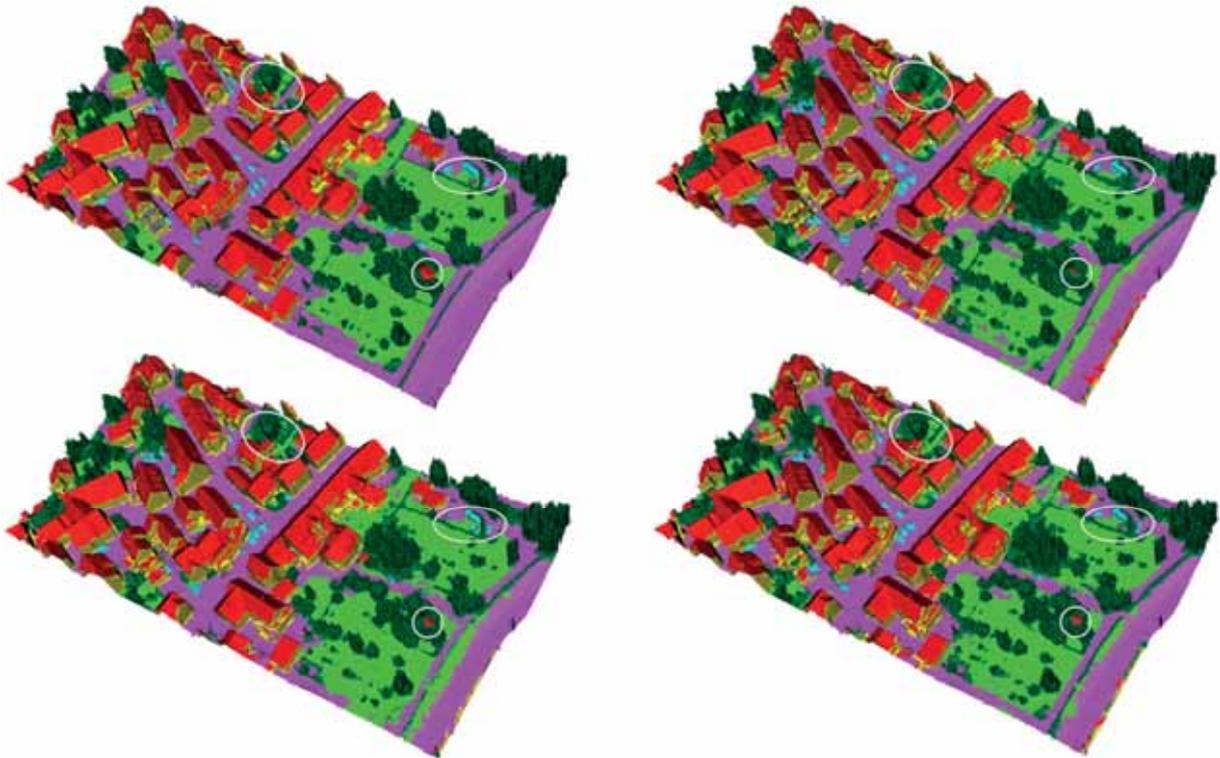


Fig. 5: Predictions of PointNet++ using geometric features and radiometric features of varying quality for tile A (cf. Tab. 1 for detailed numeric values). *From top left, to bottom right*: ground truth, prediction using *mean HSV (vertices)*, prediction using *median HSV (vertices)* and prediction using *median HSV (face)*. The marked areas show best results for *median HSV (face)*. Chimneys and antennas are best detected with the last configuration.

*Median HSV (face)* helps detect classes with low support. Per-class recall and IoU for *chimney/antenna* and *clutter* outperform the configurations using inferior radiometric feature quality to a large extent (up to factor 2 for class *chimney/antenna* or 3 for class *clutter*). This is the reason for a slightly better OA for this configuration. However, OA relies mainly on recall of the dominant class *mid and high vegetation*, which is almost on par for all configurations (~98-99 %). We deduce using per-face texture is beneficial for tackling imbalanced data sets.

Regardless of the radiometric feature quality, separation of classes *green space* and *impervious surface* improves significantly when color information is used. To give an example, 21 % of class *impervious surface* are mislabeled as class *green space* when only geometric features are used. The false positive rate drops by 15 % to 6 % when color information is incorporated. This is in accordance with the findings of (TUTZAUER et al. 2019). Furthermore, color information helps to detect buildings surrounded by high vegetation (cf. circle in Fig. 5). *Median HSV (face)* performs best since the colors are not smeared and better represent borders (cf. Fig 3).

To summarize, the best radiometric feature quality, i.e. *median HSV (face)*, globally achieves the best results. By this, using per-face color outperforms mimicked colored meshed point clouds that leverage per-vertex colors only. This shows the importance of available texture. Generally, the robust median features are better than the mean version. Nonetheless, on the class-level, the configuration using *mean HSV (vertices)* may outperform supposedly better configurations. For instance, it achieves the highest per-class precision for classes *impervious surface* and *green space*. At first glance, this may look counterintuitive. However, we claim that the 2.5D geometry of the mesh and occlusions in imagery are the main reasons why the feature configuration using *median HSV (face)* is not consistently the best (cf. ovals in Fig. 6). Both, “wrong” geometry and occlusions cause not-correctly textured faces. For instance, bushes in front of a building may occlude the facade. In that case, predicting class *mid and high vegetation* becomes more likely with increasing quality of the radiometric feature. The better radiometric feature makes it more likely to vote for a wrong class. To that extent, a better texture measure may dampen performance. In this context, the performance of textural features is naturally limited by occlusions, non-textured faces and the quality of texturing. The latter directly depends on the quality of geometric reconstruction and available imagery.

Regardless of the used feature vector configuration, prediction uncertainty increases at class borders (cf. Fig. 6). This is due to the interpolation of features in PointNet++. However, this is typical behavior in machine learning when vicinity information is incorporated. Furthermore, the effect of different color derivations (like sketched in Fig. 3) can be seen for the face that represents the garage door (cf. circle in Fig. 6).

In general, an increased color feature quality leads to more certain predictions. However, the correct predictions are more or less on par for the visualized fraction. For instance, for the garage door (cf. circle in Fig. 6), predictions using *median HSV (face)* are more certain than predictions based on *median HSV (vertices)*. However, we also see that improved color feature quality introduces uncertainty due to better color representation. For example, the dormers (cf. ovals in Fig. 6) become very uncertain for the configuration using *median HSV (face)*. In fact, a wrong label is predicted. Conversely, the other configurations achieved more stable results in this area. We assume that the better radiometric feature nudges the prediction to class *building mass/facade*.

Nevertheless, the dormers are annotated as *roof*. Thanks to this study, we reveal label noise on a fine-grained level or at least we detected a discussable class definition.

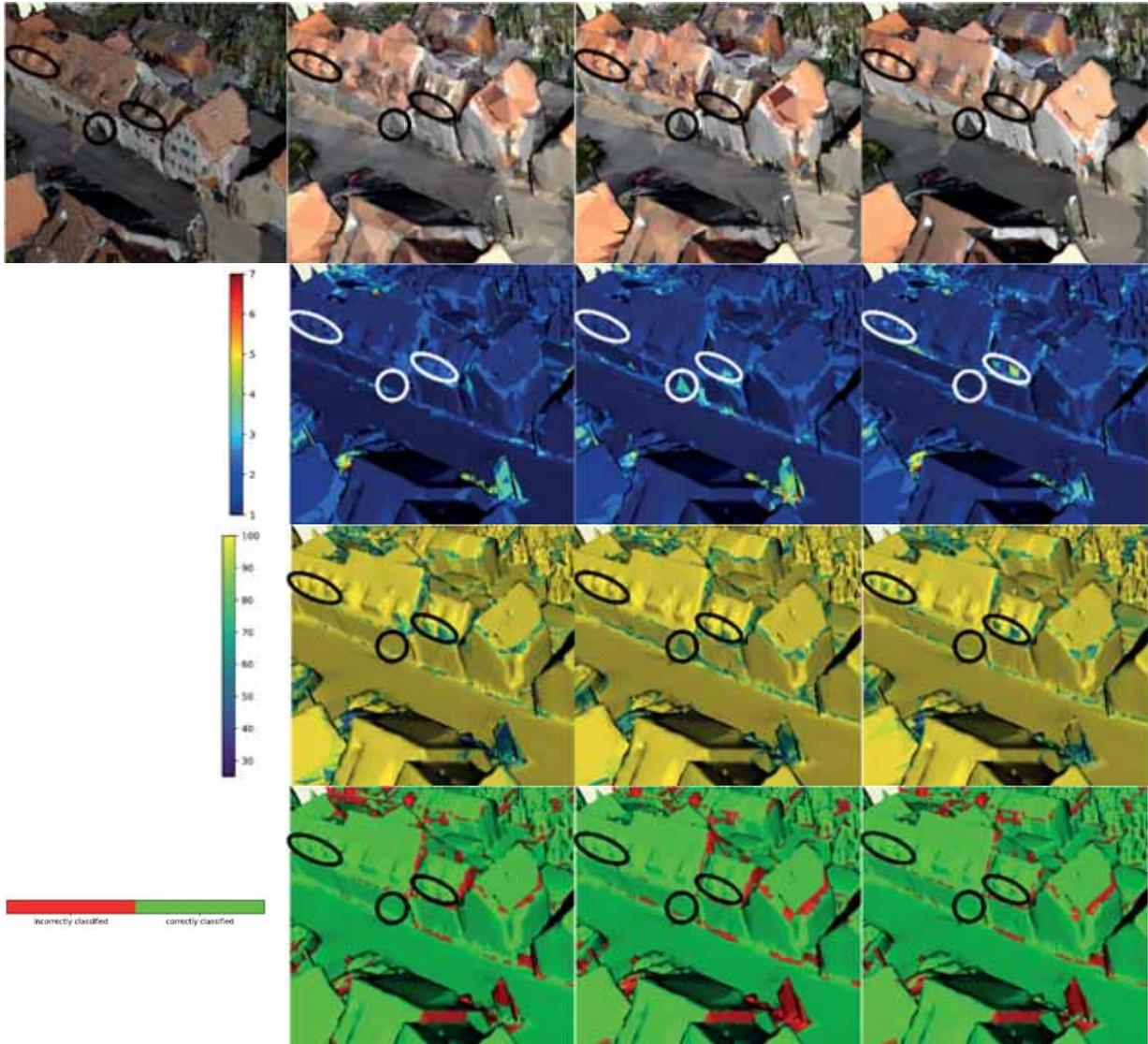


Fig. 6: Visualization of inconsistencies (2<sup>nd</sup> row) and maximum of averaged probabilities (3<sup>rd</sup> row) in aggregated predictions for PointNet++. First column depicts the textured view and the color schemes. The consecutive columns refer to feature vector configurations using *mean HSV (vertices)*, *median HSV (vertices)* and *median HSV (face)* respectively. The inconsistencies are depicted as number of predicted classes per face of overlapping mini-batches. The encircled area (garage door) shows the effect of different color derivations (like sketched in Fig. 3). The ovals show that improved color feature quality introduces uncertainty due to better color representation.

## 5 Conclusion & Outlook

We established a pipeline for semantic segmentation of textured meshes in urban scenes as generated from imagery and LiDAR data. Key idea is to represent each face by its COG associated with features. This enables using point-based classifiers. We compared several classifiers (RF, PointNet, PointNet++ and multibranch-1D CNN) with different capabilities of context mapping. PointNet++ performed best due to its capability of hierarchical feature learning (cf. section 3.2). Approximately 89 % of the surface area of a dedicated test tile has been predicted correctly.

In this study, we investigated in detail the importance/influence of radiometric feature quality. To this end, we derived radiometric features that utilize per-vertex or per-face color information. Regardless of the radiometric feature quality, separation of classes *green space* and *impervious surface* improves significantly. The false positive rate for class *green space* improves by 15 % when color information is used. We showed that increased radiometric feature quality improves performance on the global scale (cf. Tab. 1). Moreover, per-face color information improves detection rates for classes with small support (*vehicle*, *chimney/antenna*, *clutter*). We conclude that texture information matters and increases performance on a global scale. However, the study relies on one-dimensional radiometric features but still shows great performance gain. We claim that utilizing the inherent high-resolution information in the images further increases performance. Therefore, we plan to do semantic image segmentation in image space and back-project the predictions and/or extracted features to the mesh. Thereby, we leverage the entire available image content. Furthermore, we have seen the natural limit of texture due to errors in the geometric reconstruction or occlusions. Some faces carry wrong texture information (e.g. bushes may occlude facades) or do not carry texture at all. For this reason, we plan to make use of LiDAR features. The additional features may help to stabilize predictions and to detect vehicles more properly. Replacing our 2.5D mesh with a 3D mesh will inevitably improve results further on since texture quality depends on the geometric reconstruction. In particular, we plan to generate 3D meshes as generated from ALS data and multi-view stereo image matching.

We are aware of the fact that the investigations at hand are limited to one data set of a rather simple urban scene. We would like to extend investigations to more complex urban data captured with different sensors and flight configurations under different conditions (e.g. different seasons). However, such annotated reference data do not exist so far.

## 6 ACKNOWLEDGEMENTS

The urban mesh is a result of a research project in collaboration with the German Federal Institute of Hydrology (BfG) in Koblenz.

## 7 Bibliography

- BOULCH, A., 2019: ConvPoint: Continuous Convolutions for Cloud Processing. CoRR (arXiv: 1904.02375).
- BOULCH, A., LESAUX, B. & AUDEBERT, N., 2017: Unstructured Point Cloud Semantic Labeling Using Deep Segmentation Networks. Eurographics Workshop on 3D Object Retrieval, The Eurographics Association.
- CRAMER, M., HAALA, N., LAUPHEIMER, D., MANDLBURGER, G. & HAVEL, P., 2018: Ultra-High Precision UAV-Based LiDAR and Dense Image Matching. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **XLII-1**, 115-120.
- GEORGE, D., XIE, X. & TAM, G. K. L., 2017: 3D Mesh Segmentation via Multi-branch 1D Convolutional Neural Networks. CoRR (arXiv: 1705.11050).
- GRAHAM, B., ENGELCKE, M. & VAN DER MAATEN, L., 2018: Semantic Segmentation with Submanifold Sparse Convolutional Networks. Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, 9224-9232.
- KALOGERAKIS, E., AVERKIOU, M., MAJI, S. & CHAUDHURI, S., 2017: 3D Shape Segmentation with Projective Convolutional Networks. CoRR (arXiv: 1612.02808).
- KÖLLE, M., LAUPHEIMER, D. & HAALA, N., 2019: Klassifikation hochaufgelöster LiDAR- und MVS-Punktwolken zu Monitoringzwecken. 39. Wissenschaftlich-Technische Jahrestagung der OVG, DGPF und SGPF, Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V., Band **28**, 692-701.
- QI, C. R., SU, H., MO, K. & GUIBAS, L. J., 2017a: PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. IEEE Conference on Computer Vision and Pattern Recognition, 652-660.
- QI, C. R., YI, L., SU, H. & GUIBAS, L. J., 2017b: PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. Proceedings of the 31<sup>st</sup> International Conference on Neural Information Processing Systems, 5105-5114.
- ROTHERMEL, M., WENZEL, K., FRITSCH, D., HAALA, N., 2012: SURE: Photogrammetric Surface Reconstruction from Imagery. Proceedings LC3D Workshop, Berlin, 8, 2.
- ROUHANI, M., LAFARGE, F. & ALLIEZ, P., 2017: Semantic Segmentation of 3D Textured Meshes for Urban Scene Analysis. ISPRS Journal of Photogrammetry and Remote Sensing **123**, 124-139.
- SCHMOHL, S. & SOERGEL, U., 2019: Submanifold Sparse Convolutional Networks for Semantic Segmentation of Large-Scale ALS Point Clouds. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences **IV-2/W5**, 77-84.
- TUTZAUER, P., LAUPHEIMER, D. & HAALA, N., 2019: Semantic Urban Mesh Enhancement Utilizing a Hybrid Model. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences **IV-2/W7**, 175-182.
- WINIWARTER, L., MANDLBURGER, G., SCHMOHL, S. & PFEIFER, N., 2019: Classification of ALS Point Clouds Using End-to-End Deep Learning. Journal of Photogrammetry, Remote Sensing and Geoinformation Science **87**, 75-90.

# Verwendung künstlicher neuronaler Netzwerke zur Bilderkennung für die Analyse von Straßenmarkierungen in Mobile-Mapping-3D-Punktwolken

JOHANNES WOLF<sup>1</sup>, RICO RICHTER<sup>1</sup> & JÜRGEN DÖLLNER<sup>1</sup>

*Zusammenfassung: Diese Arbeit stellt einen Ansatz zur Verwendung eines etablierten künstlichen neuronalen Netzwerks aus der Bilderkennung für die effiziente semantische Klassifizierung von Straßenmarkierungen aus 3D-Mobile-Mapping-Punktwolkenscans in Straßenumgebungen vor. 3D-Punktwolken werden häufig als Datenquelle für Analysen von Geodaten verwendet. Für ihre Verarbeitung wird in vielen Anwendungsfällen eine semantische Klassifikation benötigt. Neben expliziten Regeln zur Unterscheidung durch geometrische Attribute können maschinelle Lerntechniken eingesetzt werden, um die Markierungen unter Verwendung zuvor trainierter künstlicher neuronaler Netzwerke zu identifizieren. Die eingegebenen 3D-Punktwolken werden um ein semantisches Attribut pro Punkt erweitert und es können Shapes für individuelle Straßenmarkierungen abgeleitet werden, die sich zur Weiterverarbeitung anbieten.*

## 1 Einleitung

Im Bereich der Geodaten finden 3D-Punktwolken breite Verwendung. Sie haben sich als wertvolle Datenquelle für Analysen etabliert, da sie ohne großen Aufwand erfasst und verarbeitet werden können und detaillierte Informationen der aufgenommenen Umgebung beinhalten (VOSSELMAN et al. 2004). Technisch werden sie als ungeordnete Sammlung einzelner Messpunkte vorgehalten, die jeweils Koordinaten in drei Dimensionen besitzen sowie mögliche weitere Attribute, wie beispielsweise Intensitätswerte bei der Erfassung durch LiDAR, haben können (RICHTER et al. 2013). Die ungeordneten und unstrukturierten Punkte einer 3D-Punktwolke benötigen in vielen Anwendungsfällen eine semantische Klassifizierung für die weitere Nutzung (NIEMEYER et al. 2012). Semantische Klassifizierung bezeichnet den Prozess, jedem einzelnen Punkt ein zusätzliches Attribut zuzuordnen, das die Art des Objektes beschreibt, zu dem dieser Punkt gehört, beispielsweise „Auto“, „Laternenpfahl“ oder „Verkehrsschild – rechts abbiegen“. Wenn individuelle Objekte und deren semantische Klassen identifiziert wurden, können diese für verschiedene Anwendungsfälle genutzt werden, wie zum Beispiel zur Erstellung oder Aktualisierung von Straßenkatastern (CAROTI et al. 2005), zur Analyse von Lichtraumprofilen (MIKRUT et al. 2016) oder 3D-Modellierung (VOSSELMAN et al. 2003). Typische semantische Klassen dienen zur grundlegenden Unterscheidung zwischen Boden, Vegetation und Gebäuden. Es können darüber hinaus aber je nach konkreter Anwendung auch viele weitere und sehr detaillierte Klassen, wie zum Beispiel Autos, Straßenmarkierungen, Verkehrsschilder oder Bordsteinkanten benötigt werden (PU et al. 2011). Abb. 1 zeigt eine 3D-Punktwolke, für die

---

<sup>1</sup> Hasso-Plattner-Institut, Digital Engineering Fakultät, Universität Potsdam, Prof.-Dr.-Helmert-Str. 2-3, 14482 Potsdam, E-Mail: [johannes.wolf, rico.richter, juergen.doellner]@hpi.de

semantische Klassen ermittelt wurden und in der die Punkte in jeweils einer Farbe pro Klasse dargestellt werden.

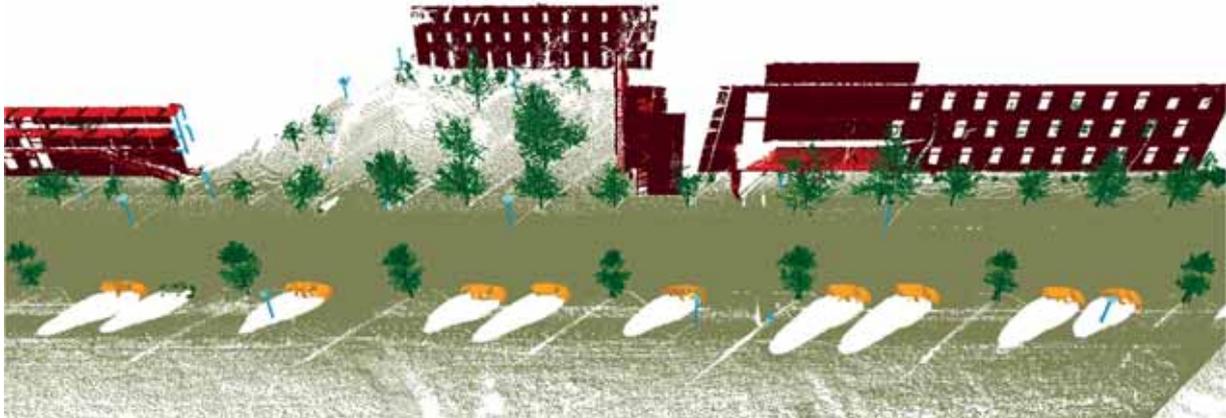


Abb. 1: Nach semantischer Klasse eingefärbte 3D-Punktvolke: Boden (braun), Vegetation (grün), Gebäude (rot), Fahrzeuge (orange), pfahlartige Strukturen (blau)

Diese Arbeit stellt einen Ansatz zur Verwendung eines künstlichen neuronalen Netzwerks aus der Bildanalyse für die semantische Klassifizierung vor. In den letzten Jahren wurden viele Techniken für die automatische Analyse von Bildern und zugehörige Frameworks entwickelt (PULLI et al. 2012). Hier wird gezeigt, dass diese auch zur Klassifizierung von 3D-Punktvolken eingesetzt werden können. Manche Objekte im Straßenraum erstrecken sich nicht in drei Dimensionen oder lassen sich ohne Verlust auch in einem zweidimensionalen Raum darstellen. Straßenmarkierungen befinden sich beispielsweise flach auf dem Boden und ihre Höhe unterscheidet sich, wenn überhaupt, nur marginal von der umgebenden Oberfläche. Sie eignen sich daher besonders für die visuelle Erkennung, weil sie ohne Schwierigkeiten in zweidimensionalen Bildern abgebildet werden können.

Die Erkennung von Straßenmarkierungen in einer Draufsicht aus 3D-Punktvolken bietet gegenüber der Erkennung in Panoramabildern den Vorteil, dass die Markierungen ohne Verzerrung erkannt und in Umringe mit präzisen Positionsangaben als Shape-Dateien überführt werden können.

## 2 Verwandte Arbeiten

3D-Punktvolken sind ein geeigneter Zwischenschritt zur automatischen Erstellung von 3D-Stadtmodellen (SCHWALBE et al. 2005) und Landschaftsmodellen für viele Anwendungsbereiche in der Stadtplanung von Kommunen, für Firmen und Einzelpersonen (VOSSELMAN et al. 2001). Katasterdaten können mit 3D-Punktvolken kombiniert werden, um interaktive Visualisierungen für die Erkundung und Analyse zu erstellen (ARINGER & ROSCHLAUB 2014). Sehr dichte Punktinformationen können analysiert werden und die Ableitung großer Modelle kann ohne großen manuellen Aufwand erfolgen (RICHTER & DÖLLNER 2013). Neben aus der Luft erfassten Daten werden auch Mobile-Mapping-Techniken zur Erfassung eingesetzt (LI 1997). Mobile-

Mapping-Scans können genutzt werden, um beispielsweise automatisch Straßennetze abzuleiten, Straßenoberflächen zu analysieren (JAAKKOLA et al. 2008) oder Gebäudefassaden zu rekonstruieren. Für viele Anwendungen ist die automatisierte Analyse der aufgenommenen 3D-Punktwolken eine zwingende Voraussetzung. Semantische Klassifizierung kann auf unterschiedliche Weise automatisiert durchgeführt werden: Semantische Pro-Punkt-Informationen können aus der Topologie einer 3D-Punktwolke explizit abgeleitet (CHEN et al. 2017) oder durch die Anwendung von Deep-Learning-Konzepten (BOULCH et al. 2017) ermittelt werden.

Im topologischen Ansatz werden explizite Regeln definiert, wie unterschiedliche semantische Klassen durch geometrische Attribute erkannt werden können (GRILLI et al. 2017). 3D-Punktwolken können in lokale Gruppen mit gleichen Eigenschaften, wie zum Beispiel ähnlich ausgerichtete Oberflächennormalen, segmentiert werden (RABBANI et al. 2006). Jedes dieser Segmente kann anschließend in Bezug auf Größe und Orientierung analysiert werden. Große, vertikale Flächen können so beispielsweise als Gebäudefassaden identifiziert werden, während Punktgruppen mit stark variierenden Oberflächennormalen in der Regel Teil von Vegetation sind (WOLF et al. 2019).

Der alternative Ansatz nutzt maschinelles Lernen, um die semantische Klasse von Objekten mit zuvor trainierten künstlichen neuronalen Netzwerken zu ermitteln (ZHOU & TUZEL 2018). Solche Netzwerke nutzen bereits klassifizierte Datensätze für das Training und lernen, die semantische Klasse für individuelle Punkte oder Punktgruppen in neuen, unbekanntem Datensätzen zu bestimmen. In den letzten Jahren wurden mehrere Ansätze entwickelt, die Struktur der 3D-Punktwolken direkt dafür zu nutzen, wie PointNet und ähnliche Netzwerke zeigen (QI et al. 2017). Sie fokussieren sich jedoch zumeist auf Einzelobjekte und benötigen große Trainingsdatensätze. Objekterkennung in Bildern ist seit vielen Jahren ein großes und aktives Forschungsgebiet. Viola et al. stellen einen oft zitierten Algorithmus zur Objekterkennung vor, der beispielsweise zur Gesichtserkennung in Bildern verwendet werden kann (VIOLA et al. 2001). U-Net, das ursprünglich im medizinischen Kontext entwickelt wurde, wird heute in vielen verschiedenen Bereichen zur Bildanalyse eingesetzt (RONNEBERGER et al. 2015). Es ermöglicht die automatische Erkennung von konkreten Bereichen in Bildern, wie beispielsweise Krebszellen aber auch Straßenmarkierungen in Luftbildern (ZHANG et al. 2018).

Anwendung auf 3D-Punktwolken findet U-Net auch bei SnapNet (BOULCH et al. 2018). Dort wird im Gegensatz zur hier vorgestellten Implementierung aus der 3D-Punktwolke ein Mesh generiert, für das zunächst eine Normalenberechnung durchgeführt werden muss. Der hier vorgestellte Ansatz benötigt diese Schritte nicht und fokussiert sich auf die Unterscheidung spezifischer Markierungstypen im Gegensatz zur Erkennung grober semantischer Kategorien.

### 3 Datensätze

In dieser Arbeit werden drei Datensätze aus unterschiedlichen Städten in Deutschland genutzt. Sie variieren im Hinblick auf Punktdichten sowie die Anzahl der Fahrzeuge, Personen und anderen Objekte, die in der Szene vorhanden sind. Das trainierte Netzwerk kann die zu erkennenden Objekte jedoch in allen Datensätzen mit ähnlicher Genauigkeit erkennen. Verschiedene Bereiche der Datensätze wurden für das Training und die Evaluation genutzt.

Eine typische Straße eines Datensatzes ist in Abb. 2 dargestellt. Die dargestellte 3D-Punktwolke besteht aus 29 000 000 Punkten und deckt etwa 670 Meter Straße mit mehreren Kreuzungen ab. Zur Erstellung eines Trainingsdatensatzes für ein U-Net-basiertes künstliches neuronales Netzwerk, das Straßenmarkierungen erkennen kann, wurden mehrere tausend Markierungen in 600 Bildern manuell klassifiziert. Data Augmentation, in diesem Fall Rotation in 90°-Schritten und Spiegelung, wurde verwendet, um den Datensatz auf das Achtfache zu vergrößern.



Abb. 2: 3D-Punktwolke einer Straße, die als Eingabe für die Erkennung von Straßenmarkierungen genutzt wurde

## 4 Konzept und Implementierung

Der hier vorgestellte Ansatz nutzt die Möglichkeiten von Algorithmen zur bildbasierten Objekterkennung, um automatisiert Straßenmarkierungen in 3D-Punktwolken zu klassifizieren. Straßenmarkierungen sind einfach abzubilden, da sie in einer Draufsicht auf die 3D-Punktwolke deutlich sichtbar sind. Die durch den Ansatz erkannten Markierungen sind in Abb. 3 dargestellt.



Abb. 3: 3D-Punktwolke, in der identifizierte Straßenmarkierungen als orangefarbene Shapes eingeblendet sind

Die Implementierung basiert auf einem Pipeline-Konzept, das große Datensätze automatisiert rendern kann, Straßenmarkierungen in den gerenderten Bildern erkennt und die Ergebnisse in die originale 3D-Punktwolke zurückprojiziert.

Zunächst werden die Eingabe-3D-Punktwolken wie in Abschnitt 4.1 beschrieben gefiltert. Ein Renderer erstellt anschließend quadratische Bilder, darauf wird in Abschnitt 4.2 näher eingegangen.

Die gerenderten Bilder werden mit dem zuvor trainierten künstlichen neuronalen Netzwerk klassifiziert und die Ergebnisse werden in die 3D-Punktwolke zurückprojiziert, wie in Abschnitt 4.3 beschrieben. Abschnitt 4.4 beschreibt die Ableitung von Shapes für individuelle Markierungen.

#### **4.1 Vorverarbeitung der 3D-Punktwolken**

Die Implementierung soll skalieren, so dass auch mehrere Milliarden Punkte, wie sie typischerweise bei der Erfassung ganzer Städte entstehen, effizient analysiert werden können. Eine geeignete Datenreduktion ist daher ein wesentlicher Aspekt der Verarbeitung. Um Straßenmarkierungen zu erkennen, muss nur die Straße selbst entlang der befahrenen Bereiche betrachtet werden. Während der Aufnahme der 3D-Punktwolken wird eine Trajektorie aufgezeichnet, die den exakten Pfad des Messfahrzeugs beschreibt. Die 3D-Punktwolke kann mithilfe dieser auf einen etwa 20 Meter breiten Streifen zurechtgeschnitten werden. Wenn keine Trajektorie vorliegt, kann der relevante Bereich durch eine Analyse der lokalen Punktdichte ermittelt werden, weil Bereiche in räumlicher Nähe zum scannenden Fahrzeug eine höhere Auflösung besitzen als weiter entfernte Bereiche. In den verbleibenden Daten werden Ausreißer durch eine Ausreißer-Erkennung entfernt, um Störungen in der 3D-Punktwolke zu entfernen, die das Rendern einer Draufsicht behindern könnten. Alle Punkte mit beispielsweise weniger als fünf benachbarten Punkten in einer Entfernung von einem halben Meter können als Ausreißer markiert und entfernt werden. Dieser Verarbeitungsschritt kann durch eine heuristische Suche beschleunigt werden, die eine räumliche Datenstruktur, wie zum Beispiel ein dreidimensionales Raster, nutzt, in das alle Punkte platziert werden. Punkte innerhalb einer Rasterzelle, die weniger als eine bestimmte Anzahl von Punkten beinhaltet, können als Ausreißer markiert werden. Für den speziellen Anwendungsfall dieser Arbeit ist ein solcher heuristischer Ansatz ausreichend, weil die wesentlichen Informationen alle in dichten Bereichen der 3D-Punktwolke liegen.

Ein Bodenerkennungs-Schritt identifiziert Bodenpunkte basierend auf deren relativer Höhe und Orientierung (MENG et al. 2009). Darüber befindliche Punkte können entfernt werden, sodass Punkte aus Gebäuden oder Vegetation nicht analysiert werden. Der Algorithmus unterteilt die zu untersuchende Fläche in ein reguläres zweidimensionales Raster. Für jede Rasterzelle wird der niedrigste Höhenwert der darin befindlichen Punkte ermittelt. Diese Werte repräsentieren ein vereinfachtes Bodenmodell. Nachdem das Raster so initialisiert wurde, werden Scanlines genutzt, um alle Bodenpunkte der 3D-Punktwolke zu finden. Diese Scanlines bewegen sich entlang der Achsen in positive und negative Richtung sowie diagonal durch das Raster. Der Algorithmus berücksichtigt, welcher Anstieg in welche Scan-Richtung ermittelt wurde und wie sich die Höhe zwischen Punkten und dem niedrigsten Wert der Umgebung verhält. Für jede Scanline werden mögliche Bodenpunkte separat ermittelt. Im Anschluss wird durch eine Mehrheitsentscheidung festgelegt, welche Punkte als Bodenpunkte betrachtet werden.

Die verbleibende 3D-Punktwolke beinhaltet nur noch Bodenpunkte entlang der Trajektorie des Messfahrzeugs ohne Ausreißer. Nach dem Vorverarbeitungsschritt enthielten die 3D-Punktwolken der Testdatensätze durchschnittlich noch etwa 60% der ursprünglichen Punkte.

## 4.2 Rendering der Bilder

Für den zu betrachtenden Bereich können nun Bilder in einer Draufsicht gerendert werden, die nacheinander durch das künstliche neuronale Netzwerk verarbeitet werden sollen.

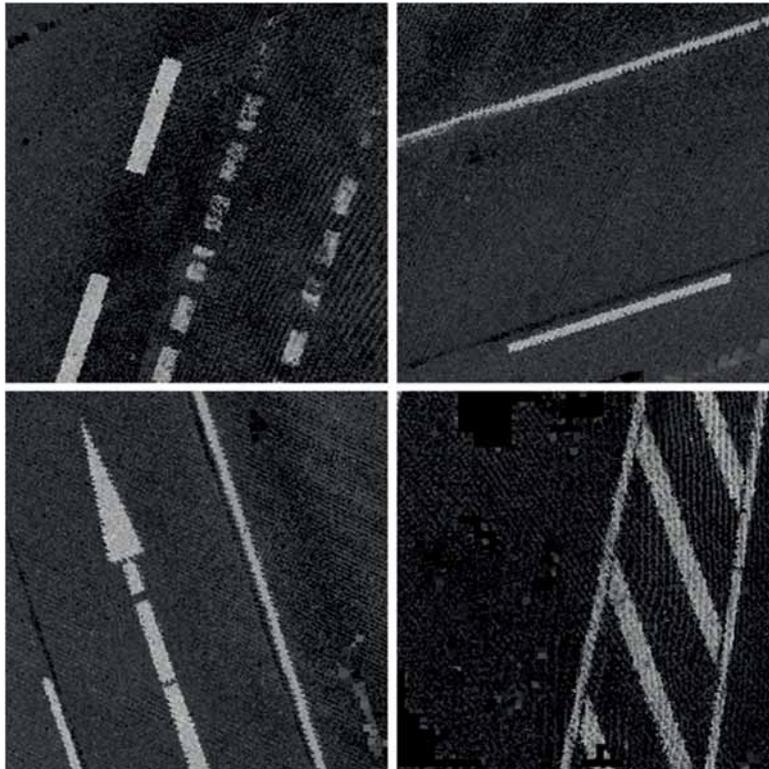


Abb. 1: Beispiele für gerenderte Bilder aus einer 3D-Punktwolke, die Fahrbahnmarkierungen zeigen

Der genutzte Renderer verwendet eine 3D-Punktwolke als Eingabe und generiert eine Menge von Bildern der Auflösung  $128 \times 128$  Pixel in orthogonaler Projektion, wie in Abb. 4 dargestellt wird. Der dargestellte Bereich hat eine Größe von etwa  $4,5 \times 4,5$  Metern. Die Position der Bilder wird so festgelegt, dass der komplette Bereich der zuvor gefilterten 3D-Punktwolke abgedeckt wird und sich die Bilder jeweils ein kleines Stück überlappen. Jedes Bild beinhaltet einen Kanal mit den Intensitätswerten des in jedem einzelnen Pixel jeweils sichtbaren Punktes sowie einen Kanal mit der ID des Punktes, der an dieser Position gerendert wurde. Zweiterer wird für die Rückprojektion der Klassifizierungsergebnisse in die 3D-Punktwolke benötigt.

Für bessere Ergebnisse werden die Punkte der 3D-Punktwolke als Paraboloiden gerendert, von denen in dichten Bereichen jeweils nur die Spitzen sichtbar sind. Rendering-Ergebnisse mit verschiedenen Primitiven werden in Abb. 5 dargestellt. Durch die Verwendung von Paraboloiden werden in weniger dichten Bereichen mehr Pixel ausgefüllt, um Löcher im generierten Bild zu vermeiden, während in dichten Bereich gleichzeitig scharfe Kanten einzelner Markierungen erhalten bleiben (siehe Abb. 5e).

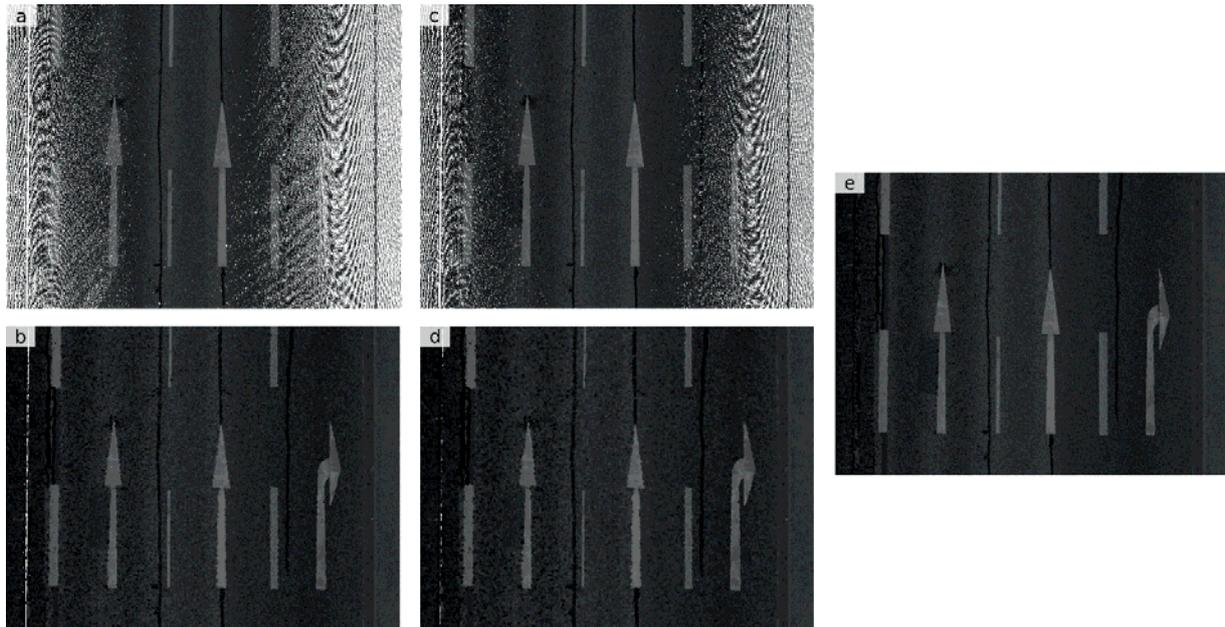


Abb. 2: 3D-Punktwolke gerendert mit unterschiedlichen Primitiven: a) Kleine Punkte (2 Pixel). b) Große Punkte (6 Pixel). c) Kleine Quads (1×1 cm). d) Große Quads (3×3 cm). e) Paraboloid

### 4.3 Klassifizierung

Die gerenderten Bilder werden als Eingabe für das zuvor trainierte künstliche neuronale Netzwerk verwendet. Das Ergebnis ist eine Ausgabemaske für jedes Eingabebild, in dem jedem Pixel die Art des dort erkannten Markierungstyps, bzw. die Information „keine Straßenmarkierung“ zugeordnet ist. Nachdem U-Net ermittelt hat, welche Pixel zu welcher Art Straßenmarkierung gehören, können die zugehörigen Punkte in der 3D-Punktwolke über den Punkt-ID-Kanal ermittelt werden. Diesen Punkten wird in einem neuen Attribut ihre semantische Klasse zugeordnet.

Die Zahl der Punkte in dem von einem Bild abgedeckten Ausschnitt ist zumeist höher als die Anzahl der Pixel des Bildes. Mehrere Punkte der 3D-Punktwolke sind demzufolge verdeckt und ihre IDs können nicht aus den klassifizierten Bildern ermittelt werden. Daher wird in einem Nachbearbeitungsschritt allen Punkten in der direkten Nachbarschaft von identifizierten Punkten ebenfalls das entsprechende semantische Attribut zugewiesen.

### 4.4 Shape-Erstellung

Für jede Gruppe zusammenhängender Punkte der gleichen semantischen Klasse kann ein Shape generiert werden (ESRI 1998). Je nach Markierungstyp werden eine konvexe Hülle oder das am besten passende Rechteck um die Punkte herum aufgespannt, wodurch Shapes für die Straßenmarkierungen erzeugt werden, wie sie in Abb. 3 dargestellt sind. Die resultierenden Dateien können anschließend in beliebigen GIS-Anwendungen für die weitere Verarbeitung genutzt werden.

## 5 Evaluierung

Der Evaluierungsdatensatz umfasst mehrere 3D-Mobile-Mapping-Punktwolken, die mit der gleichen Erfassungshardware in den gleichen Städten aufgenommen wurden, aus denen auch die Trainingsdaten stammen. Das künstliche neuronale Netzwerk wurde fünf Stunden auf einer Nvidia GeForce 1080 Ti trainiert. Nach manueller Prüfung wurden bei einem Durchsatz von 7,5 Millionen Punkten pro Minute 91% der Straßenmarkierungen korrekt identifiziert. Das entspricht einer Verarbeitung von etwa 300 Metern erfasster Straße pro Minute. Abb. 6 zeigt einen Ausschnitt einer automatisch generierten Shape-Datei mit erkannten Straßenmarkierungen.

Hauptsächlich wird die Geschwindigkeit der Analyse durch Lese- und Schreiboperationen beschränkt, wenn die einzelnen Bilder gerendert und auf die Festplatte geschrieben werden. Eine In-Memory-Lösung könnte den Ansatz um ein Vielfaches beschleunigen.

Fehler in der Erkennung lassen sich zum größten Teil in nicht erkannte Straßenmarkierungen in Bereichen allgemein hoher Intensitätswerte in nassen Bereichen der Fahrbahn und in falsch zugeordnete semantische Klassen, wie zum Beispiel den falschen Linientyp, zusammenfassen.

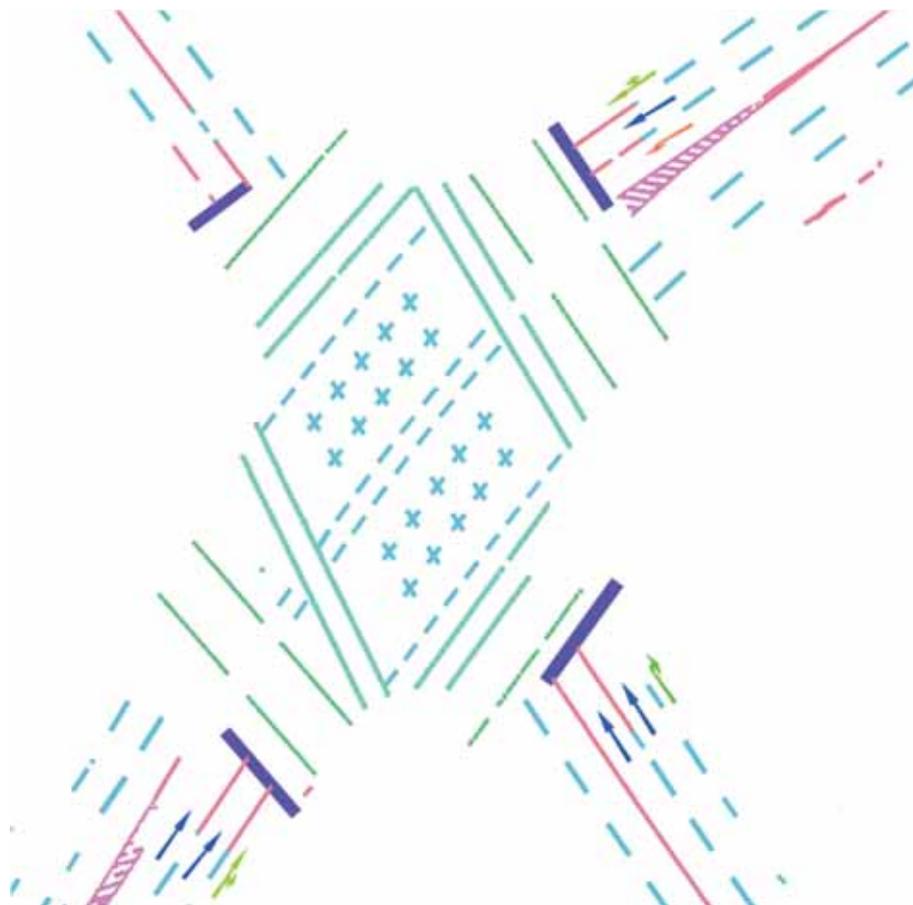


Abb. 3: Ausschnitt einer Shape-Datei mit Shapes pro erkannter Straßenmarkierung. Unterschiedliche Farben geben die jeweilige semantische Klasse an, die für die Markierung erkannt wurde.

## 6 Fazit und Ausblick

Die in dieser Arbeit vorgestellte Implementierung zeigt, dass ein künstliches neuronales Netzwerk für Bildanalyse zur Klassifizierung von Straßenmarkierungen in 3D-Mobile-Mapping-Punktwolken eingesetzt werden kann. Durch die Verwendung einer geeigneten Rendering-Technik werden detaillierte Bilder des erfassten Bodens generiert, die als Eingabe für das künstliche neuronale Netzwerk verwendet werden können. Es ist möglich, die identifizierten Markierungen in die 3D-Punktwolke zurück zu projizieren und Shape-Dateien abzuleiten, die in GIS-Anwendungen verwendet werden können.

Die Implementierung kann neben Straßenmarkierungen auch auf Kanaldeckel und ähnliche Objekte auf der Straße verwendet werden, dafür wird lediglich ein entsprechender Datensatz für das Training des künstlichen neuronalen Netzwerkes benötigt.

Es ist davon auszugehen, dass ein ähnlicher Ansatz auch für dreidimensionale Objekte wie Autos oder Verkehrsschilder eingesetzt werden kann, wenn ein paar Anpassungen vorgenommen werden. Dies soll in zukünftigen Arbeiten evaluiert werden. Mehrere Bilder können in horizontaler Projektion aus verschiedenen Winkeln um zuvor ermittelte Segmente herum gerendert werden, (zum Beispiel vier Bilder von jeder Seite). Ein zusätzlicher Tiefenkanal im gerenderten Bild enthält einen Teil der dreidimensionalen Information und ermöglicht eine einfache Unterscheidung zwischen Vorder- und Hintergrund. Ein solcher Ansatz könnte dann viele weitere Objekte im Straßenraum klassifizieren, die bisher nicht abgedeckt sind.

## 7 Danksagung

Wir danken dem Amt für Geoinformation, Vermessung und Kataster der Stadt Essen für die Bereitstellung von Daten, die im Rahmen der hier vorgestellten Arbeit verwendet wurden.

Außerdem bedanken wir uns bei Tobias Pietz und dem Bachelorprojekt 2018/2019 für ihren Beitrag zu diesem Thema.

## 8 Literaturverzeichnis

- ARINGER, K. & ROSCHLAUB, R., 2014: Bavarian 3D Building Model and Update Concept Based on Lidar, Image Matching and Cadastre Information. *Innovations in 3D Geo-Information Sciences*, Springer, 143-157.
- BOULCH, A., GUERRY, J., LE SAUX, B., & AUDEBERT, N., 2018: SnapNet: 3D Point Cloud Semantic Labeling with 2D Deep Segmentation Networks. *Computers & Graphics* **71**, 189-198.
- BOULCH, A., LE SAUX, B. & AUDEBERT, N., 2017: Unstructured Point Cloud Semantic Labeling Using Deep Segmentation Networks. *Proceedings of 3DOR* **2**.
- CAROTI, G., PIEMONTE, A. & PUCCI, B., 2005: Terrestrial Laser Scanning as Road's Cadastre Revision and Integration Support. *ISPRS Workshop Italy-Canada 2005, 3D Digital Imaging and Modeling: Applications of Heritage, Industry* **1**, 1-3.

- CHEN, D., WANG, R. & PEETHAMBARAN, J., 2017: Topologically Aware Building Rooftop Reconstruction from Airborne Laser Scanning Point Clouds. *IEEE TGRS* **55**(12), 7032-7052.
- ESRI, 1998: ESRI Shapefile Technical Description. <https://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>, zuletzt abgerufen: 2019-11-25.
- GRILLI, E., MENNA, F. & REMONDINO, F., 2017: A Review of Point Clouds Segmentation and Classification Algorithms. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* **42**, 339-344.
- JAAKKOLA, A., HYYPPÄ, J., HYYPPÄ, H. & KUKKO, A., 2008: Retrieval Algorithms for Road Surface Modelling Using Laser-Based Mobile Mapping. *Sensors* **8**, 5238-5249.
- LI, R., 1997: Mobile Mapping: An Emerging Technology for Spatial Data Acquisition. *Photogrammetric Engineering and Remote Sensing* **63**(9), 1085-1092.
- MENG, X., WANG, L., SILVÁN-CÁRDENAS, J.L. & CURRIT, N., 2009: A Multi-Directional Ground Filtering Algorithm for Airborne LIDAR. *ISPRS Journal of Photogrammetry and Remote Sensing* **64**(1), 117-124.
- MIKRUT, S., KOHUT, P., PYKA, K., TOKARCZYK, R., BARSZCZ, T. & UHL, T., 2016: Mobile Laser Scanning Systems for Measuring the Clearance Gauge of Railways: State of Play, Testing and Outlook. *Sensors*, **16**(5), 683-701.
- NIEMEYER, J., ROTTENSTEINER, F. & SOERGEL, U., 2012: Conditional Random Fields for LiDAR Point Cloud Classification in Complex Urban Areas. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **1**(3), 263-268.
- PU, S., RUTZINGER, M., VOSSELMAN, G. & ELBERINK, S.O., 2011: Recognizing Basic Structures from Mobile Laser Scanning Data for Road Inventory Studies. *ISPRS Journal of Photogrammetry and Remote Sensing* **66**(6), 28-39.
- PULLI, K., BAKSHEEV, A., KORNYAKOV, K. & ERUHIMOV, V., 2012: Real-Time Computer Vision with OpenCV. *Communications of the ACM* **55**(6), 61-69.
- QI, C. R., SU, H., MO, K. & GUIBAS, L.J., 2017: Pointnet: Deep Learning on Point Sets for 3D Classification and Segmentation. *IEEE Conference on Computer Vision and Pattern Recognition*, 652-660.
- RABBANI, T., VAN DEN HEUVEL, F. & VOSSELMANN, G., 2006: Segmentation of Point Clouds Using Smoothness Constraint. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* **36**(5), 248-253.
- RICHTER, R., BEHRENS, M. & DÖLLNER, J., 2013: Object Class Segmentation of Massive 3D Point Clouds of Urban Areas Using Point Cloud Topology. *International Journal of Remote Sensing* **34**(23), 8408-8424.
- RICHTER, R. & DÖLLNER, J., 2013: Concepts and Techniques for Integration, Analysis and Visualization of Massive 3D Point Clouds. *Computers, Environment and Urban Systems* **45**, 114-124.

- RONNEBERGER, O., FISCHER, P. & BROX, T., 2015: U-Net: Convolutional Networks for Biomedical Image Segmentation. International Conference on Medical Image Computing and Computerassisted Intervention, Springer, 234-241.
- SCHWALBE, E., MAAS, H.-G. & SEIDEL, F., 2005: 3D Building Model Generation from Airborne Laser Scanner Data Using 2D GIS Data and Orthogonal Point Cloud Projections. The International Archives of Photogrammetry and Remote Sensing **3**, 12-14.
- VIOLA, P., JONES, M., 2001: Rapid Object Detection Using a Boosted Cascade of Simple Features. CVPR (1), 511-518.
- VOSELMAN, G., 2003: 3D Reconstruction of Roads and Trees for City Modelling. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences **34**, 3-8.
- VOSELMAN, G. & DIJKMAN, E., 2001: 3D Building Model Reconstruction from Point Clouds and Ground Plans. International Archives of Photogrammetry and Remote Sensing **34**(3/W4), 37-43.
- VOSELMAN, G., GORTE, B. G., SITHOLE, G. & RABBANI, T., 2004: Recognising Structure in Laser Scanner Point Clouds. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences **46**(8), 33-38.
- WOLF, J., RICHTER, R. & DÖLLNER, J., 2019: Techniques for Automated Classification and Segregation of Mobile Mapping 3D Point Clouds. 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 201-208.
- ZHANG, Z., LIU, Q. & WANG, Y., 2018: Road Extraction by Deep Residual U-Net. IEEE Geoscience and Remote Sensing Letters **15**(5), 749-753.
- ZHOU, Y. & TUZEL, O., 2018: Voxelnet: End-to-end Learning for Point Cloud Based 3D Object Detection. IEEE Conference on Computer Vision and Pattern Recognition, 4490-4499.

# Ermittlung urbaner Überwärmungsbereiche auf der Basis multitemporaler Satellitendaten am Beispiel der Hansestadt Rostock

GÖRRES GRENZDÖRFFER<sup>1</sup>

*Zusammenfassung: Datengrundlage für die Ermittlung urbaner Überwärmungszonen ist eine langfristige Auswertung der Oberflächentemperaturen der Stadt. Dazu sind über einen Zeitraum von 20 Jahren 47 Satellitenszenen (Landsat 5,7,8) der Sommermonate (Ende März – Anfang September) ausgewählt worden. Der langfristige Datensatz gewährleistet, dass die daraus abgeleiteten Überwärmungszonen auch stabil und aussagekräftig sind. Der Vergleich verschiedener jeweils fünfjähriger Perioden zeigt die zwischenzeitliche Veränderung der Oberflächentemperaturen im Stadtgebiet von Rostock. Da die Bodenbedeckung, bzw. Landnutzung einen Großteil der Überwärmung ausmacht, wird zusätzlich auf den Zusammenhang zwischen der Landnutzung und der Überwärmung eingegangen. Als Werkzeug dient dazu der Vergleich zwischen dem NDVI, d.h. einem Vegetationsindex, der auch sehr gut zwischen versiegelter und nicht versiegelter Fläche unterscheiden kann.*

## 1 Einleitung – urbane Überwärmungsbereiche

Bei der laufenden Diskussion um Klimaerwärmung, die etwaige Zunahme von Hitzestreß auch im Hinblick auf außergewöhnlich warme Sommer, rücken auch die so genannten Überwärmungsbereiche einer Stadt in den Fokus der öffentlichen Diskussion. Diese haben aber nur wenig mit der konkreten Hitzebelastung auf die Stadtbewohner zu tun. Diese konkrete bzw. gefühlte Hitzebelastung hängt nämlich von vielen individuellen Faktoren ab, beispielsweise von der Lage bzw. Exposition der jeweiligen Wohnung zur Sonne (Süd- vs. Nordseite), der baulichen Umgebung, der Dämmung, der Klimatisierung, dem gesundheitlichen Allgemeinzustand der betroffenen Person etc. Diese Differenzierung kann und will dieser Beitrag nicht leisten.

Vielmehr untersucht der Artikel auf einer größeren Maßstabsebene den Wärmeinseleffekt anhand eines Beispiels, der Hansestadt Rostock. Damit sollen Bereiche und deren Veränderungen über einen Zeitraum von 20 Jahren identifiziert werden, die sich bei entsprechenden Wetterlagen, d.h. sonnige Sommertage, besonders erwärmen. Auch soll an dieser Stelle auf die durchaus positiven Wirkungen städtischer Überwärmung hingewiesen werden. Dazu zählen z.B. das Vorhandensein vieler wärmeliebender Pflanzen in den innerstädtischen Gärten und Grünanlagen in den gemäßigten Klimazonen und eine allgemein verlängerte Vegetationsperiode, die als positive Auswirkungen der Wärmeinsel gelten. Noch dazu sind die winterliche Frostperiode, die Frostintensität und die Schneedeckendauer durch die höheren Temperaturen in Agglomerationen erheblich verkürzt, was sich wiederum auf die Vegetationsentwicklung vorteilhaft auswirkt. Nicht zuletzt reduzieren sich dadurch auch die Heizkosten der (Innen-) Stadtbewohner. Eine bioklimatisch extrem negative

---

<sup>1</sup> Universität Rostock, Professur für Geodäsie und Geoinformatik, J.-v.-Liebig Weg 6, D-18059 Rostock, E-Mail: goerres.grenzdoerffer@uni-rostock.de

Wirkung hat die Überwärmung, wenn sich in den Sommermonaten abends und in der Nacht in der Innenstadt die Hitze staut und durch eine enge Bebauung kein Luftaustausch stattfinden kann. Die erste Feststellung aller Bearbeiter von Satelliten-Thermalbildern ist, dass sich (große) Städte aufgrund der erhöhten Oberflächentemperaturen in Satelliten-Thermalbildern als „hot spots“ zeigen. Die urbane Überwärmung nimmt dabei mit der Größe der Städte grundsätzlich zu. So werden im Jahresmittel durchschnittlich 0,5 bis 1,5 °C mehr gemessen. An windschwachen, wolkenarmen Tagen kommen Differenzen von 2-6 °C vor. Die urbane Wärmeinsel variiert lokal allerdings sehr stark, wobei die Bebauung, bzw. die Landnutzung einen erheblichen Einfluss darauf hat, ZHOU ET AL., 2018. Mit der Überwärmung wird auch viel Energie umgesetzt, die auch in Form von Wärmeströmen quantifiziert werden können, z.B. KATO, S. UND YASUSHI, Y. (2005). Allerdings können jegliche Temperaturdifferenzen schnell durch aufkommenden Wind aufgelöst werden, was für die Küstenstadt Rostock häufig zutrifft.

## 2 Multitemporale Satellitenbildauswertung

Zur Ermittlung der thermischen Überwärmung wird üblicher Weise eine einzige Thermalszene, bzw. eine Kombination aus einer Tag- und Nachtszene verwendet. Aufgrund verschiedener Einflussfaktoren (z.B. Jahreszeit, Vorwitterung, nicht perfektes Wetter bei der Thermalbefliegung, Atmosphäreinfluss etc.) kann diese unter Umständen nicht repräsentativ sein. Aus diesem Grund ist ein Mittelwert über mehrere Szenen repräsentativer und darüber hinaus auch valider, denn der Mittelwert einer Zeitreihe beschreibt die thermische Überwärmung über einen längeren Zeitraum und nicht nur zu einem einzigen Zeitpunkt. Aus diesem Grund wurde eine Zeitreihe von thermalen Tagaufnahmen 1998 – 2018 gebildet. Diese lange Zeitreihe lässt darüber hinaus auch Rückschlüsse über die Veränderungen innerhalb der letzten 20 Jahre zu.

Die Landsat Satelliten überfliegen die Hansestadt Rostock in den Orbits 193 und 194 etwa alle 8 Tage, so dass maximal für die Sommermonate (Ende März – Anfang September) ca. 15 Szenen pro Jahr aufgenommen werden können. Da Wolken sehr häufig den freien Blick zur Erdoberfläche behindern konnten für den Zeitraum von 1998 bis 2018 jedoch nur insgesamt 47 wolkenfreie Landsat-Szenen der Monate (Ende) März – (Anfang) September identifiziert werden. Auswahlkriterien waren in erster Linie Wolkenfreiheit, was auch Schleierwolken oder Nebel miteinschließt. So kann es passieren, dass in einem Jahr nur eine wolkenfreie Satellitenszene zur Verfügung steht und einem anderen Jahr fünf Szenen.

Über diesen langen Zeitraum sind Thermaldaten der Landsat Satelliten 5, 7 und 8 verwendet worden. Bei der Abtastung des Landsat 7 Satelliten, kam es im Jahr 2003 zu einem Ausfall der Scanwinkelkorrektur. Das bedeutet, dass in den nach dem Jahr 2003 aufgenommenen Szenen Daten fehlen und sich dies in Form von Streifen von 1 – 20 Pixel Breite bemerkbar macht. Aus diesem Grund wurden von diesem Satelliten so wenige Szenen wie möglich verwendet. Landsat 8 liefert seit 2013 Daten. Obwohl die an Bord der Satelliten befindlichen Thermalsensoren mit einer Bodenauflösung von 120 m bei Landsat 5, bzw. 60 m bei Landsat 7 und 100 m für Landsat 8 prinzipiell ähnliche Spezifikationen aufweisen, sind die Daten (Temperaturen) der drei Satelliten nicht direkt vergleichbar, bzw. müssen unterschiedlich aufbereitet werden, um aus ihnen vergleichbare Oberflächentemperaturen abzuleiten (PARASTATIDIS et al. 2017).

Der Download der bereits auf die Oberflächentemperatur korrigierten Thermaldaten erfolgte über die Webseite: [http://rslab.gr/downloads\\_LandsatLST.html](http://rslab.gr/downloads_LandsatLST.html). Diese Webseite greift auf die Google Earth Engine zu und nutzt die Metadaten zu den Satellitenszenen, um in wenigen Schritten die Grauwerte (DN) in Strahldichten umzurechnen, die Strahldichten anschließend in Strahlungstemperaturen umzuwandeln und unter Berücksichtigung der aus einem Vegetationsindex NDVI die Emissivität der Oberflächen (LSE) zu berechnen. Darauf aufbauend kann abschließend die Oberflächentemperatur (engl. LST, bzw. Land Surface Temperature) zuverlässig geschätzt werden. Die Abweichungen zwischen der berechneten Oberflächentemperatur und am Boden gemessenen Werten sind üblicher Weise  $< 3^{\circ}\text{C}$  (PARASTATIDIS et al. 2017).

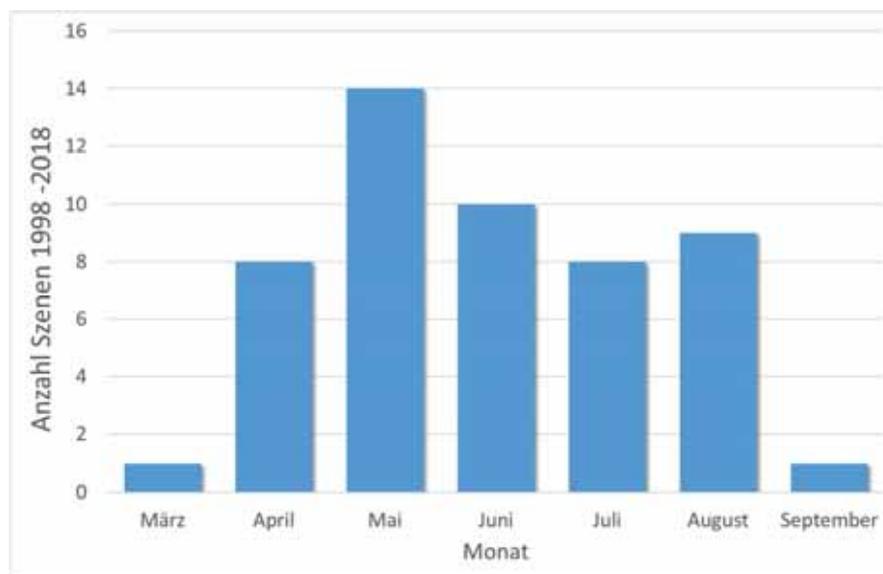


Abb. 1: Aufnahmemonate der 47 Thermalszenen der Zeitreihenanalyse von 1998 - 2018

Anhand der zeitlichen Verteilung der in die Analyse aufgenommenen Satellitenszenen über die Monate sieht man, dass in Rostock gerade der Mai häufig Wolkenfreiheit gewährleistet und der Juli normalerweise der Monat mit den meisten Wolkentagen ist, was dazu führt, dass in diesem Sommermonat die wenigsten Aufnahmen in die Zeitreihenanalyse aufgenommen werden konnten (Abb. 1).

Grundsätzlich sind die Oberflächentemperaturen, die aus der Satellitenperspektive gemessen werden, in erster Linie eine Funktion der Einstrahlung, die im Laufe des Jahres erheblich schwankt, vgl. Abb. 2. Wie aus der Abb. 3 ersichtlich ist, steigen aufgrund der geographischen Lage Rostocks, die Mittelwerte der Landoberflächen bei Sonnenschein, wie er auch auf den Satellitenbildern zu sehen ist, im Jahresgang ab dem Mai stark an. Wie aus Abb. 3 ebenfalls deutlich wird, können die Lufttemperaturen innerhalb eines Monats durchaus schwanken.

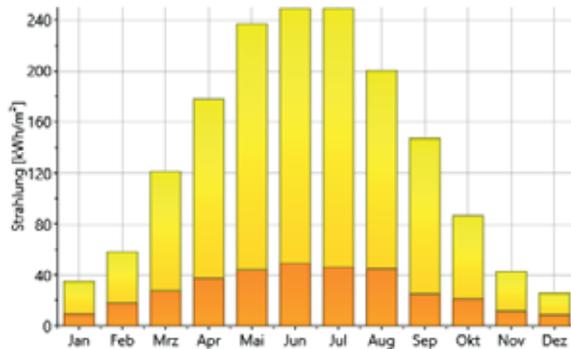


Abb. 2: Einstrahlung im Jahresverlauf bei Schönwetterlagen für Rostock (Meteonorm)



Abb. 3: Jahresgang der Lufttemperatur für Rostock bei Schönwetterlagen (30 Jahresmittel 1980 – 2010), Quelle DWD u. Meteonorm

Nun stellt sich die Frage, ob und inwieweit die in den Satellitenszenen gemessenen Oberflächentemperaturen auch mit dem Strahlungs- und Temperaturverlauf der Luft übereinstimmen. Um das zu untersuchen wurden jeweils 20 Punkte mit bekannter Landnutzung (Industrie, Innenstadt, Großwohnsiedlung, Grünland und Wald) bestimmt. Anschließend wurden die Oberflächentemperaturen aller Szenen abgegriffen und dem Datum nach sortiert (Abb. 4).

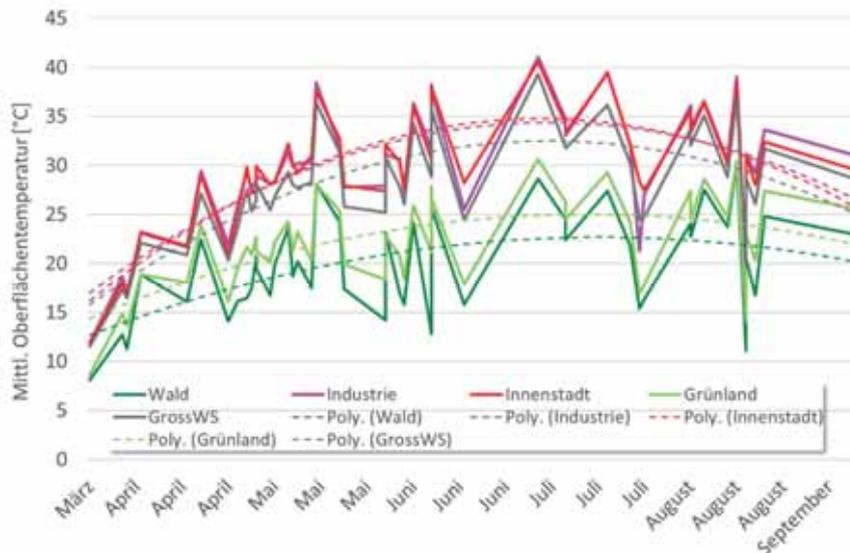


Abb. 4: Jahreszeitlicher Verlauf der Oberflächentemperaturen aus 47 Landsat Szenen von 1998 – 2018 in Abhängigkeit der Landnutzung

Wie aus der Abb. 4 gut zu erkennen ist, schwanken die gemessenen Oberflächentemperaturen im Jahresgang und zwischen den Szenen deutlich. Das ist auch nicht weiter verwunderlich, da diese nicht nur von der Einstrahlung abhängen, sondern auch vom vorhergehenden Witterungsverlauf, der Feuchte, atmosphärische Bedingungen etc. Generell nehmen die gemessenen Oberflächentemperaturen erwartungsgemäß in den Sommermonaten zu und damit verbunden, wird der Unterschied zwischen den verschiedenen Landnutzungsklassen auch größer. Daher ist im Sommer der

Unterschied der Oberflächentemperaturen z.B. zwischen Wald und Industrieflächen größer als im Frühjahr oder Herbst. In Waldgebieten ist der Jahresgang wesentlich geringer ausgeprägt. Um den saisonalen Effekt zu minimieren und die Szenen untereinander besser vergleichbar zu machen, sind die Szenen normiert worden, d.h. die gemessenen Oberflächentemperaturen werden durch den Mittelwert der Szene geteilt und in Prozent umgerechnet. Da die Oberflächentemperaturen von Ackerflächen im Jahresgang sehr stark schwanken und damit auch den Mittelwert beeinflussen, werden diese Flächen nicht in die Mittelwertbildung mit einbezogen, bzw. ausmaskiert. Im Ergebnis der Normierung fällt der jahreszeitliche Effekt weg, wie die Abb. 5 sehr anschaulich verdeutlicht. Somit können alle Szenen gleichgewichtet in die Mittelwertbildung mit einbezogen werden.

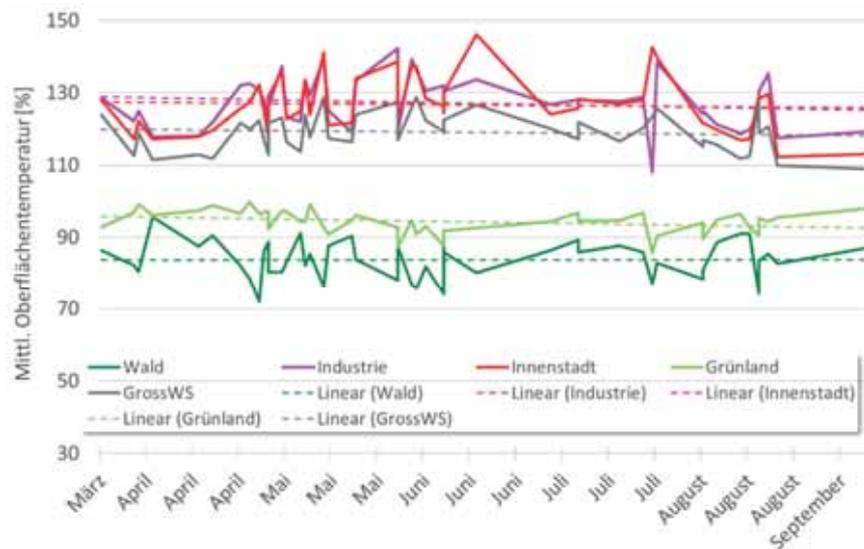


Abb. 5: Jahreszeitlicher Verlauf der Oberflächentemperaturen aus 47 Landsat Szenen von 1998 – 2018 in Abhängigkeit der Landnutzung, als Prozentangabe, normiert in Abhängigkeit der mittleren Oberflächentemperatur ohne Acker- und Wasserflächen

### 3 Ergebnisse

#### 3.1 Thermische Überwärmung

Die Berechnung der thermischen Überwärmung durch die vorliegende Zeitreihe wurde in zwei Varianten durchgeführt. Zum einen durch die Berechnung des Mittelwerts aus den Oberflächentemperaturen der 47 Szenen, in dem der Detailliertheitsgrad der Daten gezeigt werden soll. Hierbei ist anzumerken, dass durch den jahreszeitlichen Effekt die Aufnahmen aus den Sommermonaten stärker gewichtet werden, als die Frühjahrs- und Herbstaufnahmen. Zum anderen durch den Mittelwert der normalisierten Szenen. Hierbei werden alle Szenen gleich gewertet. Allerdings ist die Skala nicht in °C, sondern ein prozentualer Wert. In beiden Varianten kann die thermische Überwärmung der Stadt, bzw. der versiegelten Bereiche innerhalb der Stadt sehr gut beobachtet werden.

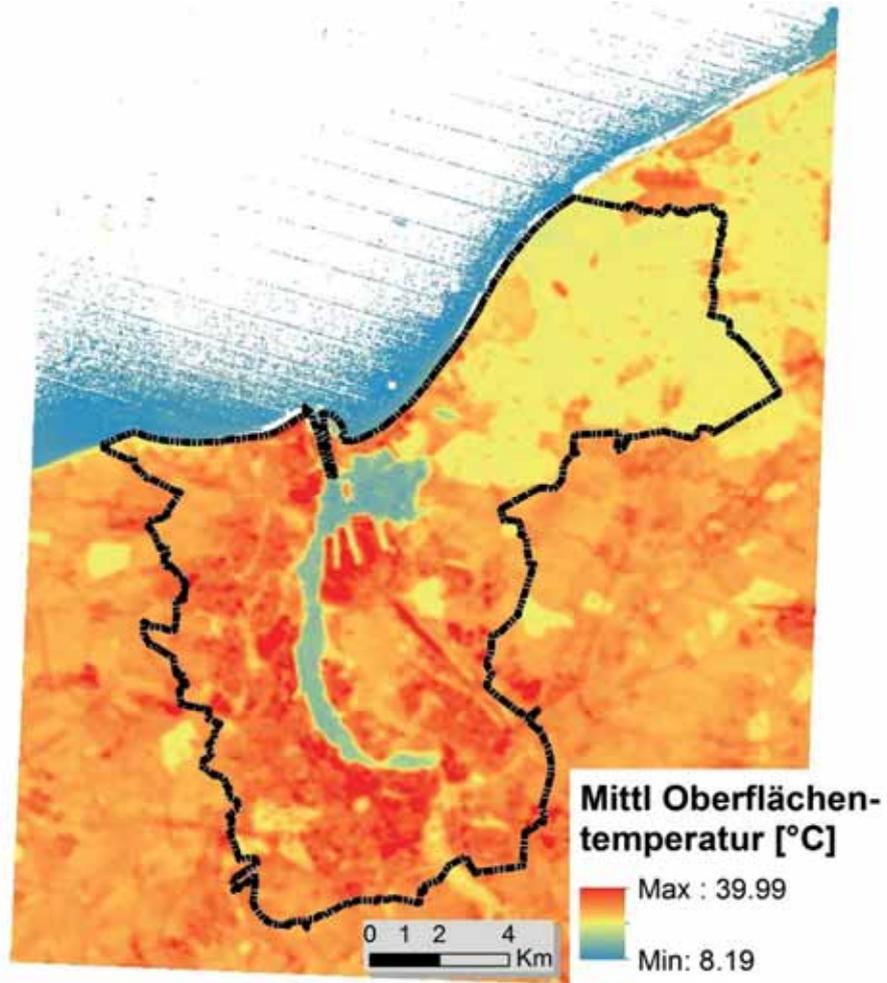


Abb. 6: Mittlere Oberflächentemperatur aus 47 Landsat-Szenen im Zeitraum von 1998 – 2018

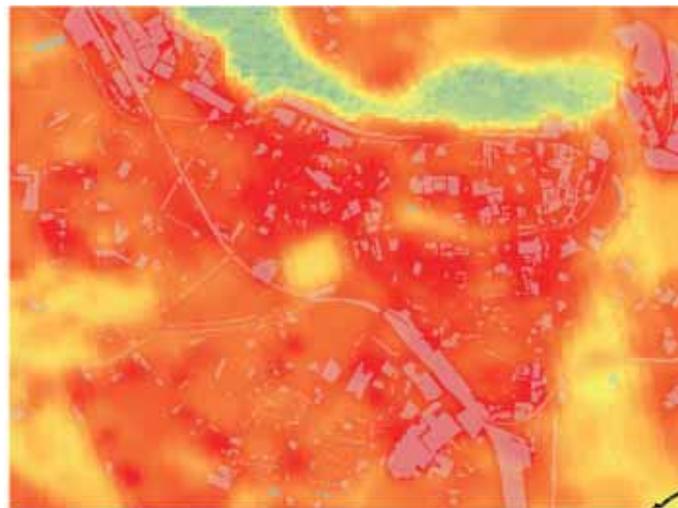


Abb. 7: Ausschnitt der Karte der mittleren Oberflächentemperatur aus 47 Landsat-Szenen im Zeitraum von 1998 – 2018 mit einer Überlagerung der versiegelten Flächen am Beispiel der Innenstadt Rostock

### 3.2 Ableitung einer Karte der thermischen Überwärmung

Die Ausprägung der städtischen Wärmeinsel besitzt abgesehen von der Abhängigkeit vom Wetter sowohl einen Jahres- wie auch einen Tagesgang: Der Jahresgang der Temperaturdifferenzen ist nur schwach ausgeprägt, die Werte der Stadt liegen stets geringfügig über denen des Umlands. Maximale Werte werden im Sommer gemessen, da nachts durch die stark erwärmten Baukörper die langwellige Abstrahlung in den Vordergrund tritt. Das ganze Jahr über sind die Unterschiede zur Mitte des Tages relativ gering, wobei es aber im Sommer passieren kann, dass die Stadt dann durch erhöhte Konvektion und den Schattenwurf der Gebäude in der Innenstadt sogar geringfügig kühler wird als das Umland. Erst gegen Abend und in der Nacht (Sommer) und kurz vor Sonnenaufgang (Herbst/Frühling) werden die hohen Temperaturgegensätze gemessen. Durch diese Abschwächung des Temperaturminimums in der Nacht ergibt sich für die Stadt schließlich eine geringe Tagesamplitude der Temperatur.

Zur Erstellung der Karte der thermischen Überwärmung mit einem Zielmaßstab von 1:50.000, die sich aus der Auflösung der Thermaldaten ergibt, ist es notwendig die signifikanten und zusammenhängenden überwärmten Bereiche zusammenzufassen. Dazu wurde die Mittelwertkarte segmentiert. Die Segmente sollten eine Mindestfläche von 1 ha. haben und kompakter Natur sein. D.h. es sollten langgestreckte, bzw. vielförmige Polygone vermieden werden. Insgesamt wurden für das Gebiet der Hansestadt Rostock 284 Polygone gebildet. Da die Erfassung der Kaltluftproduktionsbereiche aus den Tagaufnahmen nicht möglich ist, wurden zusätzlich die durch das Gelände vorgegebenen Kaltluftproduktionsbereiche einschließlich der existierenden Kaltluftbahnen, die sich durch eine besonders niedrige abendliche Temperatur auszeichnen mit in die Berechnung der thermischen Überwärmung eingearbeitet. Dazu wurden diese Bereiche um eine Überwärmungsstufe herabgestuft, so dass diese in der Regel als starke Kaltluftproduktionsbereiche gelten. Eine Quantifizierung der Überwärmungszonen in °C ist nicht sinnvoll, da das tatsächliche Ausmaß der Überwärmung sowohl im Tagesgang als auch zwischen den sonnigen Tagen, in denen die Überwärmung auftritt stark schwanken kann. Insgesamt wurden 6 Stufen gebildet. Zwei Stufen, die sich auf die Kaltluftproduktionsbereiche beziehen. Wie aus der Abb. 8 ersichtlich, sind das in erster Linie Wälder und Grünlandbereiche. Die Überwärmung des Stadtkörpers wurde in vier Stufen vorgenommen, von keiner Überwärmung bis starker Überwärmung, je nach der Intensität der Versiegelung, bzw. Überbauung,

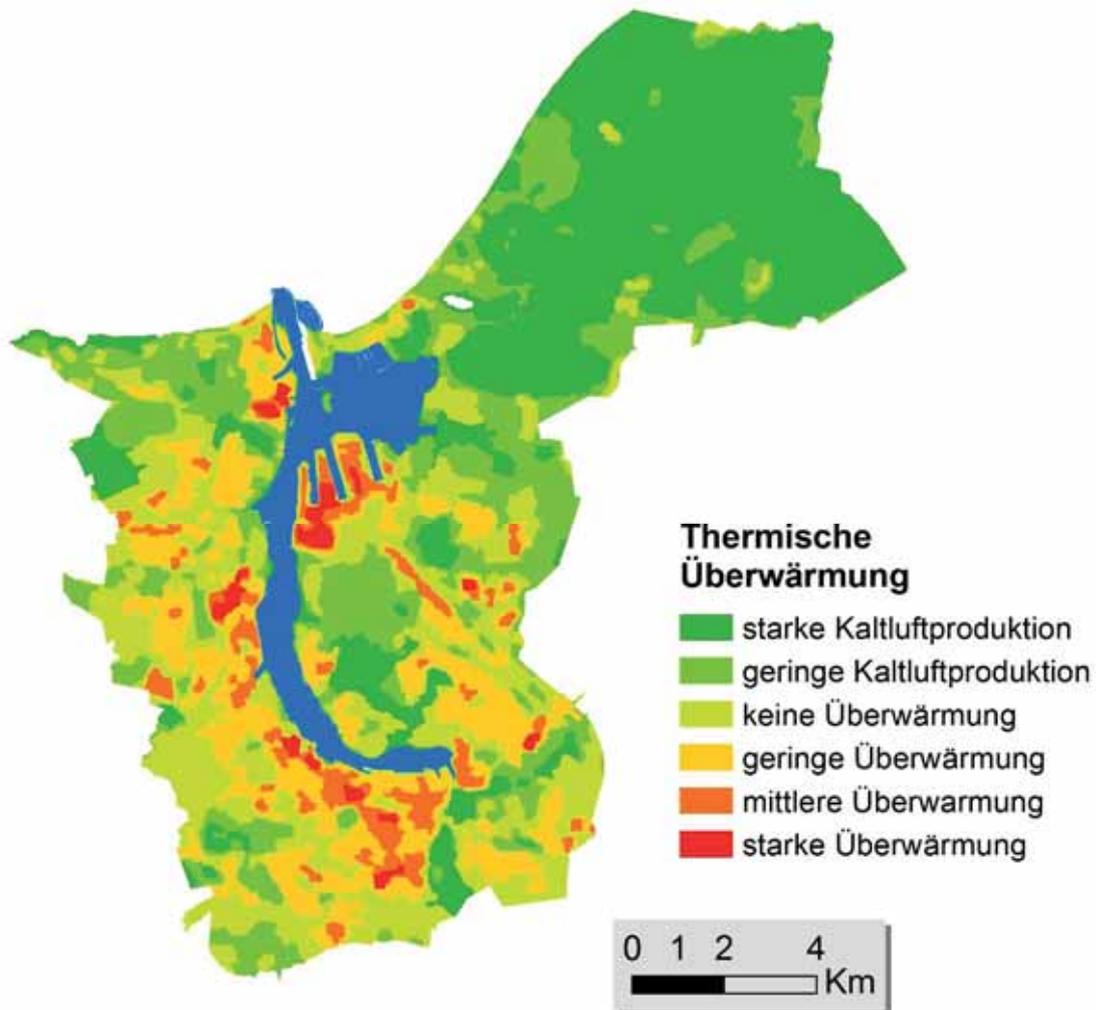


Abb. 8: Thermische Überwärmung der Hansestadt Rostock bei sonnigen Tagen auf der Grundlage langjähriger Thermalsatellitendaten

Zur Abschätzung der Bedeutung der thermischen Überwärmung und der ausgleichenden Kaltluftareale für die Hansestadt Rostock, stellt Abb. 9 diese bezogen auf das Gesamtgebiet der Hansestadt Rostock dar. Demnach tritt eine thermale Überwärmung nur auf 21 % der Stadtfläche auf. Der weit größere Teil des Stadtgebiets ist nicht von der städtischen Wärmeinsel betroffen.

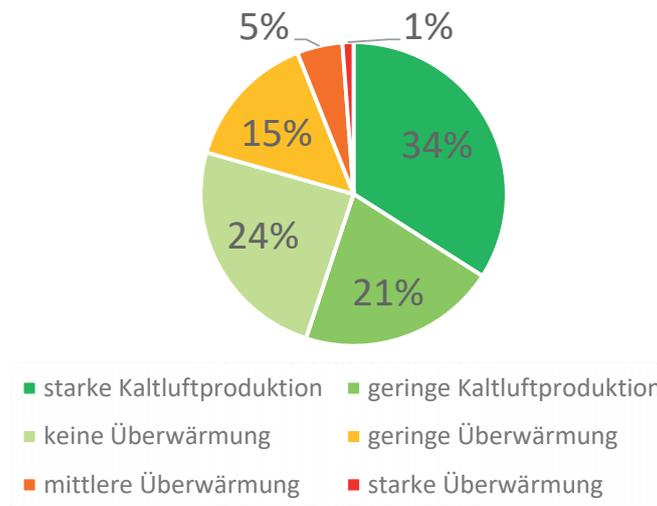


Abb. 9: Verteilung des Grads der thermischen Überwärmung, bezogen auf das gesamte Stadtgebiet

Bezogen auf die Bewohner der Stadt, die sich über das Ausmaß der thermischen Überwärmung informieren wollen, ist diese Statistik wenig zielführend. Vielmehr ist für die Bürger von Interesse, wie groß der Anteil der thermischen Überwärmung an dem urban geprägten Stadtraum ist. Dazu wurde aus der noch aktuellen Realnutzungskartierung von 2014 alle Siedlungs- und Freizeitflächen, inkl. Kleingartenanlagen mit einer Gesamtfläche von 45.7 km<sup>2</sup> zusammengefasst und mit der Karte der thermischen Überwärmung verschnitten. Danach ergibt sich folgendes Bild, Abb. 10. D.h. auf über 50 % der bewohnten und urbanen Fläche, ist bei sonnigem Wetter mit einer thermischen Überwärmung zu rechnen.

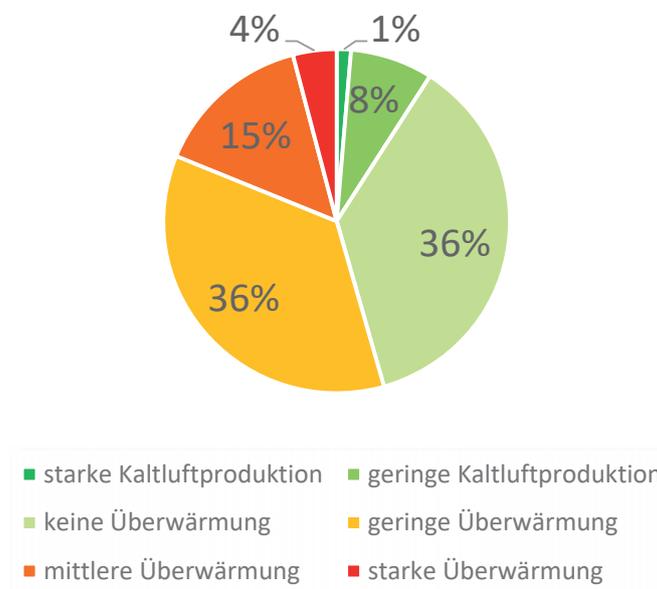


Abb. 10: Verteilung des Grads der thermischen Überwärmung, bezogen auf die teil und voll versiegelten Flächen gemäß der Realnutzungskartierung aus dem Jahr 2014

### 3.3 Bedeutung des Versiegelungsgrads an der urbanen Überwärmung

Wenn man die Überwärmungskarte betrachtet, wird rasch deutlich, dass die Bebauungsdichte, bzw. der Grad der Flächenversiegelung eng mit den Oberflächentemperaturen und der Temperaturamplitude zwischen Tag und Nacht korreliert. Die Flächenversiegelung lässt sich mit dem NDVI-Vegetationsindex recht zuverlässig beschreiben. Der dabei zugrundeliegende Gedanke ist, dass das Verhältnis zwischen roter und nahinfraroter von vegetationsbestandenen Flächen sich deutlich von unbedeckten vegetationsfreien, d.h. versiegelten Flächen unterscheidet. Diese Annahme ist korrekt, solange man ackerbaulich genutzte Flächen von der Berechnung ausschließt. Diese sind nämlich Ende August oftmals unbestellt und werden somit als versiegelte Flächen eingestuft (Abb. 11).

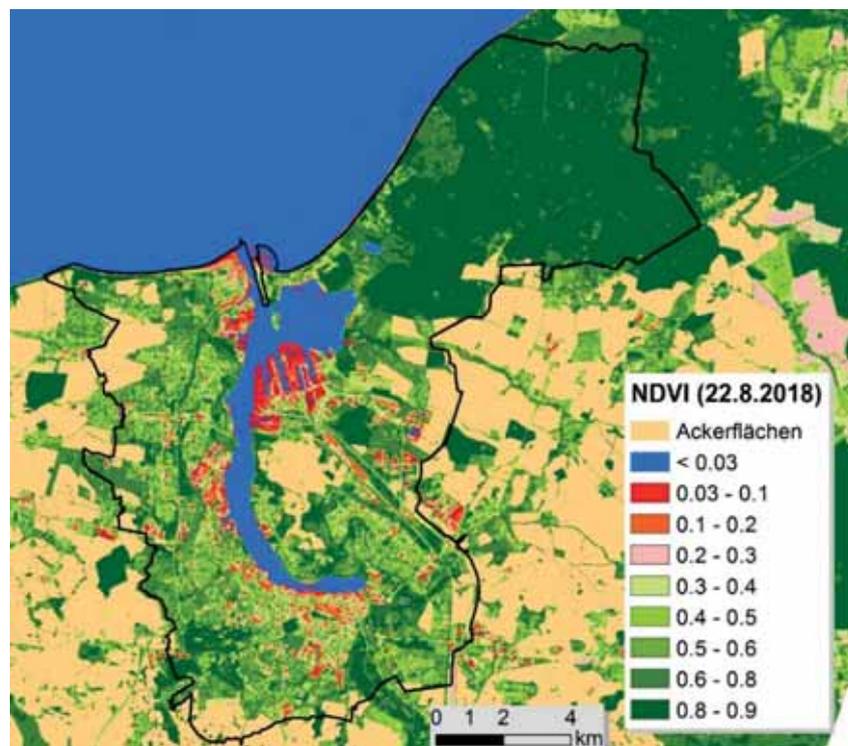


Abb. 11: Vegetationsindex NDVI – abgeleitet für das Landsat 8 Satellitenbild vom 22.08.2018. Ackerflächen sind maskiert

Um den Zusammenhang statistisch beschreiben zu können, wurde für jedes Bildelement der Satellitenszene der korrespondierende NDVI-Wert ermittelt und anschließend zusammengefasst. Grundsätzlich sind NDVI-Werte  $> 0$  Wasserflächen vorbehalten. Werte zwischen  $0 - 0.25$  sind, mit abnehmender Tendenz versiegelte, vegetationslose Bereiche, wie Straßen, Häuser etc. NDVI-Werte  $> 0.25$  stehen für Flächen auf denen vitale Vegetation zu finden ist. Dabei nimmt generell mit zunehmenden Werten der Blattflächenindex zu. So weisen in der Regel Grünland oder Getreideflächen die höchsten NDVI-Werte auf. Der NDVI von Nadelwäldern ist geringer als von Laubwäldern. Abb. 12 zeigt den Zusammenhang zwischen NDVI und Oberflächentemperatur.

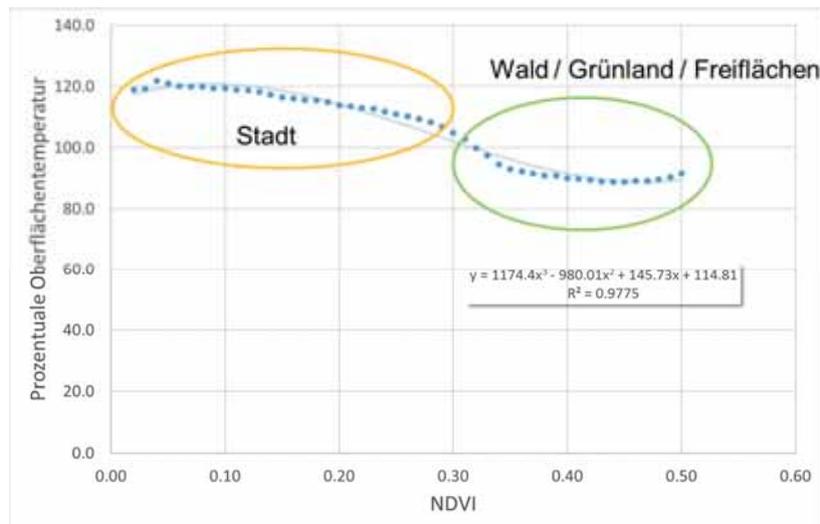


Abb. 12: Oberflächentemperatur in Relation zur Versiegelung (NDVI)

### 3.4 Oberflächentemperaturdifferenz 1998 – 2018

Um die Auswirkungen der Landnutzungsveränderungen der letzten 20 Jahre im Hinblick auf die städtische Überwärmung und die etwaige Beeinträchtigung von Kaltluftbahnen zu untersuchen, wurden die Satellitenbilder des Zeitraums von 1998 – 2002 mit dem prozentualen Mittel der Jahre 2013 – 2018 verglichen (Abb. 13).

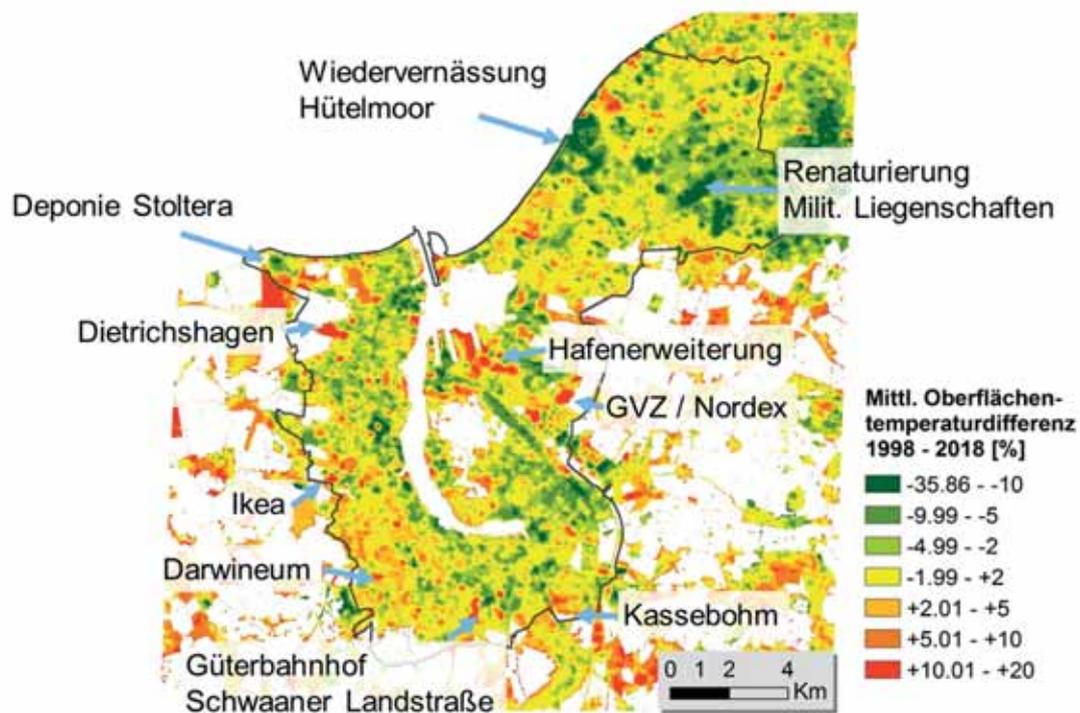


Abb. 13: Mittlere Oberflächentemperaturdifferenz für den Zeitraum 1998 – 2018 in Prozent für die Hansestadt Rostock und Umgebung

In Rostock ist im Beobachtungszeitraum viel gebaut und neue Flächen versiegelt worden. Das ist nicht neu und z.B. auch in der Realnutzungskartierung hinreichend dokumentiert. Dabei zeigt sich, dass größere Bauprojekte zu Veränderungen geführt haben. Wie aus der Karte ebenfalls deutlich wird, hat die Oberflächentemperatur nicht nur zugenommen, vielmehr ist die mittlere Oberflächentemperatur durch eine veränderte Landnutzung auch vielfach gesunken. Das trifft z.B. auf das Hütelmoor oder renaturierte militärische Liegenschaften in der Rostocker Heide zu. Auch sind in diesem Zeitraum die vorhandenen Bäume gewachsen und das städtische Grün hat in den letzten Jahren eher zugenommen als abgenommen. Das ist unter anderem ganz gut in den Großwohnsiedlungen zu beobachten, in denen die Oberflächentemperatur um einige Prozentpunkte zurückgegangen ist.

## 4 Fazit & Ausblick

Stichpunktartig können die Ergebnisse der multitemporalen Satellitenbilddatenauswertung zur Ermittlung der urbanen Überwärmung wie folgt zusammengefasst werden:

- Aus dem langjährigen Mittel der Oberflächentemperaturen zeichnet sich städtische Überwärmung der Hansestadt Rostock bei sonnigen Wetterlagen gegenüber ihrer Umgebung deutlich ab.
- Der Versiegelungsgrad bzw. die Bebauungsdichte bestimmen das Ausmaß der urbanen Überwärmung
- Bau- und Renaturierungsprojekte der letzten 20 Jahre haben die Oberflächentemperaturen, und damit auch das Stadtklima punktuell verändert.

## 5 Literaturverzeichnis

- KATO, S. & YASUSHI, Y., 2005: Analysis of urban heat-island effect using ASTER and ETM+ Data: Separation of anthropogenic heat discharge and natural heat radiation from sensible heat flux. *Remote Sensing of Environment* **99**(1-2), 44-54, doi:10.1016/j.rse.2005.04.026.
- ISAYA NDOSSI, M. & AVDAN, U., 2016: Application of Open Source Coding Technologies in the Production of Land Surface Temperature (LST) Maps from Landsat: A PyQGIS Plugin. *Remote Sensing* **8**(5), 413, doi:10.3390/rs8050413.
- PARASTATIDIS, D.; MITRAKA, Z.; CHRYSOULAKIS, N. & ABRAMS, M., 2017: Online Global Land Surface Temperature Estimation from Landsat. In: *J Remote Sensing* **9**(12), 2072-4292, doi:10.3390/rs9121208.
- ZHOU, D, XIAO, J., BONAFONI, S., BERGER, C., DEILAMI, K. ZHOU, Y., FROLKIN, S. , YAU, R., QIAO, Z. & SOBRINO, J.A., 2018: Satellite Remote Sensing of Surface Urban Heat Islands: Progress, Challenges, and Perspectives. *Remote Sensing* **11**(1), 48, doi: 10.3390/rs11010048.

## Prozessierung bundesweiter Geodaten für hydraulische Modellierung

**MICHAEL H. WIMMER<sup>1</sup>, MARKUS HOLLAUS<sup>1</sup>, GOTTFRIED MANDLBURGER<sup>1</sup>, ANDREAS BUTTINGER-KREUZHUBER<sup>1,2</sup>, JÜRGEN WASER<sup>2</sup>, JÜRGEN KOMMA<sup>1</sup> & NORBERT PFEIFER<sup>1</sup>**

*Zusammenfassung: Hochauflösende räumliche Daten und davon abgeleitete Informationen sind ein bedeutender Input für hydraulische Simulationen, müssen aber vorab für die entsprechenden semantischen und algorithmischen Erfordernisse adaptiert werden. In dieser Arbeit wurden Höhenmodelle und Vektordaten des gesamten österreichischen Bundesgebiets für die hydraulische Simulation vorbereitet bzw. davon ausgehend weitere Datensätze abgeleitet (z.B. Wasser-Land-Grenzen). Angesichts der Menge und Heterogenität der Daten lag dabei besonderer Fokus auf der Automatisierung und der Anpassung der vorhandenen Datensätze untereinander. Im Sinne höchster Zuverlässigkeit der Ergebnisse verlangen die meisten Arbeitsschritte zumindest ein geringes Maß manueller Kontrolle, wobei der Aufwand durch Qualitätsindikatoren eingegrenzt werden kann.*

### 1 Motivation und Einleitung

Das Projekt HORA 3.0 hat eine österreichweit konsistente Hochwasserrisikoanalyse durch hydraulische Simulation zum Ziel. Konsistenz erfordert eine möglichst homogene Datengrundlage, andererseits aber auch eine einheitliche Methodik, sodass die Ergebnisse bundesweit vergleichbar sind. Im Gegensatz zu lokalen Detailuntersuchungen ergibt sich somit der Vorteil einer großen Gebietsabdeckung bei gleichzeitig hoher räumlicher Auflösung im Vergleich zu anderen überregionalen Modellen.

Die hydraulische Simulation benötigt neben der Hydrologie (Abflusswerte, ...) vor allem topographische oder andere raumbezogene Daten als Input. Konkret stehen ein Digitales Geländemodell (DGM, Auflösung 1 m), ein Digitales Oberflächenmodell (DOM, 1 m), das Flussachsen-Netzwerk (Abb. 1), gemessene bzw. berechnete Gewässerprofile, eine Digitale Katastralmappe (DKM) sowie weitere Vektor-Datensätze für Seen, Kraftwerke etc. zur Verfügung. DGM und DOM stammen von zahlreichen separaten Airborne Laser Scanning (ALS) Kampagnen, die in den entsprechenden Bundesländern durchgeführt und ausgewertet wurden. Die Flussachsen sind Teil des österreichischen Berichtsgewässernetzes (BGN) und liegen als Vektordatensatz vor. Das BGN ist der österreichische Referenzdatensatz für die Darstellung gewässerbezogener Daten gemäß EU Wasserrahmenrichtlinie (EU 2000) und EU Hochwasserrichtlinie (EU 2007). Andere Datensätze wie Gewässer-Profile oder Seen stammen von Gewässerbezirken, Ingenieurbüros oder Kraftwerksbetreibern und sind folglich in sich relativ heterogen.

---

<sup>1</sup> Technische Universität Wien, Karlsplatz 13, A-1040 Wien, Österreich  
E-Mail: [Michael.Wimmer, Markus.Hollaus, Norbert.Pfeifer, Gottfried.Mandlburger]@geo.tuwien.ac.at, andreas.buttinger-kreuzhuber@tuwien.ac.at, komma@hydro.tuwien.ac.at

<sup>2</sup> VRVis Zentrum für Virtual Reality und Visualisierung, Donau-City-Str. 11, A-1220 Wien, Österreich,  
E-Mail: jwaser@vrvis.at

Aber auch die bundesweiten Daten sind weder bezüglich des Urhebers, noch hinsichtlich der Methodik und des Zeitpunktes der Erhebung vollständig konsistent. Um entsprechend dem Projektziel eine einheitliche Prozessierung und – sofern möglich – vergleichbare Datenqualität gewährleisten zu können, müssen sie vorab zusammengeführt und harmonisiert werden.

Davon ausgehend können die Daten für die Weiterverwendung in der Hydraulik und Visualisierung vorbereitet werden. Das umfasst einerseits die Anpassung bestehender Daten an die Erfordernisse der hydraulischen Simulation, aber andererseits auch basierend darauf die Erzeugung neuer Datensätze, die in der weiteren Verarbeitung benötigt werden.

Die Verwendung ALS-basierter topographischer Daten und deren Aufbereitung und Bedeutung für Hydrologie und Hydraulik wird in der Literatur umfangreich thematisiert (z.B. MANDLBURGER 2006; MANDLBURGER & BRIESE 2007; HOLLAUS et al. 2005; CASAS et al. 2006; MARKS & BATES 2000; PETROSELLI 2013). Der hauptsächliche Unterschied zu dieser Arbeit liegt darin, dass der Fokus meist auf einzelnen oder mehreren vergleichsweise kleinen Testgebieten liegt, in denen mit homogener Datengrundlage zu rechnen ist. Damit kommt Flexibilität und Automatisierungsgrad der verwendeten Methoden teilweise etwas geringere Bedeutung zu.

Zudem variieren die konkreten Anforderungen an Aufbau und Inhalt der weiterzuverarbeitenden räumlichen Datensätze. Am besten vergleichbar ist die Aufgabenstellung bei der Wasser-Land-Grenze, die als Übergang von der Wasseroberfläche zu trockener Topographie recht klar definiert ist. Die veröffentlichten Methoden basieren häufig auf (semi-)automatischer Auswertung von Luftbildern (GÜNERALP et al. 2013; ALI et al. 2015) oder auf ALS-Daten und davon abgeleiteten Modellen (MANDLBURGER 2006; PÉNARD & MOREL 2012; HÖFLE et al. 2009).

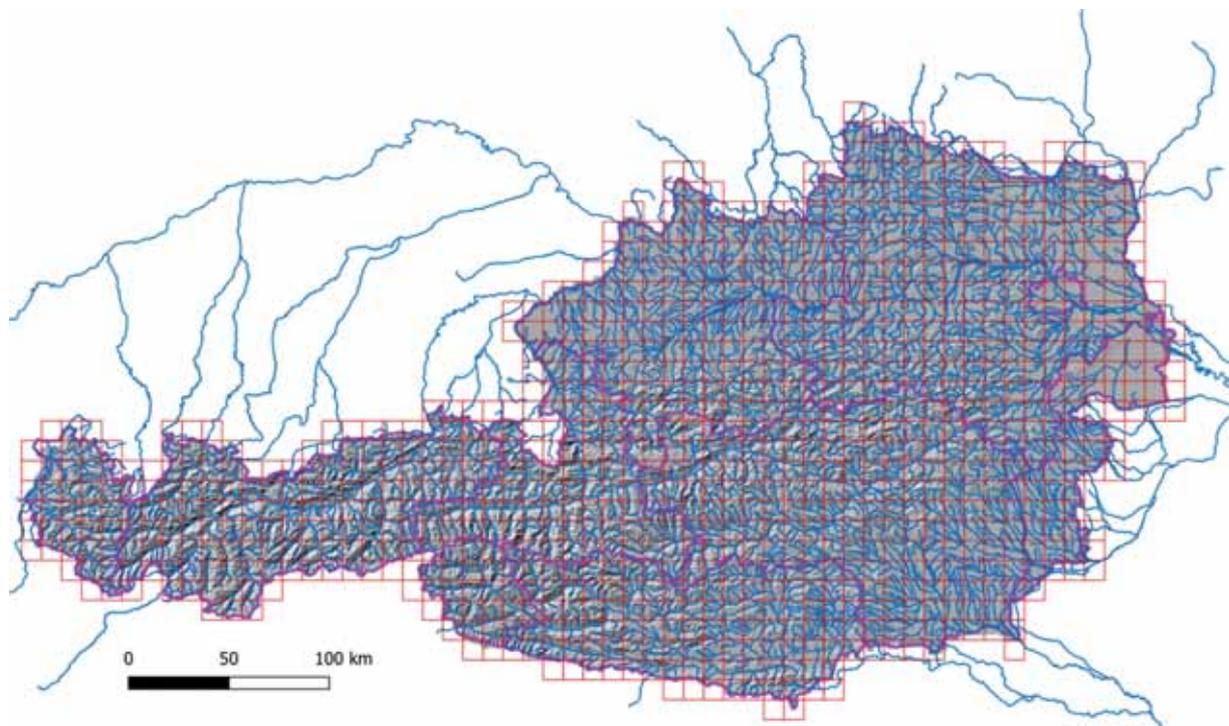


Abb. 1: Das Projektgebiet. Überblick über das verwendete Flussnetzwerk (blau) sowie das österreichweit vorliegende DGM das in Kacheln zu je 10 km Seitenlänge organisiert wird (Kachelgrenzen: rot)

## 2 Methoden

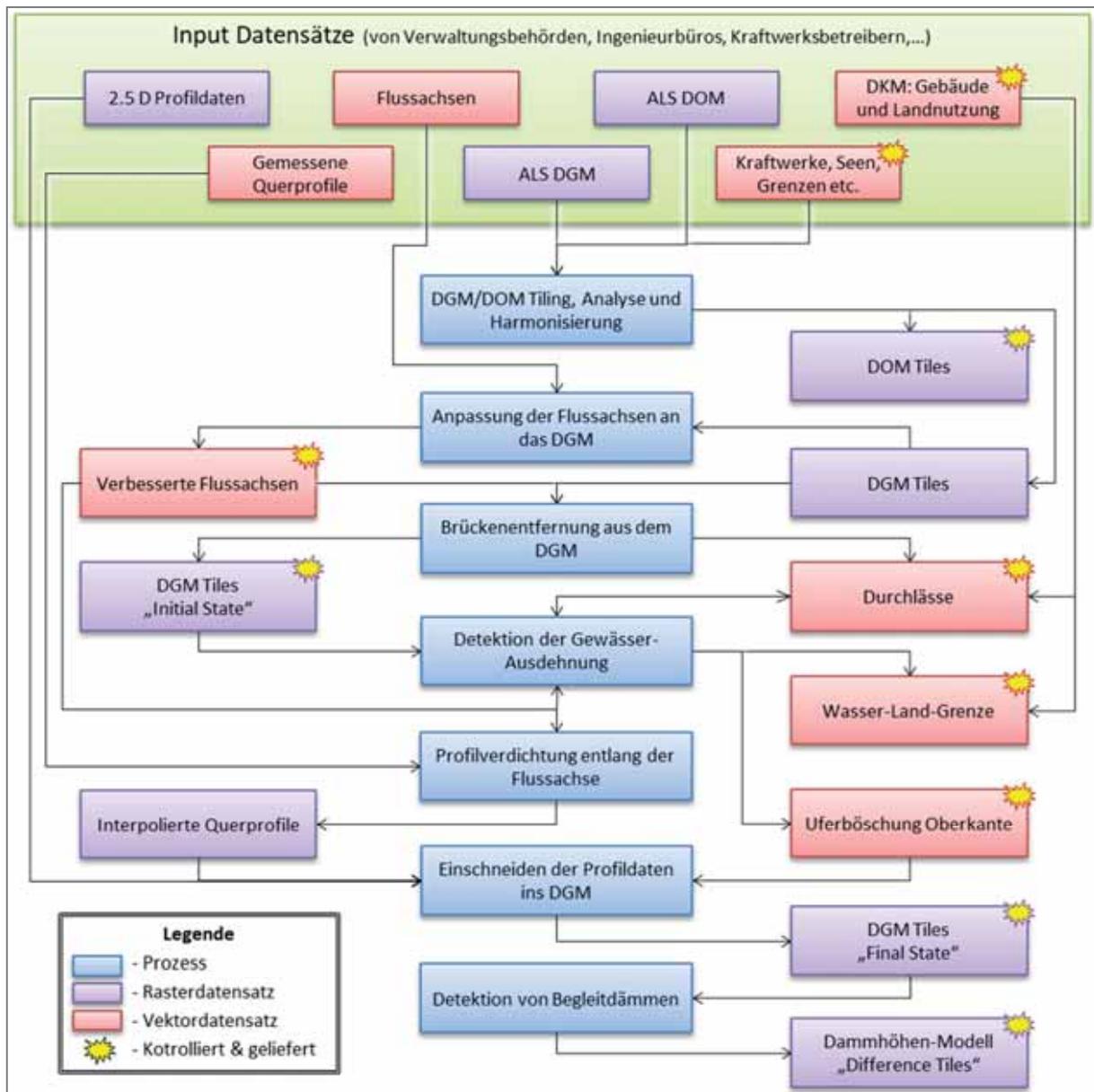


Abb. 2: Übersicht über die wichtigsten Bearbeitungsschritte und Ergebnisse. Jene Produkte, die für Weiterverarbeitung gedacht sind, wurden gelb markiert, der Rest sind interne Zwischenergebnisse

Die besondere Herausforderung bei diesem Projekt besteht in der großen Ausdehnung des Studiengebiets und der damit verbundenen Datenmenge. Konkret werden Flüsse mit einer Gesamtlänge von knapp 40.000 km sowie DGM und DOM mit rund 84.000 km<sup>2</sup> Fläche verwendet; die DKM weist über 3 Mio. Gebäude oder Gebäudeteile aus. Dies erfordert einen hohen Automatisierungsgrad in der Prozessierung, aber auch aussagekräftige Qualitätskriterien, um allfällige manuelle Korrekturen zielsicher einsetzen zu können und somit eine hohe Zuverlässigkeit aller

Ergebnisse zu gewährleisten. Die entscheidenden Arbeitsschritte werden im Folgenden näher beschrieben. Die Implementierung setzt sich aus Matlab- und Python-Scripts zusammen, wobei Letztere stark auf eine Einbindung der Punktwolken-Prozessierungssoftware OPALS (PFEIFER et al. 2014) setzen. Für GIS Operationen bzw. manuelle Kontrolle und allfällige Korrekturen kommt QGIS zum Einsatz. Der grobe Ablauf und die Zusammenhänge zwischen den Arbeitsschritten sind in Abb. 2 dargestellt.

## 2.1 Datenharmonisierung

Wie schon angesprochen, sind die Ausgangsdaten a priori nicht miteinander konsistent. Zwischen den verschiedenen Bundesländern bestehen beispielsweise Unterschiede hinsichtlich der Methoden zur DGM Erzeugung. Zudem verteilt sich die Erhebung der Daten auf einen Zeitraum von insgesamt 15 Jahren, was temporale Dekorrelation zur Folge hat. Diese Unterschiede müssen (i) innerhalb der Höhenmodelle an Landes-, Tile- oder Flugkampagnen-Grenzen berücksichtigt werden (siehe 2.1.1), vor allem aber auch (ii) zwischen unterschiedlichen Datensätzen (2.1.2).

### 2.1.1 Kontrolle und Harmonisierung des Geländemodells

Ziel der Datenaufbereitung ist idealerweise ein flächendeckend konsistentes DGM, in dem gleiche Objekte in der gleichen Art und Qualität repräsentiert sind, was insbesondere bei Gewässern (stehend und fließend), Brücken oder Bruchkanten initial nicht der Fall ist. Aufgrund der Datenmenge und mangels Informationen über Prozessierung und Datengrundlage ist jedoch eine flächendeckende Nachbearbeitung des DGM nicht sinnvoll realisierbar. Stattdessen werden bestehende Unterschiede intern dokumentiert und bei der Implementierung bzw. Parametrisierung weiterer Arbeitsschritte berücksichtigt.

Tatsächliche Anpassung erfolgt etwa bei aufgestauten Flüssen oder Seen an Landesgrenzen, da oftmals von den jeweiligen Bundesländern unterschiedliche Wasserhöhen modelliert werden, woraus sich am Übergang Sprungstellen ergeben. Mithilfe der Umriss des entsprechenden Gewässers wird der gesamte Wasserspiegel durch einfache GIS-Operationen auf eine konstante, gemittelte Höhe gebracht. Kleinräumigere Artefakte, die oftmals an den Grenzen der originalen Tiles auftreten, werden ebenfalls eliminiert, falls sie hoch genug sind, um auf die Interpretation des DGM Einfluss zu haben. Dazu werden die entsprechenden Artefakte aus dem Modell entfernt und durch Interpolation der Umgebung wieder befüllt.

Im Gegensatz zum DGM wird das Oberflächenmodell hauptsächlich als Hilfestellung bei der manuellen Kontrolle und zur Ableitung näherungsweise Gebäudehöhen für die Interpolation verwendet. Dementsprechend sind die Genauigkeitsansprüche im Vergleich zu anderen Datensätzen (z.B. Flussachsen) geringer und eine grobe Plausibilitätskontrolle der resultierenden Gebäudehöhen ist ausreichend.

### 2.1.2 Anpassung von Vektordaten an das DGM

Die aus unterschiedlichen Quellen stammenden und mit unterschiedlichen Methoden erhobenen Vektordaten stimmen in der Regel nicht ausreichend gut mit dem DGM überein. Das betrifft beispielsweise die Ausdehnung von Seen oder hydrologische Daten mit explizitem Raumbezug, vor allem aber die qualitativ sehr variablen Flussachsen. Da von diesen fast jedes Folgeprodukt

abhängt, ist ihre Anpassung an das DGM essentiell. Der vorrangige Grund, wieso die Vektor- an die Rasterdaten angepasst werden (und nicht umgekehrt), ist die global deutlich zuverlässigere Lagegenauigkeit der Rasterdaten. Zudem ist die Korrektur von 2D Polylinien (Flussachsen) mit weitaus weniger unsicheren Modellannahmen verbunden als das Verlegen ganzer Flusstäler im DGM.

Methodisch wird diese Korrektur mittels Querprofilen umgesetzt. Diese werden normal an die originalen Flussachsen angelegt. Ihre Ausdehnung wird etwas breiter als der höchste anzunehmende Fehler gewählt, damit das im DGM ersichtliche Flussbett jedenfalls enthalten ist. Im vorliegenden Fall hat sich aus Begutachtung der Daten eine Breite von 250 m ergeben.

Diese Querprofile werden beginnend mit der Quelle eines Flusses durchlaufen und für jedes wird aus verschiedenen Gewichtsfunktionen eine Detektor-Funktion kombiniert, deren Maximum den wahrscheinlichsten Schnittpunkt zwischen Querprofil und Flussachse laut DGM ausweist. Die Höhe des Maximums, absolut bzw. relativ zu anderen lokalen Maxima, gibt Aufschluss über die Zuverlässigkeit der Schätzung.

Im Folgenden sind kurz die Kriterien beschrieben, die für die einzelnen Gewichtsfunktionen verantwortlich sind. Diese sind so geordnet, dass die ersten Kriterien stärker diskriminieren, i.e. die Gewichte haben dort starke relative Unterschiede bzw. werden Bereiche sogar explizit ausgeschlossen. Die unteren Kriterien werden zwar immer mit einbezogen, zeigen aber vor allem dann Einfluss, wenn nicht schon eine klare Vorauswahl getroffen ist, z.B. bei breiten oder kaum definierten Flussbetten.

- Geländehöhe: Niedrige Abschnitte liegen, zumindest was die lokale Umgebung betrifft, wahrscheinlicher im Flussbett als höhere und werden daher stärker gewichtet.
- Durchlaufenes Höhenprofil: Speziell bei sehr nahe benachbarten Flüssen, aber auch in anderen Fällen kann nicht notwendigerweise davon ausgegangen werden, dass der niedrigste Punkt im Querprofil automatisch auch der Flussachse entspricht. Um möglicherweise tiefer liegende, benachbarte Täler auszuschließen, wird der Höhenverlauf seit dem letzten ermittelten Punkt der Flussachse berücksichtigt. Dieser sollte monoton fallend sein, je stärker das Gefälle, umso höher das Gewicht. Geringfügige Anstiege im Bereich der DGM Genauigkeit werden ebenfalls toleriert, größere Anstiege (z.B. mehr als wenige dm akkumuliert über 5 m Wegstrecke) führen zu einem Gewicht 0.
- Flussbettgeometrie: In der Mitte des Flussbetts wird relativ horizontales Gelände oder - bei kleinen Gewässern – ein lokales Minimum erwartet (das ALS DGM repräsentiert bestenfalls die Wasseroberfläche, nie das Flussbett). Entsprechend wird von einer geringen ersten Ableitung und einer geringen oder positiven zweiten Ableitung ausgegangen. Eine betragsmäßig große erste Ableitung bzw. eine negative zweite Ableitung führen zu geringerem Gewicht.
- Horizontale Distanz: Bei ähnlichen Voraussetzungen bezüglich Geländegeometrie wird erwartet, dass Wasser den kürzeren Weg nimmt, der folglich höher gewichtet wird. Dieses Kriterium hilft vor allem, allzu starken Zick-Zack Verlauf der korrigierten Achse innerhalb des Flussbetts zu vermeiden.
- Prädiktion aus dem bisherigen Verlauf: Neben einer kurzen Weglänge ist auch der bisherige Verlauf der Flussachse relevant. Dazu wird durch die zuletzt korrigierten Abschnitte

der Flussachse wird eine genäherte Kurve gelegt und mit dem aktuellen Querprofil verschnitten. Der Schnittpunkt ist das Zentrum einer Gauß-förmigen Gewichtsfunktion.

- Berücksichtigung der Näherungswerte: Das originale Gewässernetz dient schon implizit als Näherung, da es die Platzierung der Querprofile definiert. Explizit wird es ebenfalls durch eine Gauß-Funktion eingeführt, die ihr Zentrum an der originalen Flussachse, also in der Mitte des Querprofils, hat. Diese Kurve wird im Flussverlauf relativ flach gewählt (große Standardabweichung), hat also auch vor allem bei breiten, horizontalen Wasseroberflächen Einfluss. Außerdem wird die Näherung im Bereich der Quelle stärker gewichtet, wo viele der anderen Kriterien mangels klarem Flussbett und vorangegangenen Verlauf noch nicht sehr zuverlässig bestimmt sind.

Die unterschiedlichen Gewichtsfunktionen werden per Multiplikation zur Detektor-Funktion kombiniert. Das hat den Zweck, dass durch ein Kriterium ausdrücklich ausgeschlossene Punkte (Gewicht 0) nicht mehr in Frage kommen. Die verbesserte Lage der Flussachse ergibt sich aus dem Maximum der Detektor-Funktion. Die Ausgeprägtheit des Maximums und seine Höhe relativ zu weiteren Maxima lassen Rückschlüsse auf die Zuverlässigkeit der Schätzung zu.

In Ausnahmefällen (z.B. ausgeprägte Hindernisse im Flussbett) kann die Detektor-Funktion für sämtliche Punkte eines Querprofils 0 ergeben, wodurch kein eindeutiges Maximum bestimmbar ist. In dem Fall wird eine geometrische Extrapolation der unmittelbar zuvor schon angepassten Achsabschnitte verwendet, vgl. „Prädiktion aus dem bisherigen Verlauf“ oben.

## 2.2 Entfernung von Hindernissen im Flussbett

Neben Unzulänglichkeiten der DGM Interpolation über Wasserflächen (z.B. bei überhängender Vegetation oder groben Fehlern) können auch korrekt abgebildete Objekte problematisch für die Hydraulik sein. So etwa erscheinen Brücken in ihrer 2.5 D Repräsentation für die verwendeten Softwarepakete wie Dämme und müssen daher detektiert und aus dem DGM entfernt werden.

Für die Detektion wird das Höhenprofil entlang der (korrigierten) Flussachse herangezogen. In diesem Höhenprofil werden Anstiege gesucht, die über Variationen im Bereich der DGM-Genauigkeit hinausgehen, entsprechend einer hohen Ersten Ableitung des Höhenprofils. In Kombination damit wird auch das wahrscheinliche Ende des Hindernisses als darauf folgender, deutlicher Höhenabfall unter das Ausgangslevel detektiert.

Der so detektierte Abschnitt beinhaltet den oberen Teil des Hindernisses. Der zu eliminierende Abschnitt beginnt und endet aber nicht im Bereich der stärksten Steigung, sondern dort, wo sich das Hindernis vom Flussbett abzuheben beginnt. Die genaue Ausdehnung wird also bestimmt, indem unmittelbar vor und nach dem detektierten Abschnitt des Höhenprofils nach Maxima des Quotienten zwischen zweiter Ableitung und Betrag der ersten Ableitung gesucht wird. Um den Abschnitt jedenfalls ausreichend zu umfassen, wird zusätzlich ein Margin von 2 m beiderseits des Abschnitts eingefügt.

Das Ausschneiden der gefundenen Objekte wird auf Basis von Querprofilen durchgeführt (vgl. MANDLBURGER 2006). Diese werden in ein Profilkordinatensystem transformiert – entsprechend einer Begradigung der Flussachse, so dass alle Querprofile parallel zu liegen kommen. In diesem System wird zwischen den Profilen vor und nach dem zu eliminierenden Objekt interpoliert. Außerdem wird aus der Geometrie des Querprofils die ungefähre Breite des Flussbetts bestimmt (analog zu 2.3.1). Innerhalb dieser Breite werden die interpolierten

Höhenwerte übernommen, also das Hindernis entfernt. Im unmittelbar angrenzenden Uferbereich werden interpolierte und originale Höhen mithilfe gewichteter Mittelwerte ineinander überführt. Das Gewicht der interpolierten Höhen nimmt mit Abstand von der Flussachse ab, sodass vom neu interpolierten Flussbett zum ursprünglichen Höhenmodell ein fließender Verlauf gegeben ist. Beispielhafte Ergebnisse der Brückendetektion und -elimination sind in Abb. 4 dargestellt.

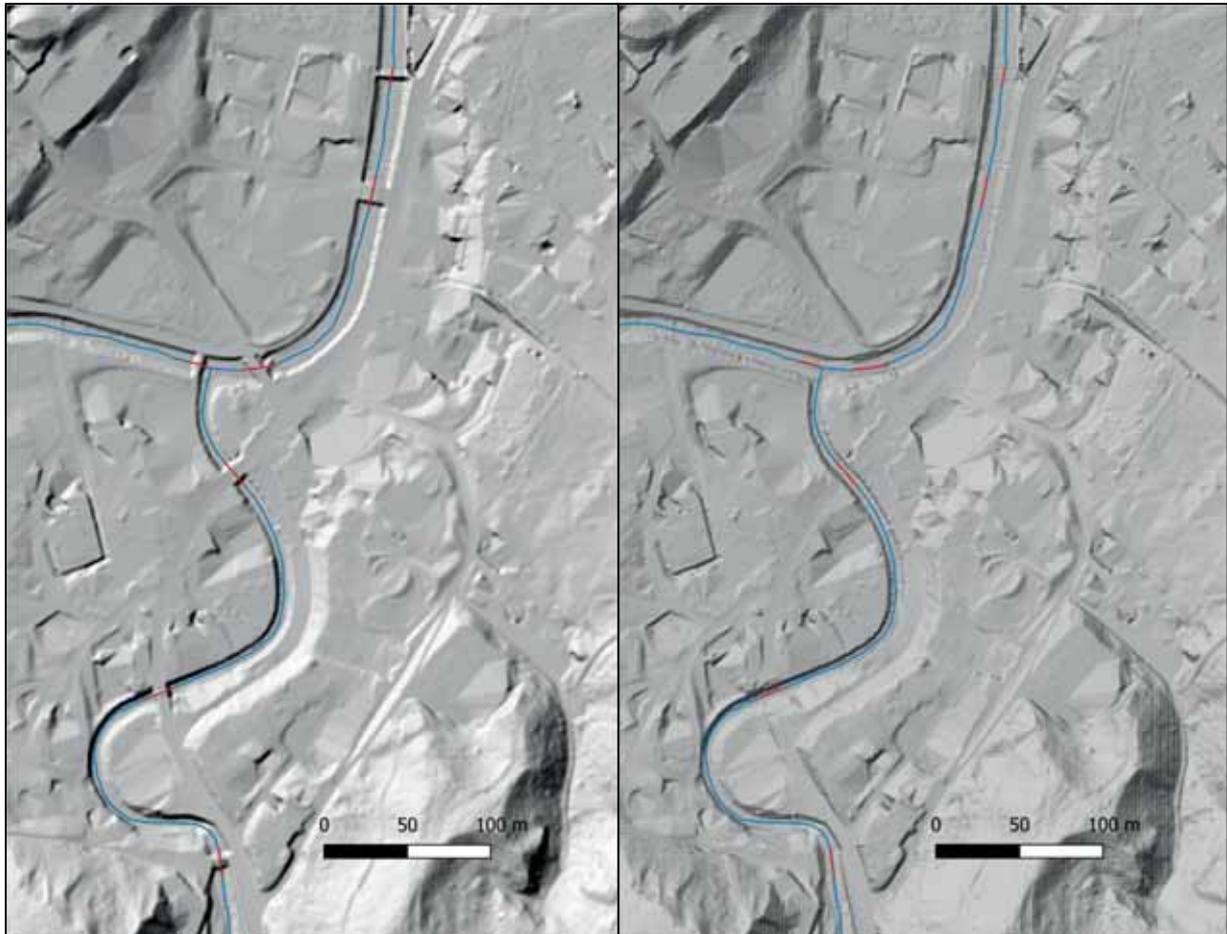


Abb. 3: Brücken-/Hindernisdetektion im Zentrum von Lech am Arlberg (links). In blau sind die Flussachsen eingezeichnet, in rot jene Achsenabschnitte, für die Hindernisse detektiert wurden. Rechts sind die Hindernisse aus dem DGM entfernt

## 2.3 Ableitung neuer Datensätze

Ausgehend von nunmehr geometrisch konsistenten Ausgangsdaten werden weitere Produkte abgeleitet, die für die hydraulische Simulation benötigt werden. Diese sind im Folgenden zusammengefasst.

### 2.3.1 Uferböschung Oberkante

Die Uferböschung Oberkante (UBOK) wird für das Setzen von Randbedingungen benötigt. Ähnlich wie bei der Bestimmung des Flussbetts im Zuge der Hinderniselimination (2.2), werden auch hierfür Querprofile verwendet. Ausgehend von der Flussachse wird das Höhenprofil in beide Richtungen geometrisch analysiert. Der erste signifikante Höhenanstieg im Querprofil ist die

Uferböschung. Dort, wo dieser Anstieg wieder endet (stark negative zweite Ableitung) ist die Oberkante. Falls eine solche Oberkante nicht klar vorhanden ist, wie etwa in alpinen Schluchten, wird stattdessen ein Grenzwert für die Höhendifferenz zum Gewässerbett angesetzt.

In der Praxis ist die zentrale Herausforderung die Definition, was unter einem signifikanten Höhenanstieg zu verstehen ist. Dabei ist einerseits die Repräsentation der Wasseroberfläche relevant, andererseits die Geländeform im Umland. Bei sehr breiten Flüssen oder überhängender Vegetation ist die interpolierte Wasseroberfläche oft sehr unruhig mit deutlichen Höhenunterschieden quer zur Flussrichtung. Im Flachland können die Höhenunterschiede innerhalb der interpolierten Wasseroberfläche in einer ähnlichen Größenordnung wie die Höhe der Uferböschung sein.

Für eine bessere Detektion der tatsächlichen Uferböschung wird die Höhenverteilung im umliegenden Gelände analysiert. Je flacher das Gelände, umso geringere Grenzwerte werden für Steigung und Höhendifferenz verwendet. Außerdem wird die ermittelte Gewässerbreite mithilfe der umliegenden Querprofile einer Plausibilitätskontrolle unterzogen. Allzu sprunghafte Veränderungen erfordern ggf. zumindest eine Nachkontrolle.

Abhängig vom Verlauf der Flussachse und vom Abstand der Querprofile können Selbstkreuzungen bei der resultierenden UBOK auftreten. Diese werden in der Nachprozessierung automatisiert gefunden und eliminiert, indem die auftretende Schleife entfernt wird. Zu Visualisierungszwecken wird die UBOK außerdem mit bestehenden GIS Tools geglättet und vereinfacht.

### 2.3.2 Wasser-Land-Grenze

Die Wasser-Land-Grenze (WLG) liegt etwa in Höhe des Wasserspiegels und wird u.a. für das Einschneiden des Flussbetts in das DGM verwendet. Die Lage der WLG wird analog zur UBOK basierend auf Gewässerquerschnitten ermittelt. Der hauptsächliche Unterschied besteht darin, dass die WLG nicht oberhalb, sondern unterhalb der Uferböschung gesucht wird, näher an der Flussachse. Als Indikator wird das positive Maximum der zweiten Ableitung verwendet, das dort auftritt, wo sich das Gelände von der Wasseroberfläche abhebt.

### 2.3.3 Durchlässe

Durchlässe kommen an Stellen zum Einsatz, wo Gewässer nicht überirdisch verlaufen, das Einschneiden ins Geländemodell aber aufgrund der Länge nicht praktikabel ist bzw. die Ergebnisse der Simulation korrumpieren würde; Beispiele dafür sind Kanalbrücken oder Durchleitungen unter Siedlungsgebieten. Die entsprechenden Abschnitte werden bei der Detektion von Hindernissen (siehe 2.2) und durch Verschnitt des Wasserkörpers innerhalb der WLG mit Gebäuden aus der DKM gefunden und unterstützt durch automatisiert feststellbare Kriterien eingegrenzt.

Da die Unterscheidung zwischen Durchlass oder einem großräumigen Hindernis, das aus dem DGM eliminiert werden kann, nicht allgemeingültig an objektiven Kriterien festgemacht werden kann, ist hier jedenfalls auch manuelle Interpretation notwendig. Dabei werden die vorab lokalisierten Achsenabschnitte durchgegangen und mittels Orthophoto, DGM und DOM analysiert.

### 2.3.4 Begleitämme

Begleitämme (Abb. 4) werden für die Simulation von Dammbuchsenarien detektiert und in ihrer Geometrie beschrieben. Sie sind dadurch gekennzeichnet, sich deutlich und steil vom Ge-

lände abzuheben, im Vergleich zu anderen Objekten mit dieser Eigenschaft sind sie aber deutlich langgezogener.

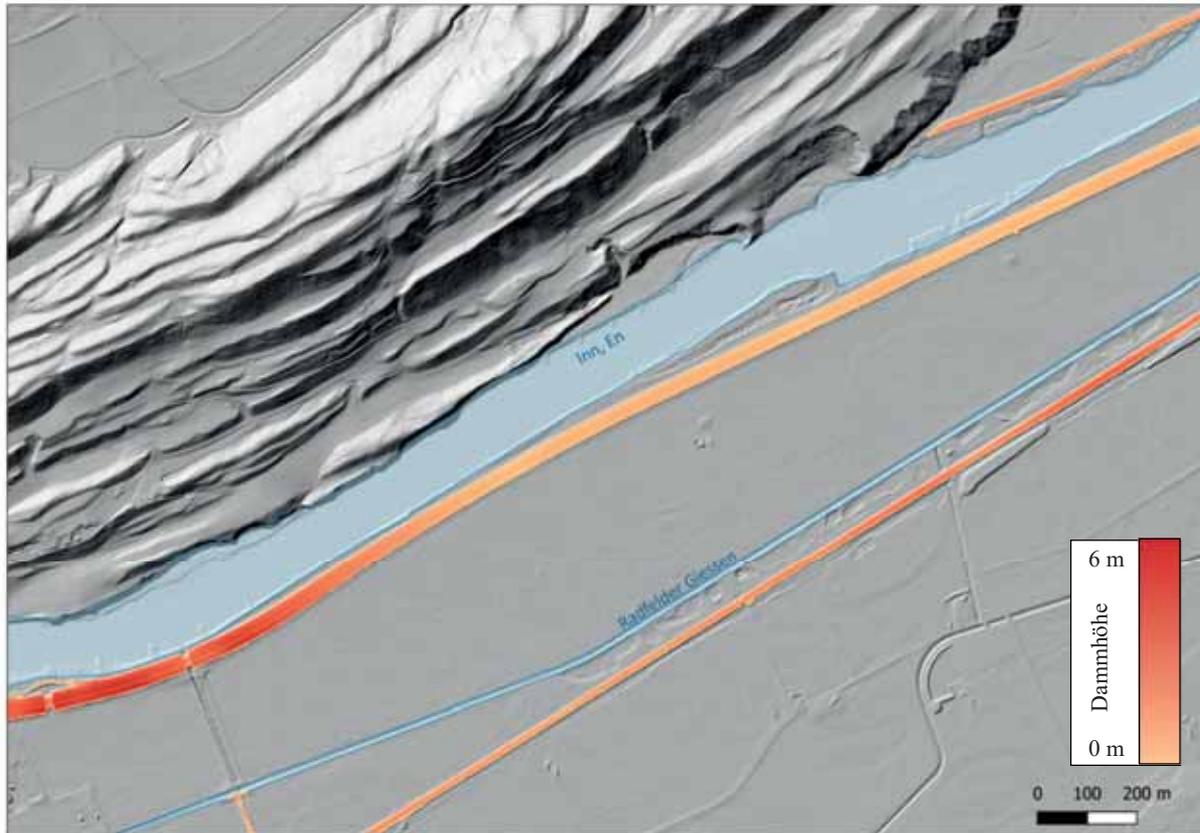


Abb. 4: Nach geometrischen Kriterien detektierte Begleiddämme im Tiroler Inntal. Die Bewertung der Stabilität hängt auch von Bauweise und Nutzung ab, benötigt also weitere Informationen

Vor der Detektion wird der Suchbereich auf das unmittelbare Umland von Flüssen eingeschränkt, da die beschriebenen geometrischen Kriterien im Grunde auch bei Berggraten erfüllt sind. Für die Eingrenzung werden die HQ300 Flächen aus vorangegangenen Untersuchungen verwendet. Im Suchbereich wird dann ein modifizierter Hampel Filter eingesetzt, um jene Objekte zu identifizieren, die sich klar vom Umland abheben. Zusammenhängende Flächen werden vektorisiert und mithilfe von Form- und Flächenkriterien auf potenzielle Dammbauwerke eingeschränkt.

Für die geometrische Beschreibung der Dammhöhen werden diese aus dem DGM entfernt und die entstandenen Lücken durch Interpolation geschlossen. Aus der Differenz zwischen dem originalen und dem auf diese Weise modifizierten DGM ergibt sich die eine geometrische Beschreibung der Dammbauwerke.

## 2.4 Integration von Profilinformatoren ins DGM

Da nicht die Wasseroberfläche sondern das darunter liegende Flussbett für die Simulationen relevant ist, muss dieses ins DGM eingeschnitten werden. Je nach Datenverfügbarkeit seitens der Verwaltungsbehörden/Kraftwerksbetreiber unterscheidet sich der Workflow:

- Bei Verfügbarkeit flächendeckender Repräsentationen des Flussbetts in Form von Rastermodellen können diese nach Prüfung der Konsistenz der Höhensysteme direkt eingefügt werden. Dabei wird das DGM innerhalb der UBOK durch das Modell des Flussbetts ersetzt.
- Falls dagegen nur gemessene Querprofile in Form von 3D Punkten vorhanden sind, müssen diese vorab entlang der Flussachse zu flächigen Höhenmodellen interpoliert werden. Die meist annähernd linear und normal zur Flussachse ausgerichteten Profilmunkte werden zuerst auf ein exaktes Querprofil des Gewässers projiziert. Im Anschluss werden die so erhaltenen Querprofile entlang der Flussachse durch Interpolation verdichtet (vgl. MANDLBURGER 2000; MANDLBURGER 2006). Dies geschieht erneut im Profilkordinatensystem analog zur Elimination von Brücken und Hindernissen (2.2). Diese verdichteten Punktwolken werden zu Rastermodellen interpoliert und ebenfalls innerhalb der UBOK ins DGM eingefügt.
- Wo keine gemessenen Daten über das Flussbett verfügbar sind, wird das DGM nicht verändert, die entsprechenden Bereiche werden aber per Maske ausgewiesen. Im Zuge der Simulation werden dort iterativ Abflussquerschnitte und auf Basis der WLG möglichst realistische Flussbett-Querschnitte ermittelt.

### **3 Ergebnisse und Fazit**

Wie schon in Abschnitt 2.1.2 ausgeführt, ist die Übereinstimmung zwischen Flussachse und DGM entscheidend für fast alle weiteren Verarbeitungsschritte. Die Annahme einer fehlfreien Lage der Flussachsachse bezüglich des DGMs ist zudem in der Folge die Voraussetzung für einen möglichst hohen Automatisierungsgrad. Obwohl die Anpassung der Flussachsen grundsätzlich ebenfalls automatisiert durchgeführt wird, muss aufgrund dieser hohen Erwartungen an die Zuverlässigkeit der Resultate flächendeckend manuell nachkontrolliert werden. Vor allem die heterogene Qualität der Ausgangsdaten macht dies erforderlich. Das betrifft einerseits die Frage, wie gut die Näherung durch die originalen Flussachsen ist, andererseits aber auch das DGM. Vegetationsüberhang, große Wasserflächen ohne ALS-Punkte, die nur mittels Interpolation gefüllt wurden und undeutliche repräsentierte Flussbetten stellen potentielle Fehlerquellen dar. Zu schmale Flussläufe werden im DGM teilweise nicht aufgelöst und können von Bereichen, wo Gewässer unterirdisch verlaufen, nicht automatisch unterschieden werden. Die Anpassung der Flussachsen an das DGM ist somit sicher der arbeitsaufwändigste Schritt.

Darauf basierend führen Korrekturen (2.2) und die Ableitung von Folgeprodukten (2.3, 2.4) automatisiert zu sehr guten Ergebnissen (siehe Abb. 5). Punktuell notwendige Korrekturen können aufgrund von Qualitätskriterien vorab eingegrenzt werden, was den Zeitaufwand für Nachbearbeitung im Vergleich zu flächendeckender Kontrolle massiv reduziert. Dennoch bestehen auch hier regionale Unterschiede, was vor allem auf die Qualität des DGMs zurückzuführen ist.

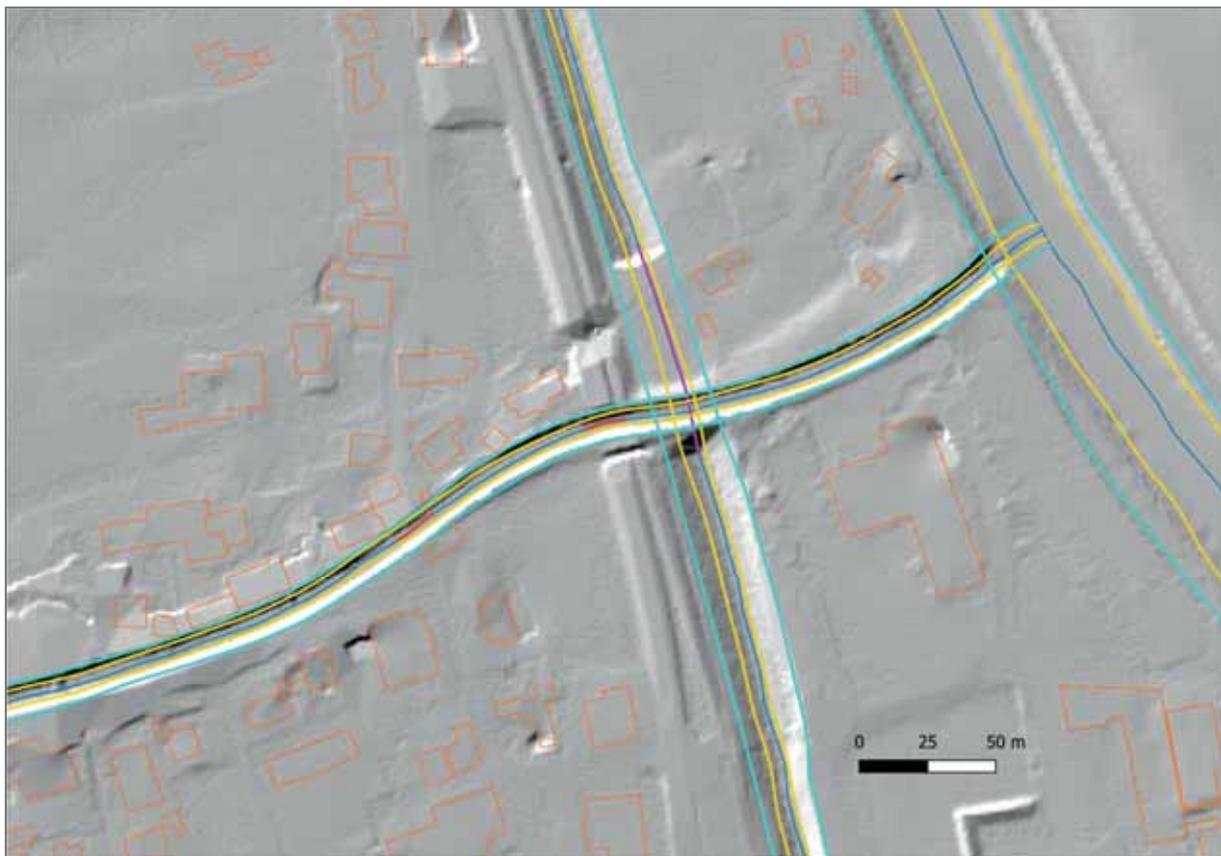


Abb. 5: Die wichtigsten Ergebnisse der Datenprozessierung. Parallel zu den Flussachsen (blau) sind die WLG (gelb) und die UBOK (türkis) eingezeichnet. Rot sind die Flussachsenabschnitte markiert, entlang derer Hindernisse entfernt wurden, in violett ein Durchlass eines Flusses unter einem anderen. Für das Flussbett des Ziller (ganz rechts) wurde das DGM durch Profildaten ersetzt.

Die Validierung erfolgt intern durch stichprobenartige bzw. durch Qualitätsmerkmale gesteuerte manuelle Kontrolle, vor allem aber auch extern im Zuge der Weiterverwendung der Ergebnisse. Der Prozess der hydraulischen Simulation sowie der Vergleich von Ergebnissen mit bestehenden lokalen Untersuchungen (z.B. Gefahrenzonenplänen) lassen unter anderem auch Rückschlüsse auf die Qualität der räumlichen Daten zu.

## 4 Literaturverzeichnis

- ALI, T.A, ATABAY, S., MORTULA, M. & MA, R., 2015: Semiautomatic Extraction of the Shoreline from High-Resolution Satellite Imagery and Coastal Terrain Model. *Surveying and Land Information Science* **74**(1), 15-21.
- CASAS, A., BENITO, G., THORNDYCRRAFT, V.R., & RICO, M., 2006: The topographic data source of digital terrain models as a key element in the accuracy of hydraulic flood modelling. *Earth Surface Processes and Landforms* **31**, 444-456.
- EU, 2000: Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. *Official Journal of the European Communities* **L327**, 1-72.
- EU, 2007: Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks. *Official Journal of the European Union* **L228**, 27-34.
- GÜNERALP, I., FILIPPI, A.M. & HALES, B.U., 2013: River-flow boundary delineation from digital aerial photography and ancillary images using Support Vector Machines. *GIScience & Remote Sensing* **50**(1), 1-25.
- HÖFLE, B., VETTER, M., PFEIFER, N., MANDLBURGER, G. & STÖTTER, J., 2009: Water surface mapping from airborne laser scanning using signal intensity and elevation data. *Earth Surface Processes and Landforms* **34**, 1635-1649.
- HOLLAUS, M., WAGNER, W. & KRAUS, K., 2005: Airborne laser scanning and usefulness for hydrological models. *Advances in Geosciences, European Geosciences Union* **5**, 57-63.
- MANDLBURGER, G., 2000: Verdichtung von Echolot Querprofilen unter Berücksichtigung der Flussmorphologie. *VGI – Österreichische Zeitschrift für Vermessung und Geoinformation* **88**(4), 211-214.
- MANDLBURGER, G., 2006: Topographische Modelle für Anwendungen in Hydraulik und Hydrologie. TU Wien, Dissertation.
- MANDLBURGER, G. & BRIESE, C., 2007: Using Airborne Laser Scanning for Improved Hydraulic Models. *Land, Water and Environmental Management: Integrated Systems for Sustainability, Proceedings*, 731-738.
- MARKS, K. & BATES, P., 2000: Integration of high-resolution topographic data with floodplain flow models. *Hydrological Processes* **14**, 2109-2122.
- PÉNARD, L. & MOREL, M., 2012: Automatic detection of gravel bars in a river channel from airborne LiDAR-derived DTM. *HIC 2012 – 10th International Conference on Hydroinformatics*, July, Hamburg, Germany.
- PETROSELLI, A., 2012: LIDAR Data and Hydrological Applications at the Basin Scale. *GIScience & Remote Sensing* **49**(1), 139-162.
- PFEIFER, N., MANDLBURGER, G., OTEPKA, J. & KAREL, W., 2014: OPALS – A framework for airborne laser scanning data analysis. *Computers, Environment and Urban Systems* **45**, 125-126.

# Leistungsfähiges portables Mobile Mapping im Außenraum – Genauigkeitspotential und Grenzen unterschiedlicher Georeferenzierungsansätze in der Stadt und im Wald

JONAS MEYER<sup>1</sup>, STEFAN BLASER<sup>1</sup>, STEPHAN NEBIKER<sup>1</sup>, LEA FRICKER<sup>1</sup> & DANIEL WEBER<sup>1</sup>

*Die fortschreitende Digitalisierung führt zu tiefgreifenden Veränderungen im Infrastrukturmanagement. Mithilfe bildbasierter Webdienste – vergleichbar mit Street View – mit 3D-Messfunktionalität können zahlreiche aufwändige und gefährliche Mess-, Beurteilungs- und Planungsaufgaben vom Feld ins Büro verlagert werden. Mit portablen bildbasierten Mobile Mapping Systemen (MMS), wie dem prototypischen Messrucksack „BIMAGE Backpack“ der Fachhochschule Nordwestschweiz, können bisher unzugängliche Orte im Außenraum, wie enge Gassen in der Innenstadt oder schmale Waldwege, kinematisch erfasst werden. In diesem Beitrag werden unterschiedliche Georeferenzierungsansätze diskutiert sowie deren absolutes Genauigkeits- und Zuverlässigkeitspotenzial aufgrund ausgedehnter Aufnahmen der Innenstadt von Basel sowie von Waldgebieten untersucht. Gegenüber der direkten Georeferenzierung, wurden mit nachträglicher bildbasierter Georeferenzierung Verbesserungen um einen Faktor 5-10 erreicht. Die Mediane absoluter Differenzen zu Referenzpunkten variierten zwischen 10.2 und 11.7 cm und die Präzision von Relativmessungen betrug jeweils 3 mm. Hiermit ergibt sich ein breites Anwendungsspektrum für portable bildbasierte MMS im Außenraum.*

## 1 Einleitung

Mit der fortschreitenden Urbanisierung sowie Digitalisierung und der baulichen Verdichtung gewinnen genaue Daten von Infrastrukturbauten zunehmend an Bedeutung. Die Digitalisierung führt zu tiefgreifenden Veränderungen im gesamten Planungs- und Bauprozess. Sie bietet insbesondere große Chancen im Management von Infrastrukturanlagen, indem zahlreiche aufwändige und potentiell gefährliche Mess-, Beurteilungs- und Planungsaufgaben vom Feld ins Büro verlagert werden können. Voraussetzung dafür bilden realitätsgetreue, genaue, zuverlässige und aktuelle 3D-Daten, die sich in kollaborativen digitalen Planungs- und Managementprozessen nutzen lassen. Im Straßen- und Schienenbereich werden dazu bildbasierte Webdienste mit 3D-Messfunktionalität – im Ansatz vergleichbar mit *Street View* – bereits verbreitet eingesetzt (NEBIKER 2017). Für die großflächige kinematische 3D-Infrastrukturerfassung eignen sich bildbasierte Mobile Mapping Systeme (MMS) mit mehreren Stereobasen (BURKHARD et al. 2012; NOVAK 1991; SCHWARZ et al. 1993), welche für die 360°-Abdeckung urbaner Gebiete um spezielle Stereo-Panorama-Konfigurationen (BLASER et al. 2017; MEILLAND et al. 2015) erweitert werden können. Die Vorteile von RGB-D Bildern von MMS mit fixen Stereobasen gegenüber texturierten Punktwolken von LiDAR-basierten MMS (HEUVEL et al. 2006; PUENTE et al. 2013) werden in

---

<sup>1</sup> FHNW Fachhochschule Nordwestschweiz, Institut Geomatik, Hofackerstrasse 30, CH-4132 Muttenz, Schweiz, E-Mail: [Jonas.Meyer, Stefan.Blaser, Stephan.Nebiker]@fhnw.ch, lea.delia.fricke@gmail.com, tapi@gmx.ch

NEBIKER et al. (2015) aufgezeigt und diskutiert. Im Außenraum findet für die direkte Georeferenzierung die Sensorkombination von GNSS und Inertialmesseinheit (INS) verbreiteten Einsatz. Deren Genauigkeit hängt jedoch stark von der jeweiligen GNSS-Abdeckung ab und kann auch mit hochwertigen INS-Komponenten auf mehrere Dezimeter bis Meter abfallen. Mit nachträglicher bildbasierter Georeferenzierung in Kombination mit Luftbildern sind Genauigkeiten im Dezimeterbereich erreichbar (JENDE et al. 2019). Mittels bedingter Bündelblockausgleichung und deren Lagerung auf Passpunkten lassen sich Genauigkeiten aus der direkten Georeferenzierung um einen Faktor 10 in den Zentimeterbereich verbessern (CAVEGN et al. 2018).

Frühere eigene Erfassungs- und Auswerterversuche in bewaldeten Gebieten mit portablen bildbasierten MMS (WITTMER 2017) scheiterten einerseits aufgrund eines damals zu groß gewählten Bildfolgeabstands von über 2 m, andererseits aufgrund noch fehlender alternativer robuster Georeferenzierungsansätze. Dank den Entwicklungen und Erkenntnissen der letzten Jahre, erscheint deren bildbasierte 3D-Dokumentation nun als realistisch.

Der ursprünglich für die Innenraumaufnahme konzipierte prototypische portable Messrucksack *BIMAGE Backpack* (BLASER et al. 2018) wurde für die Anwendung im Außenraum um ein mit GNSS gekoppeltes Inertialmesssystem (INS) erweitert. Somit besteht die aktuelle Systemkonfiguration aus einer Mehrkopf-Panoramakamera *Ladybug5*, zwei Multiprofillaserscannern *Velodyne VLP-16*, einer GNSS-Antenne, sowie einem INS *NovAtel SPAN CPT7*. Sämtliche Sensoren werden hardwarebasiert mit dem INS synchronisiert, welches die Zeitbasis des MMS bildet.

Die Bildposen der einzelnen Panoramakameraköpfe können nun einerseits, wie in BLASER et al. (2018) beschrieben, mit dem LiDAR-SLAM und anschließender 6 DoF-Transformation über Passpunkte, oder andererseits – dank der erweiterten Systemkonfiguration – mit direkter Georeferenzierung durch die eng gekoppelte Sensordatenfusion von GNSS und INS berechnet werden. CAVEGN et al. (2018) zeigten auf, dass sich sowohl die äußeren Orientierungen des LiDAR-SLAM, als auch jene der direkten Georeferenzierung signifikant verbessern lassen, indem die Einzelbilder der Panoramakamera mit einer leistungsfähigen Structure-from-Motion (SfM)-Pipeline prozessiert werden und somit eine Bündelblockausgleichung durchgeführt wird. Die zuvor kalibrierten relativen Orientierungen zwischen den einzelnen Panoramakameraköpfen wurden als Bedingungen in die Bündelblockausgleichung eingeführt und fixiert. Das Genauigkeitspotenzial des bildbasierten portablen MMS wurde bereits im Innenraum (BLASER et al. 2018) und unter Tage (BLASER et al. 2019a) in verschiedenen kleinräumigen Testgebieten untersucht und aufgezeigt.

In diesem Beitrag werden großflächige Untersuchungen der erweiterten Systemkonfiguration im Außenraum durchgeführt, um das Potenzial der verschiedenen zur Verfügung stehenden Georeferenzierungsansätze in unterschiedlichen Umgebungen aufzuzeigen. Neben der SLAM-basierten Georeferenzierung und der nachträglichen bildbasierten Georeferenzierung, ist dank der Systemerweiterung, zusätzlich die direkte Georeferenzierung mit GNSS und INS möglich. Ein Testgebiet befindet sich im Wald und ein weiteres Testfeld liegt in einer historischen Innenstadt mit jeweils unterschiedlichen Straßen- und Wegbreiten sowie GNSS-Signalabschattungen durch Bäume und enge, teils unbefahrte Straßenschluchten. In beiden Testgebieten bestehen tachymetrisch eingemessene Referenzpunktnetze mit 80 und 89 Fixpunkten für aussagekräftige Genauigkeitsuntersuchungen. Die großen Ausdehnungen beider Testgebiete erlauben Trajektorien mit Gesamtlängen zwischen 700 und 800 m.

Der Beitrag ist wie folgt aufgebaut: In Kapitel 2 wird das erweiterte portable bildbasierte MMS beschrieben, in Kapitel 3 werden die verschiedenen Georeferenzierungsansätze diskutiert, Kapitel 4 beinhaltet die umfangreichen Untersuchungen in der Stadt und im Wald und in Kapitel 5 werden die erzielten Resultate aufgezeigt und beurteilt.

## 2 Portables bildbasiertes Mobile Mapping System

Mit Hilfe des bildbasierten Mobile Mapping Systems (MMS) *BIMAGE Backpack* (BLASER et al. 2018) sollen nebst Innenräumen und unterirdischen Infrastrukturanlagen (BLASER et al. 2019b) auch enge innerstädtische Umgebungen und Wald kinematisch erfasst werden. Die Repräsentation dieser Umgebungen erfolgt mittels georeferenzierter 3D-Bilder, analog zu existierenden stereobildbasierten MMS im Außenraum (BURKHARD et al. 2012; BLASER et al. 2017). In NEBIKER et al. (2015) werden das Konzept und der Aufbau der daraus resultierenden georeferenzierten 3D-Bildräume beschrieben sowie deren besondere Eigenschaften diskutiert. Die für nahezu jedes Pixel verfügbare Tiefeninformation eines 3D-Bilds ermöglicht denkbar einfache, direkte 3D-Messungen im Bild per Mausklick. Dank bekannter äußerer Orientierung der Stereobilder, können aus den 3D-Messungen absolute 3D-Koordinaten berechnet werden. Der Ansatz des 3D-Bilds beschränkt sich jedoch nicht nur auf die zusätzliche Ebene mit Tiefeninformation, ein 3D-Bild lässt sich um weitere pixelbasierte Informationsebenen (z. B. Qualität der Tiefeninformation, Klassifikation, etc.) erweitern.

Das prototypische portable MMS *BIMAGE Backpack* wurde als Rucksack konzipiert. Dessen erste Sensorkonfiguration für die Innenraumaufnahme ist in BLASER et al. (2018) im Detail dokumentiert. Sämtliche Sensoren wurden an einen robusten Aluminiumrahmen montiert, welcher wiederum an ein Rucksacktraggestell befestigt wurde. Die Mehrkopf-Panoramakamera *Ladybug5* von *FLIR* (Abb. 1, Nr. 3), bestehend aus sechs einzelnen Kameraköpfen mit Fischaugenoptik und einer Auflösung von je 5 MP, bildet die Umgebungssensorik. Als Navigationssensorik dienen zwei Multiprofillaserscanner *VLP-16* von *Velodyne*, wobei je einer der beiden Laserscanner horizontal und der zweite vertikal ausgerichtet ist (Abb. 1, Nr. 2 & 5). Diese Konfiguration verleiht der fusionierten Punktwolke die notwendige geometrische Stabilität – auch in beengten Verhältnissen, wie etwa in Treppenhäusern. Des Weiteren war in der ersten Konfiguration für die Innenraumaufnahme die Inertialmesssensorik (INS) *XSens MTI-300* in das MMS integriert. Mit dem 3D-LiDAR-SLAM-Algorithmus *Google Cartographer* (HESS et al. 2016) werden die Rohdaten beider Laserscanner und jene der INS fusioniert, woraus sich die relativen 3D-Systemposen ergeben. Die Sensorsynchronisation wurde in der ersten Konfiguration mit dem Einplatinencomputer *Arduino Nano* gesteuert. Mit einem elektrischen Impuls pro Sekunde werden die Navigationssensoren (LiDAR & INS), die kontinuierlich Daten erfassen, synchronisiert. Ein zweiter leistungsstarker Mini-Computer dient sowohl der Ausführung der Erfassungssoftware sowie der Speicherung der Rohdaten für den nachträglichen Auswertungsprozess. Der flexible mechanische Aufbau, wie auch die modulare Architektur der Erfassungssoftware mit dem weit verbreiteten open-source Framework Robot Operating System (ROS) (QUIGLEY et al. 2009), ermöglichen den Austausch von einzelnen Sensorkomponenten mit geringem Aufwand.

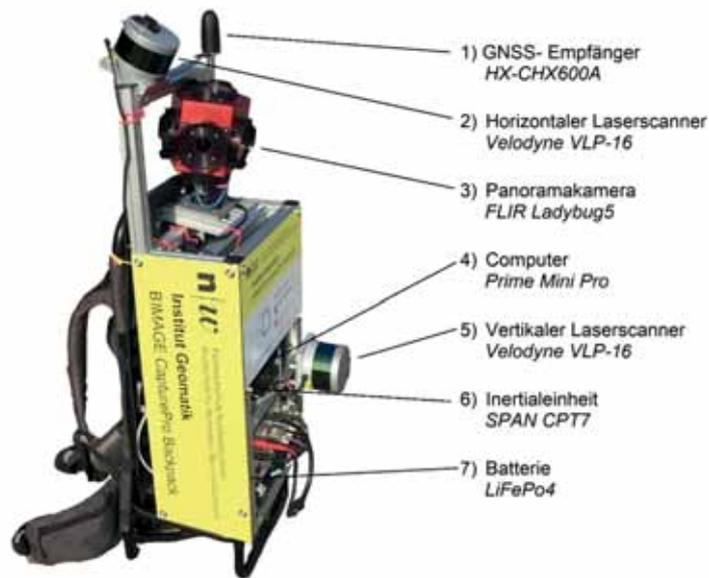


Abb. 1: Außen- und innenraumfähige Systemkonfiguration des portablen Mobile Mapping Systems *BIMAGE Backpack* mit Bezeichnung der Komponenten (FRICKER & WEBER, 2019)

Für Aufnahmen sowohl im Außen- wie auch im Innenraum, wurde einerseits das INS *XSens MIT-300* durch das mit GNSS gekoppelte INS *NovAtel SPAN CPT7* (NOVATEL INC. 2020) (Abb. 1, Nr. 6) ersetzt und dementsprechend um die GNSS-Antenne *HX-CHX600A* erweitert (Abb. 1, Nr. 1). Zusätzlich zur SLAM-basierten Systemnavigation für Innenräume, steht mit dem neuen INS, nun die mit GNSS und INS gekoppelte Systemnavigation zur Verfügung. Da das neue INS die Zeitbasis des MMS bildet und die hardwarebasierte Sensorsynchronisation unterstützt, konnte der Einplatinencomputer *Arduino Nano* weggelassen werden. Für die nachträgliche Datenauswertung wurde die in Tab. 1 aufgeführte Hardware verwendet.

Tab. 1: Verwendete Hardware für nachträgliche Datenauswertungen

	Prozessor	Grafikkarte	Arbeitsspeicher
Dell Precision Tower 7910	Intel Xeon E5-2623, 4 Core @ 3.0 GHz	nVidia GeForce GTX 1080	32 GB

### 3 Georeferenzierungsansätze

Die **direkte Georeferenzierung** (SCHWARZ, 1993) wird heutzutage verbreitet für die Erfassung von georeferenzierten Bilddaten mittels MMS eingesetzt. Aus der Sensordatenfusion von INS- und GNSS-Daten resultieren die Trajektorien, über den Bildauslösezeitpunkt werden die Bildposen in die Trajektorien interpoliert. Aufgrund der Verwendung von GNSS-Daten ist die direkte Georeferenzierung nur im Außenraum anwendbar. Sie weist ein Genauigkeitspotential von einigen Zentimeter bis mehreren Dezimeter auf (CAVEGN et al. 2016). Die erreichbare Genauigkeit sowie die Robustheit dieses Ansatzes sind stark von der Verfügbarkeit von GNSS-Signalen abhängig. Die direkte Georeferenzierung ist echtzeitfähig und geht mit dem geringsten Aufwand

für die nachträgliche Prozessierung einher. Die Prozessierungszeit ist kürzer als die Erfassungszeit und es werden aufgrund der Integration von GNSS-Daten keine Passpunkte benötigt (Tab. 2). Einen alternativen Georeferenzierungsansatz bieten SLAM-Algorithmen (DURRANT-WHYTE & BAILEY, 2006). Ursprünglich für die Navigation von Robotern in Innenräumen entwickelt, lokalisieren SLAM-Algorithmen die Position eines Sensors in der gleichzeitig erfassten Umgebung. Zur Entstehung und Ausprägungen von SLAM-Algorithmen wie beispielsweise Visual SLAM oder LiDAR SLAM geben CADENA et al. (2016) einen guten Überblick. Mit einem Genauigkeitspotential von einem Dezimeter ist die **SLAM-basierte Georeferenzierung** (Blaser et al. 2018) ähnlich genau wie die direkte Georeferenzierung. Die erreichbare Genauigkeit und Robustheit sind einerseits von der Geometrie der erfassten Umgebungsobjekte (z.B. regelmäßig oder unregelmäßig) und andererseits von der Distanz zu den Objekten abhängig. Ebenfalls eine Auswirkung auf die Genauigkeit hat die Länge der erfassten Schleifen. Im Gegensatz zur direkten Georeferenzierung ist die SLAM-basierte Georeferenzierung im Außenraum, im Innenraum sowie im Untergrund anwendbar. Wie die direkte Georeferenzierung ist die SLAM-basierte Georeferenzierung echtzeitfähig. Bei der nachträglichen Prozessierung ist mit einer leicht höheren Prozessierungszeit als der Erfassungszeit zu rechnen. Zusätzlich sind Passpunkte für die Transformation in den globalen Bezugsrahmen zu bestimmen (Tab. 2).

Tab. 2: Gegenüberstellung der zu untersuchenden Georeferenzierungsansätze

	Direkt	SLAM-basiert (LiDAR)	Bildbasiert
Genauigkeitspotential (3D-Position)	Dezimeter bis Zentimeter	Dezimeter	Zentimeter
Anwendungsspektrum	Außenraum	Außenraum Innenraum Untergrund	Außenraum Innenraum Untergrund
Robustheit	Abhängig von GNSS-Verfügbarkeit	Abhängig von geometrischer Umgebungsbeschaffenheit	Abhängig von radiometrischer Umgebungsbeschaffenheit, Qualität der Näherungsposen
Aufwand	Echtzeitfähig, nachträglicher Prozessierungsaufwand < Erfassungszeit, Passpunkte nicht notwendig	Echtzeitfähig, nachträglicher Prozessierungsaufwand > Erfassungszeit, Passpunkte notwendig für Transformation in globalen Bezugsrahmen	Nicht echtzeitfähig, nachträglicher Prozessierungsaufwand >> Erfassungszeit, Passpunkte notwendig für Transformation in globalen Bezugsrahmen

Für die **bildbasierte Georeferenzierung** werden die Bilder mit zugehörigen Näherungsposen sowie kalibrierten Hebelarmen und Fehlansichtungen zwischen den Kameraköpfen in eine leistungsfähige Structure-from-Motion (SfM)-Pipeline eingeführt. Durch Anwendung einer Bündelblockausgleichung, werden die Bilder über zuvor automatisch generierte Verknüpfungspunkte gegenseitig orientiert. Unter Einbezug von Passpunktmessungen können die Bildposen in den absoluten Bezugsrahmen überführt werden. Mit der bildbasierten Georeferenzierung sind Genauigkeiten im Zentimeterbereich erreichbar (CAVEGN et al. 2018). Wobei die Genauigkeit und die Robustheit dieses Ansatzes von den Lichtverhältnissen, der Beschaffenheit der kartierten

Umgebung (Strukturierung, Reflexionen, Schatten etc.) sowie der Qualität der Näherungsposen abhängig ist. Die bildbasierte Georeferenzierung findet im Außenraum, im Innenraum und im Untergrund Anwendung. Sie benötigt ein Mehrfaches der Erfassungszeit für die nachträgliche Prozessierung sowie die zusätzliche Bestimmung von Passpunkten für die Transformation in den globalen Bezugsrahmen (Tab. 2). Somit ist die bildbasierte Georeferenzierung der rechenaufwändigste Ansatz. Untersuchungen haben gezeigt, dass die Näherungsposen aus der direkten (CAVEGN et al. 2018) und aus der SLAM-basierten Georeferenzierung (BLASER et al. 2018) mithilfe der bildbasierten Georeferenzierung signifikant verbessert werden können.

## 4 Untersuchungen

Für die Untersuchungen wurde jeweils ein ausgedehntes repräsentatives Untersuchungsgebiet in einem Waldgebiet im Raum Basel sowie in der historischen Innenstadt von Basel ausgewählt. Im Wald wie auch in der Innenstadt ist die Befahrbarkeit oftmals nicht gegeben. Zudem ist der GNSS-Empfang, aufgrund von Signalabschattungen durch Gebäude oder Bäume, in beiden Szenarien oft sehr stark eingeschränkt. Dies hat einen bedeutenden Einfluss auf die direkte Georeferenzierung, welche den momentan üblichen Ansatz für MMS im Außenraum darstellt. Die folgenden Untersuchungen mit dem portablen MMS *BIMAGE Backpack* sollen die Grenzen und das Genauigkeitspotential unterschiedlicher Georeferenzierungsansätze unter realistischen Bedingungen im Außenraum aufzeigen. Wir untersuchten, inwiefern die SLAM-basierte Georeferenzierung eine Alternative zur direkten Georeferenzierung darstellt. Des Weiteren untersuchten wir die Genauigkeitssteigerung der direkten Georeferenzierung durch die nachträgliche Anwendung einer bildbasierten Georeferenzierung. Das Genauigkeitspotential der bildbasierten Georeferenzierung mit Näherungswerten aus der SLAM-basierten Georeferenzierung wurde bereits im Innenraum (BLASER et al. 2018) und unter Tage (BLASER et al. 2019b) in verschiedenen kleinräumigen Testgebieten nachgewiesen.

### 4.1 Innenstadt

Das Testgebiet in der Basler Innenstadt verfügt über typische Ausprägungen von Plätzen, Straßen und Gassen wie sie in historisch gewachsenen Städten zu finden sind. Die schleifenförmige Trajektorie, mit einer Länge von 800 m (Abb. 2, Mitte & Tab. 3), beginnt und endet auf dem Barfüsserplatz (Abb. 2, Bild 1). Der ca. 70 x 50 m große Platz, weist eine gute GNSS-Abdeckung auf und eignet sich daher optimal für die Initialisierung des portablen MMS. Das Testgebiet Innenstadt beinhaltet unterschiedlich breite Straßen und Gassen. Auf Abb. 2, Bild 2 ist eine enge Gasse mit alten Stadtgebäuden auf der einen Seite und einer hohen Mauer mit wenig Struktur auf der anderen Seite zu erkennen. Die Gasse weist eine Steigung von rund 16 % auf. Weiter verfügt das Testgebiet über Fußgängerzonen gesäumt von modernen Gebäuden mit Gewerbenutzung im Erdgeschoss (Abb. 2, Bild 3). Im Bereich von Abb. 2, Bild 4 wird eine breite Straße mit historischen und modernen Stadtgebäuden beidseitig erfasst. Diese Straße hat ein hohes Tram- und Fahrradverkehrsaufkommen.

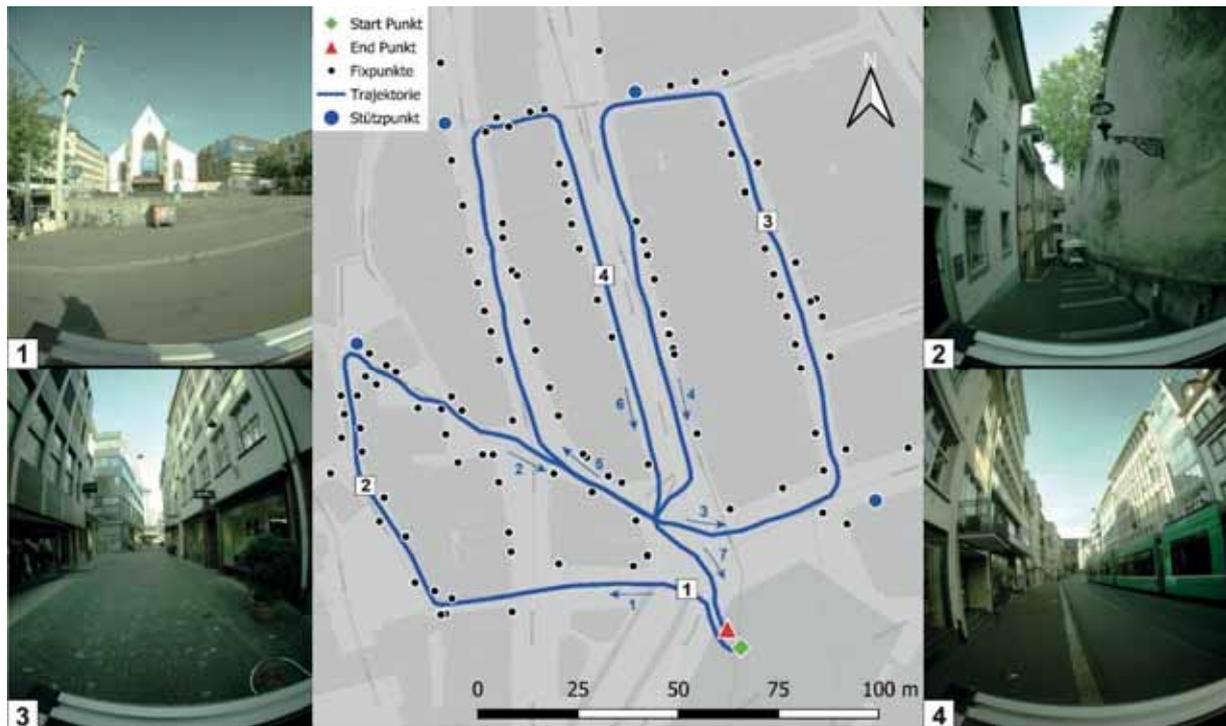


Abb. 2: Karte des Testgebiet Innenstadt mit Messbildern zur Veranschaulichung der Bedingungen vor Ort

Vor der Erfassung wurde im Untersuchungsgebiet (Abb. 2, Mitte) ein Fixpunktnetz als Referenz für die Genauigkeitsuntersuchungen angelegt (FRICKER & WEBER, 2019). Die 80 größtenteils natürlichen Fixpunkte (klar definierte Punkte an Fassaden, Verkehrsschildern etc.) und die wenigen mit photogrammetrischen Zielmarken versicherten Zusatzpunkte wurden tachymetrisch eingemessen und an das sehr genaue Fixpunktnetz der amtlichen Vermessung Basel-Stadt angeschlossen, wobei eine Fixpunktkoordinatengenauigkeit von ca. 5 mm erreicht wurde (Tab. 3).

Zu Beginn und am Ende der Kampagne wurde das portable MMS jeweils auf dem Barfüsserplatz initialisiert. Dies ermöglicht die nachträgliche beidseitige Auswertung der Trajektorie des Inertialmesssystems. Bei der Initialisierung des Inertialmesssystems wird dessen Ausrichtung gegenüber dem Weltkoordinatensystem basierend auf GNSS-Messungen berechnet (INS-Lösung) (NOVATEL INC. 2020). Für eine ausreichend genaue INS-Lösung sind einige Runden mit schnellem Gehen in einem Gebiet mit guter GNSS-Abdeckung erforderlich, wobei Abrupte Dreh- und Schwenkbewegungen zu vermeiden sind.

Nach erfolgreicher Initialisierung wurde die Trajektorie mit einer Gesamtlänge von 800 m in einem Guss abgeschrieben und wiederum auf dem Barfüsserplatz abgeschlossen (Abb. 2). Die Speicherung der sechs einzelnen Panoramarohbilder nimmt jeweils etwa eine Sekunde in Anspruch. Um den angestrebten Bildfolgeabstand von einem Meter zu erzielen, musste die Aufnahme mit einer langsamen Gehgeschwindigkeit von maximal 1 m/s erfolgen.

Tab. 3: Fakten zum Testgebiet Innenstadt

Bilddaufnahmestandorte:	721	Anzahl Fixpunkte:	80
Einzelbilder:	4326	Genauigkeit Fixpunkte:	5 mm
Länge Trajektorie	800 m	Anzahl Schleifen:	3
Datenmenge Punktwolke	14 GB	Erfassungszeit:	25 min

## 4.2 Wald

Im Testgebiet Wald wurde die Trajektorie so gelegt, dass sowohl verschiedenartige Wege wie auch unterschiedlich dichter Bewuchs erfasst werden. Ausgangs- und Endpunkt der Trajektorie befinden sich außerhalb des Waldes auf einer Autobahnbrücke (Abb. 3, Bild 1). Dies ermöglichte aufgrund der dortigen geringen GNSS-Signalabschattung eine gute Systeminitialisierung. Die definierte Trajektorie führt über breite Forstwege (Abb. 3, Bilder 2, 4 & 5) sowie schmale Trampelpfade (Abb. 3, Bilder 3 & 5), zusätzlich variiert die Dichte des Waldes stark. Abb. 3, Bilder 3 & 5 zeigen dichten Wald mit vielen Büschen während bei Abb. 3, Bild 2 lichter, vereinzelt mit Sträuchern durchmischter Wald vorherrscht.

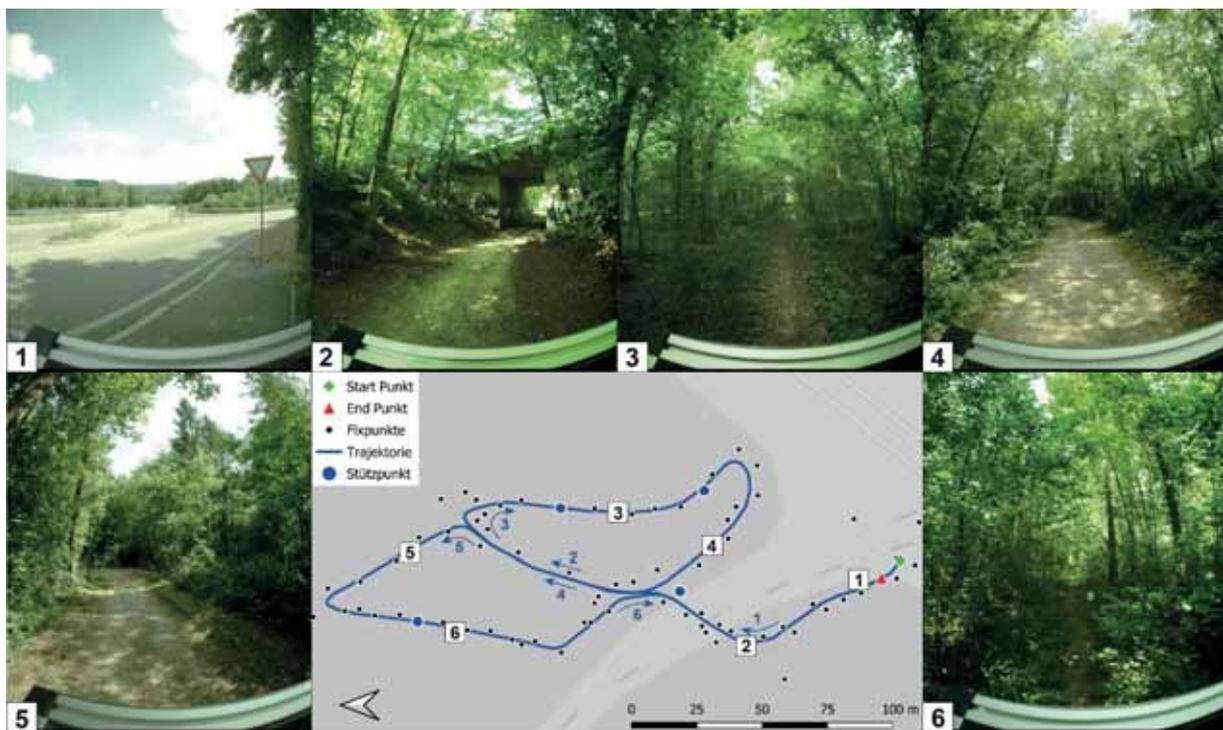


Abb. 3: Karte des Testgebiet Wald mit Messbildern zur Veranschaulichung der Bedingungen vor Ort

In FRICKER & WEBER (2019) ist die tachymetrische Bestimmung und Auswertung des Fixpunktnetzes dokumentiert. Sämtliche Fixpunkte wurden mit A4-grossen laminierten photogrammetrischen Zielmarken versichert, welche an Baumstämme sowie an eingeschlagene Pfähle befestigt wurden. Das Fixpunktnetz umfasst 89 Fixpunkte, welche tachymetrisch mit einem geschlossenen Polygonzug eingemessen wurden. Die Ausgleichung erfolgte in einem Guss nach der Methode der

kleinsten Quadrate. Es resultierte eine Genauigkeit von 5 mm in Lage und Höhe (Tab. 4). Der Ablauf der Datenerfassung mit dem portablen MMS erfolgte analog dem Testgebiet in der Stadt.

Tab. 4: Fakten zum Testgebiet Wald

Bildaufnahmestandorte:	843	Anzahl Fixpunkte:	89
Anzahl Einzelbilder:	5052	Genauigkeit Fixpunkte:	5 mm
Länge Trajektorie	740 m	Anzahl Schleifen:	2
Datenmenge Punktwolke	16 GB	Erfassungszeit:	24 min

### 4.3 Untersuchungssystematik

Mit den nachfolgenden Untersuchungen sollen primär die Genauigkeitspotenziale für absolute 3D-Koordinatenmessungen sowie für lokale relative 3D-Messungen (z. B. 3D-Distanz oder 3D-Flächenmessung) der verschiedenen Georeferenzierungsmethoden ermittelt und einander gegenübergestellt werden.

#### 4.3.1 3D-Koordinatenbestimmung

Das Genauigkeitspotenzial für absolute 3D-Koordinatenmessungen wurde, analog BLASER et al. (2018), mittels Differenzen zwischen Koordinatenmessungen aus den erfassten Daten und den Referenzkoordinaten der Fixpunkte ermittelt. Die Koordinatenmessungen in den Daten des portablen MMS erfolgten mit räumlichen Vorwärtseinschnitten aus Bildmessungen mit jeweils vier verschiedenen Kamerastandorten. Dabei wurden für die Untersuchung sämtlicher Georeferenzierungsmethoden stets dieselben Bildmessungen verwendet und jeweils die entsprechenden Bildposen ausgetauscht.

Wie bei vorhergehenden Untersuchungen (BLASER et al. 2018; BLASER et al. 2019a) wurde das Genauigkeitspotenzial für relative 3D-Messungen mit den zur Georeferenzierungsmethode und Untersuchungsgebiet zugehörigen Standardabweichungen der Vorwärtseinschnitte beurteilt.

#### 4.3.2 Lagerung

Die lokalen Bildposen der SLAM-basierten sowie der bildbasierten Georeferenzierung wurden pro Untersuchungsgebiet mit jeweils vier identischen Stützpunkten (Abb. 2 & Abb. 3) mit einer 6 DoF-Koordinatentransformation ins Referenzkoordinatensystem überführt. Die Bildposen der direkten Georeferenzierung wurden nicht transformiert, da sie bereits im Referenzkoordinatensystem vorlagen. Allfällig systematisch vorliegende Translationen und Rotationen der direkten Georeferenzierung wurden bei der Untersuchung der direkten Georeferenzierung somit nicht berücksichtigt.

#### 4.3.3 Auswerteprozesse

Die Datenauswerteprozesse für die SLAM-basierte sowie die bildbasierte Georeferenzierung sind bereits in BLASER et al. (2018) im Detail beschrieben (Abb. 4). Für die Auswertung der direkten Georeferenzierung (Abb. 4, Mitte) mit *NovAtel Inertial Explorer* werden die GNSS-Rohdaten des portablen MMS, die GNSS-Rohdaten einer Referenzstation sowie die Rohdaten des INS benötigt. Die Trajektorie wird mit eng gekoppelter Sensordatenfusion (engl. *tightly-coupled*) in einem Kalman-Filter beidseitig prozessiert.

Bei der SLAM-basierten Georeferenzierung mit *Google Cartographer* wird in einem ersten Schritt die Trajektorie vom *Cartographer Status* extrahiert. Anschließend werden die Bildposen mit den Zeitstempeln der Kameraauslösung linear interpoliert (Abb. 4, links). Durch entsprechende 6 DoF-Transformation und Anbringen der Hebelarme und Fehlrichtungen, lassen sich die äußeren Orientierungen Bilder der einzelnen Kameraköpfe im Referenzkoordinatensystem berechnen. Für die bildbasierte Georeferenzierung mit der SfM-Pipeline *Agisoft Metashape* werden die Kameraposen der direkten Georeferenzierung als Initialwerte verwendet. Die verzeichnungsfreien Einzelbilder im äquidistanten Kameramodell (ABRAHAM & FÖRSTNER 2005) werden als Mehrkameranystem mit festgesetzten relativen Orientierungen zwischen den Kameraköpfen in die SfM-Pipeline mit Bündelblockausgleichung eingeführt.

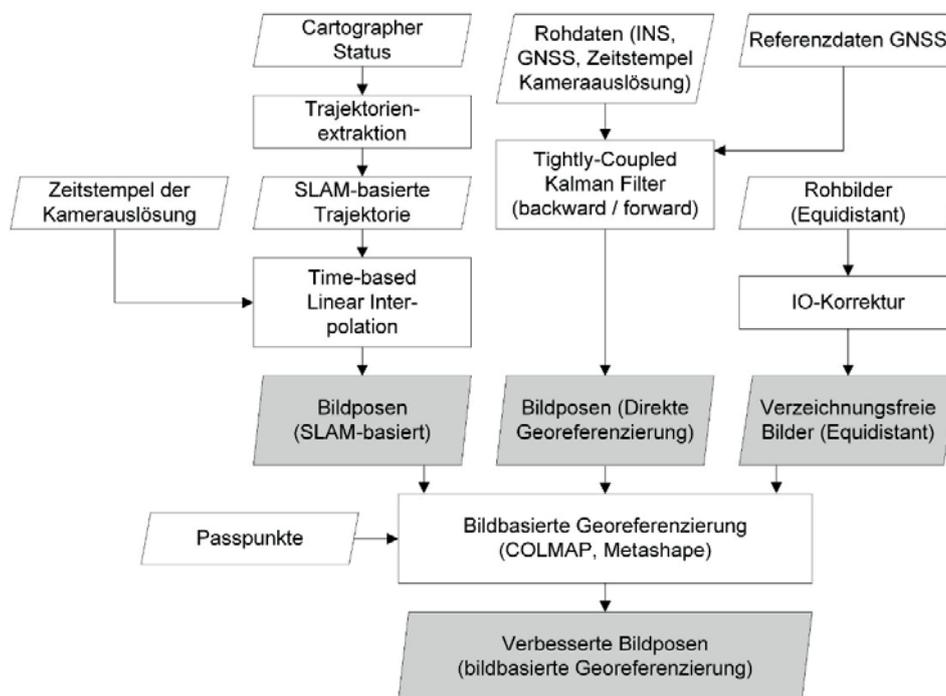


Abb. 4: Flussdiagramm des nachträglichen Datenauswertungsprozesses

## 5 Resultate

Als erstes wurden die empirischen Standardabweichungen der Vorwärtseinschnitte berechnet. Sie repräsentieren die Präzision einer 3D-Koordinatenbestimmung. Die Präzision ist vorwiegend von der relativen Orientierungsgenauigkeit aufeinanderfolgender bzw. benachbarter Bilder abhängig. Aufgrund der Präzision lässt sich auf die erreichbare Genauigkeit einer für die praktische Nutzung typische Relativmessung, beispielsweise einer 3D-Distanz schließen.

Die GNSS-Verfügbarkeit, die Geometrie der erfassten Objekte, die Lichtverhältnisse und die Beschaffenheit der kartierten Umgebung (Tab. 2), führten bei sämtlichen Georeferenzierungsansätzen in beiden Testgebieten zu Ausreißern. Daher wurden der Median sowie der MAD (engl.: *median absolute deviation*) als robuste Qualitätsmasse verwendet.

Sowohl in der Stadt als auch im Wald weist die direkte Georeferenzierung einen um Faktor drei verbesserten Medianwert der Präzision gegenüber der SLAM-basierten Georeferenzierung auf (Abb. 5 & Tab. 5). Die Präzision der direkten Georeferenzierung lässt sich mithilfe der bildbasierten Georeferenzierung von 0.034 m auf 0.003 m in der Stadt und von 0.023 m auf 0.003 m im Wald verbessern, was in etwa einem Faktor zehn entspricht (Tab. 5).

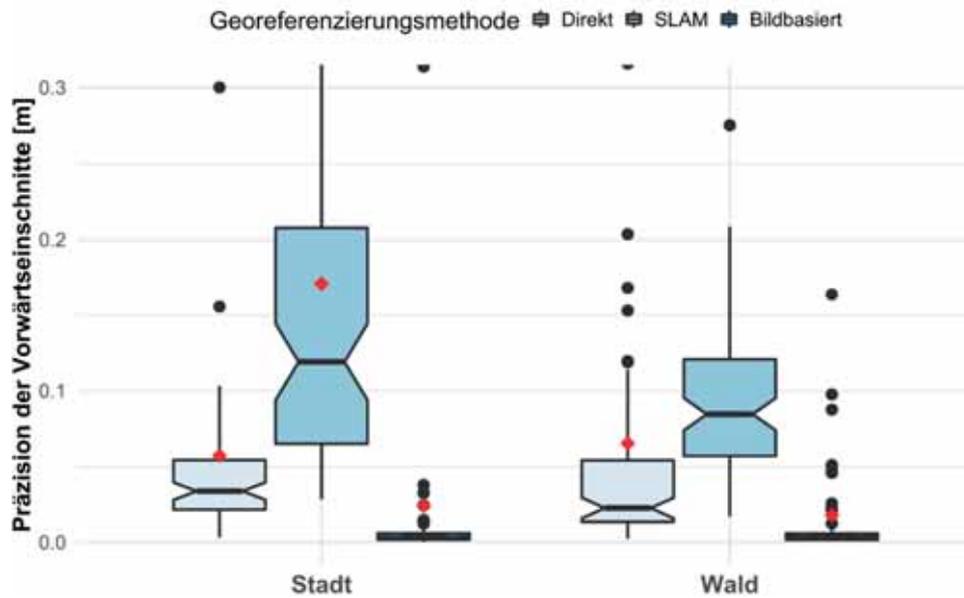


Abb. 5: Boxplot der Präzision der Vorwärtseinschnitte, aufgeteilt nach Testgebiete Stadt und Wald und geordnet nach Georeferenzierungsmethoden Direkt, SLAM, bildbasiert. Der rote Punkt definiert den Mittelwert, schwarze Punkte repräsentieren Ausreißer.

Tab. 5: Zusammenstellung der statistischen Kennzahlen zur Präzision der Vorwärtseinschnitte. Die direkte Georeferenzierung wurde nicht transformiert, die SLAM-basierte sowie die bildbasierte Georeferenzierung wurden auf je vier Stützpunkten gelagert.

	Stadt			Wald		
	Direkt	SLAM	bildbasiert	Direkt	SLAM	bildbasiert
n	80	80	80	89	89	89
min(x) [m]	0.003	0.028	0.000	0.002	0.017	0.001
Max(x) [m]	1.223	1.146	1.202	0.712	101.431	0.747
Mittelwert $\bar{x}$ [m]	0.057	0.171	0.024	0.065	1.278	0.018
Stdabw( $\bar{x}$ ) [m]	0.138	0.191	0.138	0.130	10.739	0.081
Median $\tilde{x}$ [m]	0.034	0.119	0.003	0.023	0.084	0.003
MAD [m]	0.015	0.056	0.002	0.012	0.031	0.002

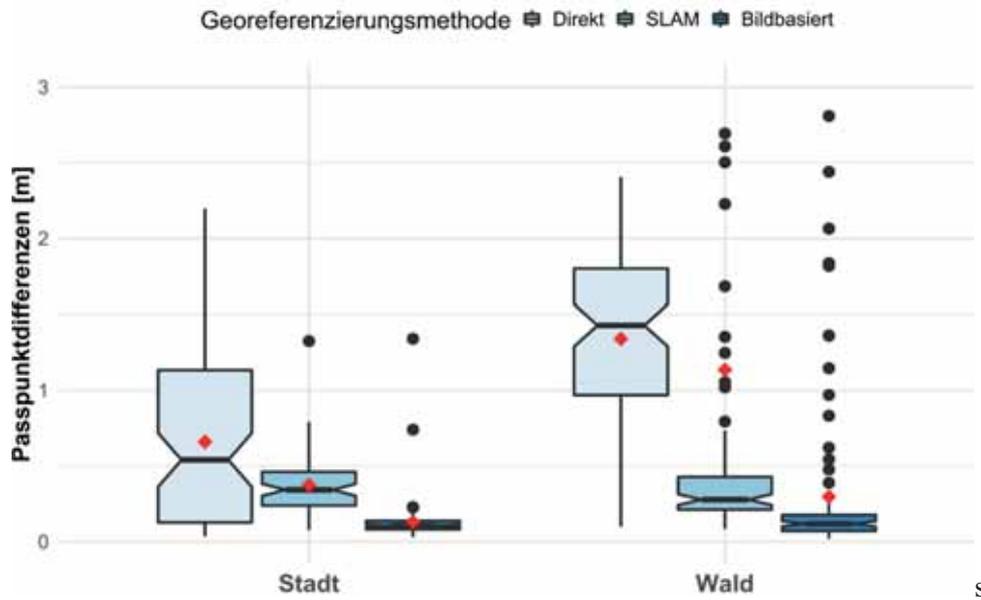


Abb. 6: Boxplot der absoluten Passpunktdifferenzen, aufgeteilt nach Testgebiete Stadt und Wald und geordnet nach Georeferenzierungsmethoden Direkt, SLAM, bildbasiert. Der rote Punkt definiert den Mittelwert, schwarze Punkte repräsentieren Ausreißer.

Tab. 6: Zusammenstellung der statistischen Kennzahlen der absoluten Genauigkeitsuntersuchungen. Die direkte Georeferenzierung wurde nicht transformiert, die SLAM-basierte sowie die bildbasierte Georeferenzierung wurden auf je vier Stützpunkten gelagert.

	Stadt			Wald		
	Direkt	SLAM	bildbasiert	Direkt	SLAM	bildbasiert
n	80	80	80	89	89	89
min(x) [m]	0.034	0.076	0.029	0.098	0.083	0.018
Max(x) [m]	2.197	1.326	1.341	2.406	56.386	2.809
Mittelwert $\bar{x}$ [m]	0.659	0.373	0.129	1.341	1.136	0.296
Stdabw( $\bar{x}$ ) [m]	0.557	0.202	0.159	0.609	5.965	0.540
Median $\tilde{x}$ [m]	0.540	0.342	0.102	1.429	0.277	0.117
MAD [m]	0.439	0.114	0.027	0.406	0.096	0.056

Zur Beurteilung der absoluten Genauigkeiten der verschiedenen Georeferenzierungsansätze wurden Passpunktdifferenzen berechnet. Im Gegensatz zur SLAM- und bildbasierten Georeferenzierung wurde die direkte Georeferenzierung nicht über die vier Stützpunkte (Abb. 2 & Abb. 3) transformiert.

Die direkte Georeferenzierung weist sowohl in der Stadt als auch im Wald die geringste absolute Genauigkeit auf, wobei der Medianwert des Testgebiets Wald um beinahe Faktor drei schlechter ist als jener in der Stadt (Abb. 6 & Tab. 6). Dies könnte die größere GNSS-Signalabschattung im Wald als Ursache haben. Des Weiteren liegen die Mediane der absoluten Genauigkeiten von der SLAM-basierten und bildbasierten Georeferenzierung beider Testgebiete in denselben Genauigkeitsbereichen. Vergleiche der Standardabweichungen und der maximalen Werte (Tab. 6) zeigen, dass die SLAM-basierte und die direkte Georeferenzierung im Wald jeweils deutlich mehr Ausreißer aufweisen als in der Stadt. Zu erklären ist dies einerseits durch den verwendeten, primär für Innenräume konzipierten, SLAM-Algorithmus *Google Cartographer* (HESS et al. 2016).

Aufgrund zahlreicher orthogonaler Kanten und Ebenen in der Innenstadt, liefert der SLAM-Algorithmus in der Stadt robustere Systemposen als im Wald. Zudem ist in der Stadt eine größere Anzahl eindeutig identifizierbarer und stabiler Verknüpfungspunkte für die bildbasierte Georeferenzierung verfügbar, was zu einer geringeren Anzahl an Ausreißern führt.

In der Stadt führte die SLAM-basierte Georeferenzierung gegenüber der direkten Georeferenzierung zu einer Verbesserung des Medians um einen Faktor 1.5 von 0.54 m auf 0.34 m, während dem die Verbesserung im Wald von 1.43 m auf 0.28 m dem Faktor 5 entspricht (Tab. 6). Mit der bildbasierten Georeferenzierung konnten die Passpunktdifferenzen im Vergleich mit der direkten Georeferenzierung in der Stadt von 0.54 m auf 0.10 m um Faktor 5 und im Wald von 1.43 m auf 0.12 m um Faktor 12 verringert werden (Tab. 6). Die Verbesserung der SLAM-basierten zur bildbasierten Georeferenzierung beträgt in beiden Testgebieten Faktor drei. Diese Verbesserung entspricht den Resultaten von BLASER et al. (2019b) welche im Untergrund erzielt wurden.

Im Vergleich mit TANG et al. (2016), welche die SLAM-basierte Erfassung von Waldbeständen untersuchten, liegen unsere Resultate ebenfalls in einem ähnlichen Genauigkeitsbereich. TANG et al. (2016) erreichen mit ihrem mobilen fahrzeugbasierten Laserscanning-System eine 2D-Positionsgenauigkeit von 0.32 m bei einer 300 m langen Trajektorie. Ihre Resultate sind in deren Größenordnung mit unseren Ergebnissen vergleichbar. Jedoch unterscheiden sich die Arbeiten im verwendeten SLAM-Algorithmus, in den unterschiedlichen Ausprägungen des Waldes sowie in den unterschiedlichen Qualitätsmassen der Genauigkeitsuntersuchungen.

## 6 Fazit & Ausblick

Im Rahmen dieses Beitrags wurde das prototypische portable bildbasiert MMS «BIMAGE Backpack» um ein mit GNSS kombiniertes INS erweitert, sodass neben der SLAM-basierten und bildbasierten Georeferenzierung auch die direkte Georeferenzierung im Außenraum ermöglicht wird. Die verfügbaren Georeferenzierungsansätze wurden diskutiert und anhand derer Einsatzgebiet, Genauigkeitspotenzial und Aufwand einander gegenübergestellt. Das Einsatzgebiet der direkten Georeferenzierung beschränkt sich auf den Außenraum und deren Genauigkeit hängt stark von der GNSS-Verfügbarkeit ab und kann zwischen dem Zentimeter- und dem Meterbereich variieren. Die direkte Georeferenzierung geht mit der geringsten nachträglichen Rechenzeit einher. Für die SLAM-basierte Georeferenzierung ist der Rechenaufwand gegenüber der Erfassungszeit leicht erhöht. Der rechenaufwändigste Ansatz ist die bildbasierte Georeferenzierung, hierbei übersteigt die nachträgliche Rechenzeit die Erfassungszeit um ein Mehrfaches. Die SLAM-basierte sowie die bildbasierte Georeferenzierung können sowohl im Innen- als auch im Außenraum eingesetzt werden. Das Genauigkeitspotenzial der SLAM-basierten Georeferenzierung ist im Wesentlichen von der geometrischen und jenes der bildbasierten Georeferenzierung von der radiometrischen Umgebungsbeschaffenheit abhängig. Die lokalen Szenen der SLAM- und bildbasierten Georeferenzierung lassen sich über Passpunkte mit einer 6 DoF-Transformation in den globalen Bezugsrahmen überführen.

Die verschiedenen Georeferenzierungsansätze wurden anhand großflächiger Aufnahmen im urbanen Raum mit engen, teilweise nicht befahrbaren Gassen und tiefen Straßenschluchten, sowie im Wald mit unterschiedlich dichtem Bewuchs empirisch evaluiert. Die Trajektorienlängen betragen zwischen 700 und 800 m. Für die Beurteilung des absoluten Genauigkeitspotenzials wurden die

Passpunktkoordinaten mit Vorwärtseinschnitten basierend auf Bildmessungen von jeweils vier unabhängigen Standpunkten berechnet. Mit den Differenzen zu den Referenzkoordinaten wurde das absolute Koordinatengenauigkeitspotenzial untersucht. Die Standardabweichungen der Vorwärtseinschnitte dienen als gutes Maß für die Genauigkeit von lokalen Relativmessungen, beispielsweise 3D-Distanzen.

Der Median der Differenzen zwischen direkter Georeferenzierung und Referenzkoordinaten betrug in der Stadt 0.54 m und im Wald aufgrund der größeren GNSS-Signalabschattung 1.43 m. Mit der SLAM-basierter Georeferenzierung wurde eine Verbesserung der Medianabweichungen um Faktor 2-5 nachgewiesen. Die entsprechenden Medianabweichungen der bildbasierten Georeferenzierung betrugen 0.10 und 0.12 m, was einer Verbesserung um Faktor 5-10 gegenüber der direkten Georeferenzierung entspricht. Somit konnten die Ergebnisse der bildbasierten Georeferenzierung im urbanen Raum von CAVEGN et al. (2018) auch mit einem portablen System bestätigt werden. Die Ergebnisse im Wald mit SLAM-basierter Georeferenzierung sind mit den Resultaten von TANG et al. (2015) vergleichbar. Im Wald wurden jedoch sowohl bei der bildbasierten als auch bei der SLAM-basierten Georeferenzierung eine große Anzahl von Ausreißern festgestellt. Dies ist vermutlich auf die herausfordernde Umgebung zurückzuführen. Die Mediane der Standardabweichungen der Vorwärtseinschnitte mit der bildbasierten Georeferenzierung betrugen sowohl in der Stadt als auch im Wald je 3 mm. Im Untergrund (BLASER et al. 2019b) sowie im Innenraum (BLASER et al. 2018) wurden vergleichbare Resultate erzielt. Die Ergebnisse zeigen das große Potenzial von alternativen Georeferenzierungsstrategien auf, wodurch die kinematische bildbasierte Erfassung von bisher unzugänglichen Gebieten ermöglicht wird.

In weiteren Untersuchungen wird die Kombination der unterschiedlichen Georeferenzierungsstrategien angestrebt, um neben der Genauigkeit auch die Robustheit der Georeferenzierung unter variierenden Umgebungsbedingungen weiter zu verbessern. Ein weiterer Schritt umfasst die Kalibrierung des Gesamtsensorsystems, wobei die analytische Schätzung der Hebelarme zwischen den Laserscannern und der Panoramakamera noch ausstehend sind. Ein aktueller Entwicklungsschritt umfasst die robuste Rekonstruktion der Tiefeninformation in den erfassten Bildern, damit künftig Koordinatenmessungen im 3D-Bild per Mausklick vorgenommen werden können, was die Handhabung von 3D-Bilddiensten für das Infrastrukturmanagement wesentlich vereinfacht. Weitere Forschungsinitiativen beinhalten die visuelle Lokalisierung auf der Basis von 3D-Bilddiensten sowie die bildbasierte Georeferenzierung von Bildern unterschiedlicher Zeitstände.

## 7 Dank

Das Forschungsprojekt *cloudIO* (32411.1 IP-ICT) wurde von der Innosuisse, der Schweizerischen Agentur für Innovationsförderung, und von der Industriepartnerin iNovitas AG finanziell unterstützt.

## 8 Literaturverzeichnis

- ABRAHAM, S. & FÖRSTNER, W., 2005: Fish-eye-stereo calibration and epipolar rectification. *ISPRS Journal of Photogrammetry and Remote Sensing* **59**(5), 278-288.
- BLASER, S., CAVEGN, S. & NEBIKER, S., 2018: Development of a Portable High Performance Mobile Mapping System using the Robot Operating System. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **4**(1), 13-20.
- BLASER, S., NEBIKER, S. & CAVEGN, S., 2017: System Design, Calibration and Performance Analysis of a Novel 360° Stereo Panoramic Mobile Mapping System. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **4**(1/W1), 207-213.
- BLASER, S., NEBIKER, S. & WISLER, D., 2019a: Portables bildbasiertes Mobile Mapping System im Einsatz unter Tage – Systemaufbau, Georeferenzierungs- und Genauigkeitsuntersuchungen. *Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V., Band 28*, T. Kersten (Hrsg.), 39. Wissenschaftlich-Technische Jahrestagung der DGPF, 20.-22. März 2019 in Wien, 458-471.
- BLASER, S., NEBIKER, S. & WISLER, D., 2019b: Portable Image-based High Performance Mobile Mapping System in Underground Environments – System Configuration and Performance Evaluation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **4**(2/W5), 255-262.
- BURKHARD, J., CAVEGN, S., BARMETTLER, A. & NEBIKER, S., 2012: Stereovision Mobile Mapping: System Design and Performance Evaluation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **39**, Part B5, 453-458.
- CADENA, C., CARLONE, L., CARRILLO, H., LATIF, Y., SCARAMUZZA, D., NEIRA, J., REID, I.D. & LEONARD, J.J., 2016: Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. *IEEE Transactions on Robotics* **32**(6), 1309-1332.
- CAVEGN, S., NEBIKER, S. & HAALA, N., 2016: A Systematic Comparison of direct Georeferencing in Challenging Urban Areas. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **41**(B1), 529-536, doi:10.5194/isprs-archives-XLI-B1-529-2016.
- CAVEGN, S., BLASER, S., NEBIKER, S. & HAALA, N., 2018: Robust and Accurate Image-Based Georeferencing Exploiting Relative Orientation Constraints. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* **4**(2), 57-64.
- DURRANT-WHYTE, H. & BAILEY, T., 2006: Simultaneous localization and mapping: part I. *IEEE Robotics & Automation Magazine* **13**(2), 99-110.
- FRICKER, L. & WEBER, D., 2019: Portables Mobile Mapping im Aussenraum. *Bachelorthesis, 2019*(2), Fachhochschule Nordwestschweiz, Institut Geomatik, (unveröffentlicht).
- HESS, W., KOHLER, D., RAPP, H. & ANDOR, D., 2016: Real-Time Loop Closure in 2D LIDAR SLAM. *IEEE International Conference on Robotics and Automation (ICRA)*, Stockholm, Sweden, 1271-1278.
- HEUVEL, F.A. VAN DEN, VERWAAL, R. & BEERS, B., 2006: Calibration of Fisheye Camera Systems and the Reduction of Chromatic Aberration. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Dresden, Germany* **36**(5).

- MEILLAND, M., COMPORT, A.I. & RIVES, P., 2015: Dense Omnidirectional RGB-D Mapping of Large-scale Outdoor Environments for Real-time Localization and Autonomous Navigation. *Journal of Field Robotics* **32**(4), 474-503.
- JENDE, P.L.H., 2019: Position estimation of mobile mapping imaging sensors using aerial imagery. Dissertation, Enschede: University of Twente, Faculty of Geo-Information Science and Earth Observation (ITC), doi: 10.3990/1.9789036548847.
- NEBIKER, S., 2017: 3D Imagery for Infrastructure Management – Mobile Mapping meets the Cloud. The 56th Photogrammetric Week. Stuttgart: University of Stuttgart, Institute for Photogrammetry.
- NEBIKER, S., CAVEGN, S. & LOESCH, B., 2015: Cloud-Based Geospatial 3D Image Spaces—A Powerful Urban Model for the Smart City. *ISPRS International Journal of Geo-Information* **4**(4), 2267-2291.
- NOVAK, N., 1991: The Ohio State University Mapping System: the stereo vision system component. Proceedings of the 47th Annual Meeting, Institute of Navigation, Williamsburg, VA, June 10–12 (1991), 121-124.
- NOVATEL INC., 2020: SPAN CPT7. <https://www.novatel.com/assets/Documents/Papers/SPAN-CPT7-PS.pdf>, letzter Zugriff 15.01.2020.
- PUENTE, I., GONZÁLEZ-JORGE, H., MARTÍNEZ-SÁNCHEZ, J. & ARIAS, P., 2013: Review of mobile mapping and surveying technologies. *Measurement* **46**(7), 2127-2145.
- QUIGLEY, M., CONLEY, K., GERKEY, B., FAUST, J., FOOTE, T., LEIBS, J., BERGER, E., WHEELER, R. & NG, A., 2009: ROS: an open-source Robot Operating System. ICRA workshop on open source software **3**(3.2).
- TANG, J., CHEN, Y., KUKKO, A., KAARTINEN, H., JAAKOLA, A., KHORAMSHAHI, E., HAKALA, T., HYYPPÄ, J., HOLOPAINEN, M. & HYYPPÄ, H., 2015: SLAM-Aided Stem Mapping for Forest Inventory with Small-Footprint Mobile LiDAR. *Forests* **6**(12), 4588-4606.
- WITTMER, S., 2017: INS- und bildbasierte Messmethoden für bewaldete Gebiete. Masterthesis, **2016**(7), Fachhochschule Nordwestschweiz, Institut Geomatik, (unveröffentlicht).
- SCHWARZ, K.P., MARTELL, H.E., EL-SHEIMY, N., LI, R., CHAPMAN, M.A. & COSANDIER, D., 1993: VIASAT - A Mobile Highway Survey System of High Accuracy. Proceedings of the Vehicle Navigation and Information Systems Conference, Ottawa, 476-481.

# Seamless Outdoor and Indoor Mapping using a LiDAR-based Multi-Sensor System: Case Study on Kalman-Filter Fusion

ARASH JAVANMARD-GH<sup>1</sup>, CHARLES TOTH<sup>2</sup> & DOROTA IWASZCZUK<sup>1</sup>

*Abstract: Due to recent developments in the field of autonomous systems, pedestrian navigation, and Building Information Modeling, increasing interest in 3D indoor maps has been observed. One of the main challenges in indoor modeling is registration with outdoor models. A possible solution is seamless outdoor-indoor mapping using a mobile system operating well in both environments. In this paper, we introduce and compare some indoor and outdoor mapping methods for fusing outputs of a mobile multi-sensor system, such as IMU and LiDAR, where we enhance classical Extended Kalman Filter (EKF) based methods. In the first place, we will show the results of a loosely coupled error-based EKF for navigation, where we exploit the IMU measurements for prediction and the output of the iterative closest point (ICP) algorithm on the point cloud of the LiDARs for correction. Furthermore, we will also examine and compare the leverage of different ICP algorithms on the task of pose estimation.*

## 1 Introduction

Accurate and georeferenced 3D maps of built areas have been gaining in importance over past years. Such concepts as smart cities rely on up-to-date models of the built-up world. Further-more, modern navigation solutions, including navigation of continuously growing number of autonomous systems and seamless pedestrian navigation, require accurate and georeferenced maps, including 3D indoor scenes. Indoor mapping is also of high interest for such applications as building maintenance and Building Information Modeling (BIM).

Georeferencing based on Global Navigation Satellite System (GNSS) can be successfully applied in outdoor. In indoor or in some outdoor environments the GNSS signal is however too weak for reliable georeferencing. Many approaches allowing the creation of indoor models using inertial measurement units (IMU) and/or point cloud matching can be found in the literature. Creating the connection between indoor and outdoor maps, however, remains challenging. Therefore, we investigate the feasibility of seamless outdoor-indoor mapping using a mobile multi-sensor mapping system. For this purpose, we compare and analyze the results of different indoor and outdoor mapping approaches using the KITTI (GEIGER et al. 2012) dataset and one where a mobile platform consisting of three laser scanners (LiDARs), six RGB cameras, two GPS receivers and an IMU (IWASZCZUK et al. 2019) was used to collect geometric and photogrammetric information of an urban environment, including road networks, vegetation, city furniture, and building facades together with detailed structures, such as windows and doors as well as indoor scene with multiple floors and rooms. Due to the lack of synchronization at the time of publishing this paper we will use just the LiDAR point cloud from this dataset.

---

1 Technische Universität Darmstadt, Remote Sensing and Image Analysis, Franziska-Braun-Straße 7, D-64287 Darmstadt, E-Mail: iwaszczuk@geod.tu-darmstadt.de

2 The Ohio State University Columbus, Dept. of Civil, Environmental and Geodetic Engineering, OH, USA

## 2 Related Works

There have been many studies on laser or camera-based localization and mapping recently. Pose-Graph based approaches (GRISSETTI et al. 2010; HESS et al. 2016; YE & LIU 2017) formulate the task of localization and mapping as a Maximum a Posteriori (MAP), where a front-end builds up the graph by connecting nodes depending on the measurements and a back-end subsequently tries to find the optimal estimation using non-linear optimization frameworks (CADEN et al. 2016). Nevertheless, Kalman-Filter based approaches are still of interest, especially an error-state based Extended Kalman Filter (ES-EKF) is due to its simplicity and the fact that its computation does not depend on growing covariance-matrix is very suitable for applications where many different sensory sources with different updates rate have to be fused to produce a more accurate trajectory estimation. Excellent work for a vision-aided inertial navigation system based on ES-EKF can be find in (MOURIKIS & ROUMELIOTIS 2007), by integrating the IMU measurement in the prediction step and a correction based on the camera triangulation. ZHEN et al. (2017) use an ES-EKF to fuse the odometry information of an IMU with Gaussian Particle Filter measurement updates. In this paper, we build up on their idea. But in contrast to their work, we will use different Iterative Closest Point (ICP) variations at the measurement step. ICP is by far one of the most popular methods for matching multiple point clouds. In the basic form, ICP is an optimization problem that is solved iteratively to find a hopefully good estimation of rotation and translation, which align the point clouds. Even though there exist many different variants, we will use two popular formulations, point-to-point (BESL & MCKAY 1992) and point-to-plane (YANG & MEDIONI 1992). Furthermore, we will use the open3d-library for point cloud processing (ZHOU et al. 2018).

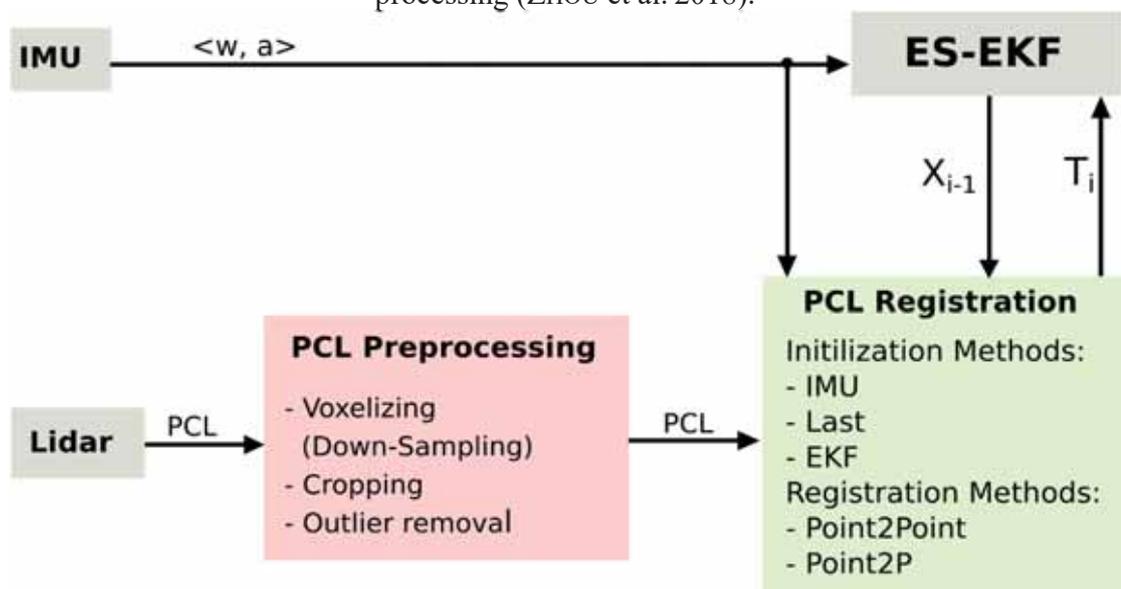


Fig. 1 gives an overview of the system, in section 3 we will give a description of the Point Cloud (PCL) processing and registration steps and in section 4 we will describe the procedure of ES-EKF for state estimation. Finally, in section 5 we present and discuss our results.

### 3 Point Cloud preprocessing and registration

In the preprocessing step the raw point cloud measurements are down sampled using a voxel grid filter based on its nearest neighbors ( $N$ ) and a specific radius ( $r$ ), which are both hyperparameters of the system and should be set a priori by the user. Since the noise of the lidar scanner increases with the distance of the points we also crop the point cloud by a spherical shape. The inner and outer radius of the shape are also hyper-parameters which should be chosen carefully by the user. At least a statistical outlier removal is applied to each point cloud. The adjusted output of the PCL-Processor is used by the PCL-Registration module. Like almost any other non-linear optimization-based approaches, also ICP needs an initial estimate of the transformation between two consecutive point clouds. We have implemented three approaches for this, the first initialization is done based on the assumption of constant velocity model, where we take the last estimate  $T_{i-1}$ . In the second case, we pre-integrate IMU measurements (FORSTER et al. 2015) between each LiDAR frame and calculated the initial transformation between two consecutive frames

$$T_{i-1,i}^b = T_{nb,i-1}^{-1} T_{nb,i} \quad (1),$$

where  $b$  is the body frame and  $n$  is the navigation frame.

Since IMU measurements suffer from noise by just integrating the measurements, our pose estimation will drift over time. To overcome this issue, we use the output of the ES-EKF to reset the last prediction of the IMU and start to integrate based on this corrected state.

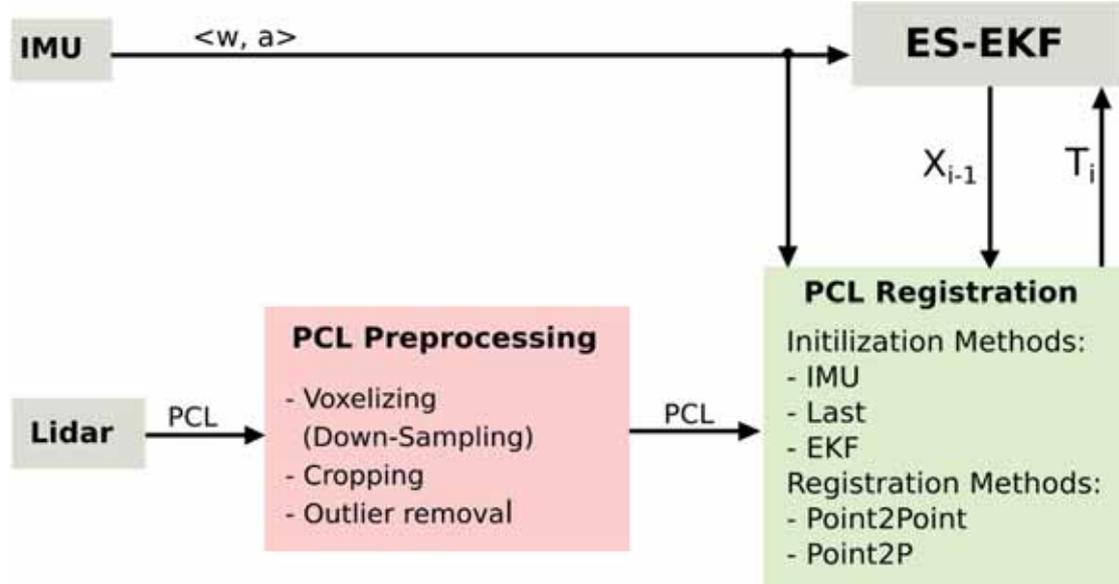


Fig. 1: System preview consisting of LiDAR preprocessing and registration, Extended Kalman Filter for state estimation and IMU measurements

## 4 IMU/Lidar fusion using ES-EKF

Dead reckoning is a widely used approach for pose estimation, wherein the goal is to find the optimal pose estimation by fusing different signals. One of the most widely used sensor combinations for pose estimation is IMU and Lidar. In the ES-EKF manner the linear acceleration and angular velocity measurements are integrated in the prediction step at the update rate of the IMU, which is typically 100-200 Hz and the output of a scan-matching algorithm which runs on much lower frequency 10Hz is fused for correction. Which is based on the LiDAR point clouds by calculating the transformation between each successive point cloud frames as it was elucidated in Sec. 3. In this work, we decided to use an Error-State EKF, due to its calculation simplicity and the fact the error states are close to zero, so that especially small-signal rotation approximation is linear and does not drift much from the nominal state (MADYASTHA et al. 2011)

$$R(\delta\theta) \approx I + [\delta\theta]_{\times} \quad (2).$$

Here is  $[\delta\theta]_{\times}$  a skew-symmetric matrix of a small rotation  $\delta\theta$ .

Our formulation of the ES-EKF is based on the detailed investigative work in (SOLÀ 2017), but in contrast to that work we will not estimate the IMU bias in the state vector. Furthermore, we assume that our initial frame is a horizontal plane so we do not need to estimate the gravity vector at the initial frame. So that the nominal state dynamic vector  $\dot{x}$  and the system error dynamics vector  $\delta\dot{x}$  are represented as

$$\dot{x} = \begin{bmatrix} \dot{p} \\ \dot{v} \\ \dot{q} \end{bmatrix} = \begin{bmatrix} v \\ Ra_m + g \\ 0.5q \otimes w_m \end{bmatrix}, \quad \delta\dot{x} = \begin{bmatrix} \delta\dot{p} \\ \delta\dot{v} \\ \delta\dot{\theta} \end{bmatrix} = \begin{bmatrix} \delta v \\ R[a_m]_{\times} \delta\theta - a_n \\ w_m \delta\theta - w_n \end{bmatrix} \quad (3),$$

where  $\dot{p}, \dot{v} \in R^3$  denote velocity and acceleration vectors and  $q \in H$  is the quaternion representation of orientation. As mentioned before constant gravity vector  $g = [0, 0, 9.81]^T$ . Linear acceleration  $a_m$  and angular velocity  $w_n$  are considered as the control inputs. Furthermore  $a_n, w_n$  are accordingly system noise, which are assumed to be zero mean Gaussian random variables, e.g.,

$$a_n \sim N(0, Q_a), w_n \sim N(0, Q_w) \quad (4).$$

The above in continuous time formulated differential equations should be discretized, since our sensor measurements are available just at some discrete time intervals  $\delta t$ . This can be done by any integration methods, like  $k^{\text{th}}$ -Order Runge-Kutta. We waive to formulate the whole integration step and refer to (SOLÀ 2017) for more information. In this work, we will use the Euler integration method, which is 1th-Order Runge-Kutta.

### 4.1 Prediction step

In the prediction, the nominal state is updated using the IMU measurements as control input as follow:

$$\check{x} = \begin{bmatrix} p_t \\ v_t \\ q_t \end{bmatrix} = \begin{bmatrix} p_{t-1} + v\Delta t + 0.5(R_{t-1}a_{m,t} + g)\Delta t^2 \\ v_{t-1} + (R_{t-1}a_m + g)\Delta t \\ q_{t-1} \otimes q(w_m\Delta t) \end{bmatrix} \quad (5),$$

$$\begin{bmatrix} \delta p_t \\ \delta v_t \\ \delta \theta_t \end{bmatrix} = \begin{bmatrix} \delta p_{t-1} + \delta v\Delta t \\ \delta v - R[a_m]_x\delta\theta\Delta t + \hat{a}_t \\ R(w_m\Delta t)^T\delta\theta + \hat{w}_t \end{bmatrix} \quad (6).$$

Error state itself can be estimated at the same time as measurement is observed, but nevertheless the error covariance  $P$  of the error state should be propagated at the same rate as IMU,

$$\check{P}_t = F_{t-1}P_{t-1}F_{t-1}^T + L_{t-1}Q_{t-1}L_{t-1}^T \quad (7).$$

With  $F$  and  $L$  being the transition matrix and the derivative of the discretized error-state dynamics w.r.t. the noise variables

$$F_t = \begin{bmatrix} I & I\Delta t & 0 \\ 0 & I & -R[a_m]_x\Delta t \\ 0 & 0 & R^T(w_m\Delta t) \end{bmatrix} Q_t = \begin{bmatrix} Q_a & 0 \\ 0 & Q_w \end{bmatrix} L_t = \begin{bmatrix} 0 & 0 \\ I & 0 \\ 0 & I \end{bmatrix} \quad (8).$$

We assume that the covariances  $Q_a, Q_w$  of the IMU are isotropic and can be read form the IMU datasheet.

$$Q_a = \sigma_a^2\Delta t^2 I [m^2/s^2], Q_w = \sigma_w^2\Delta t^2 I [rad^2] \quad (9).$$

The units are the results of integrating the covariances of zero mean Gaussian process variables (MAYBECK & SIOURIS 1980).

## 4.2 Correction step

Using the pose estimation of the PCL-Registration, the measurement model is formulated as

$$y_t = Hx_t + v_t, H = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \end{bmatrix} \in R^{6 \times 9}, v \sim N(0, V_{lidar}) \quad (10).$$

Now by calculating all necessary information, the Kalman Gain and the error made during prediction step are

$$K_t = P_t H_t^T (H_t P_t H_t^T + V)^{-1} \quad (11),$$

$$\delta \hat{x}_t = K_t (y_t - \check{x}_t)$$

In the Closed-Loop correction Kalman Filter (GROVES 2008) the errors estimated by the filter are fed back to the current nominal state and at the last step the error covariance is also corrected as follows

$$\begin{aligned}
\hat{p} &= \check{p} + \delta p_k \\
\hat{v} &= \check{v} + \delta v_k, \hat{P} = (1 - K_t H_t) \check{P} \\
\hat{q} &= q(\delta\theta) \otimes \check{q}
\end{aligned} \tag{12}.$$

## 5 Evaluation and Discussion

In this paper, we present preliminary results of our investigation by evaluating different combinations of the mentioned approaches in indoor and outdoor environments. We investigate, the quality of point cloud registration based on the of pure IMU mechanization, also we show how IMU can aid the ICP registration, especially when, due to a large transformation ICP does not converge to a good solution. Since IMUs are affected by noise, we examine the effect of Kalman Filter ICP registration. At least we compare two different registration methods, namely *point2point* and *point2plane* for outdoor to indoor transition.

Table 1 shows the results of the proposed approaches on the KITTI dataset, where a vehicle was driving for about 90 sec.

In Fig. 2 two different maps made by the proposed approaches are shown, on the right map a strong misalignment, due to drift around the roll axis can be seen. In contrast, however, in the left image, a map and a corresponding trajectory were built.

Tab. 1: Results of different approaches performed on ~90 [sec] of KITTI dataset.

Initialization	Regist. Method	Position Error	Orientation Error
Imu	Imu	31%	12%
Imu	Point2Plane	10%	9%
Imu	Point2Point	19%	19%
Imu + ES-KF	Point2Point	<b>3.5%</b>	3%
Last	Point2Plane	4.5%	8%
Last	Point2Point	4.5%	<b>2.8%</b>

We also test here described methodology on our outdoor-indoor dataset captured by the backpack mobile-mapping system (IWASZCZUK et al. 2019). The Velodyne VLP-16, which was used in this mapping system, consists of 16 lines, according to this, the point clouds collected at each revolution are sparse, unlike the LiDAR systems with many more lines, like one used in KITTI dataset. This sparsity makes an appropriate registration much harder, specially *point2plane* registration produces more outliers and bad registrations, since the radius for estimating the normals at each point must be large enough to make a plane estimation accurate enough. But on the other side, large radius may result in bad normal estimations since the points may not belong to the same surface anymore. As can be seen in Fig. 3 and Fig. 4, *point2point*-method produces fewer misalignments compared to *point2plane* (red area). Nevertheless, both approaches fail to keep a good estimate of the height; especially in an indoor area the floors are collapsed after some iterations.

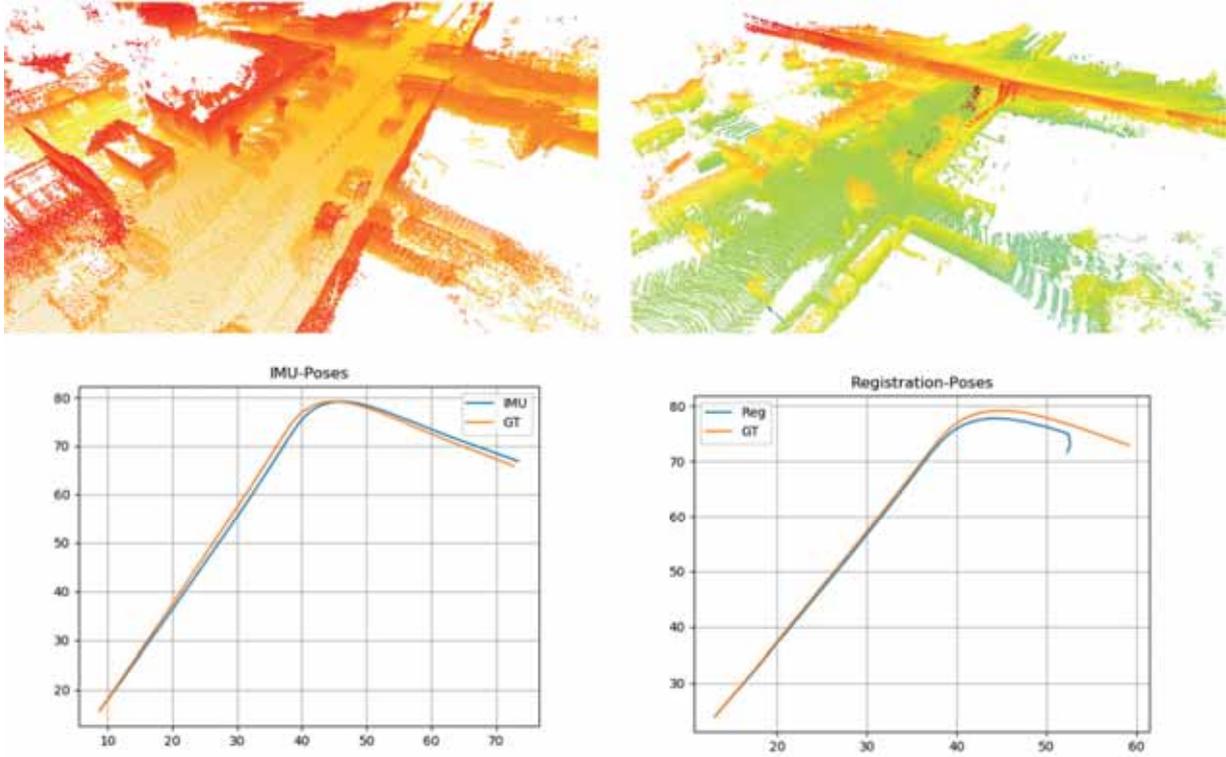


Fig. 2: Top: A 3D map of the scene, right) good registration results by IMU+EKF, left) Registration by pure IMU, where a drift about the roll axis caused a bad registration result. Bottom: Comparison of the estimated trajectory against ground truth

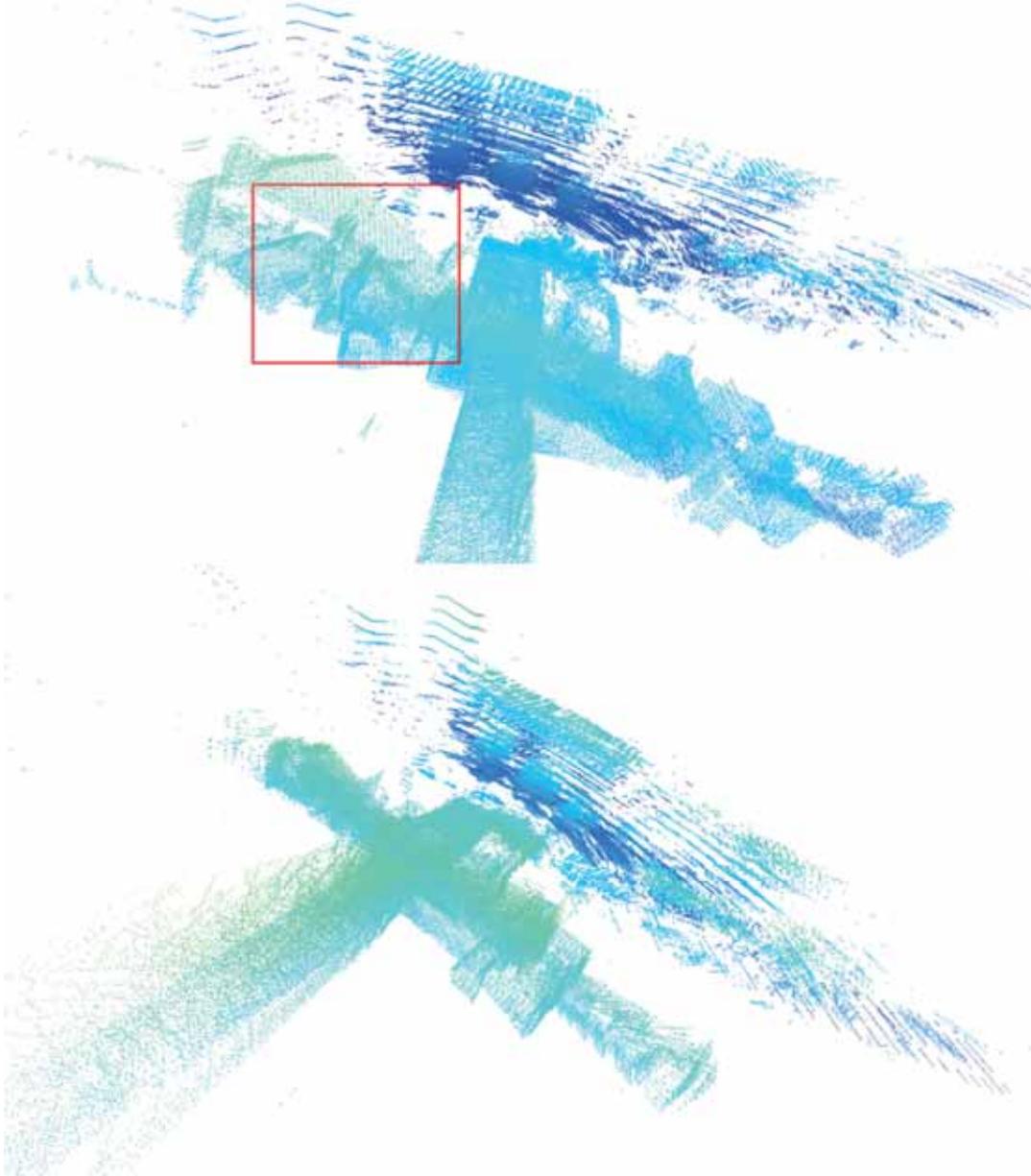


Fig. 3: 3D-Map created by *point2plane* registration (top), 3D-Map created by *point2point* registration (bottom)

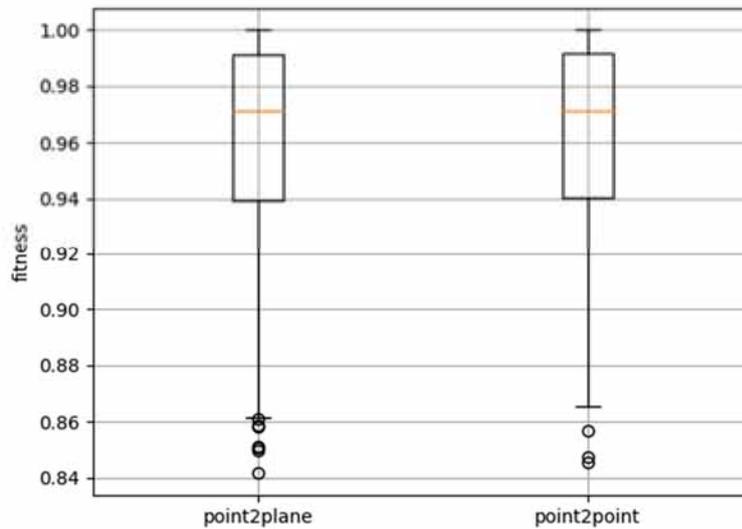


Fig. 4: Comparison between *point2point* and *point2plane* fitness values

## 6 Future Work

In this study, we have removed the estimation of the IMU bias from ES-EKF formulated in this paper, and assumed it is given and constant. This assumption is rather rarely true, unless we have a highly accurate navigation grade sensor. We expect that moving the IMU bias into the dynamics of the error state, provided that they are observable, would result in more accurate state estimation. This issue will be the subject of an in-depth future investigation. Moreover, one drawback of using EKF in both, loosely coupled or tightly coupled modes, is that due to the linearization and the presence of potentially strong non-linear effects, the map becomes more inconsistent over time. To overcome this issue, we plan to explore the possibilities of combining Pose Graph-based approaches; especially for building a more accurate map during the outdoor-indoor transition. Another issue we have observed during this work is the weakness of current ICP based methods when registering sparse Point Clouds, like those generated by Velodyne VLP-16. Therefore, we will also investigate the benefits of using more sophisticated feature-based registration methods. For example, we are going to perform an unsupervised classification based on the geometrical features extracted from point clouds. Afterward, feature association followed by registration is performed. We hope to get more satisfactory registration results than pure ICP. We also plan to investigate deep learning methods either as a part of the whole system, e.g. for registration, or in an end-to-end network for the whole task of odometry estimation.

## 7 References

- BESL P. J. & MCKAY, N. D., 1992: A Method for Registration of 3-D Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**(2), 239-256.
- CADEN, C., CARLONE, L., CARRILLO, H., LATIF, Y., SCARAMUZZA, D., NEIRA, J., REID, I. & LEONARD, J.J., 2016: Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on robotics* **32**(6), 1309-1332.
- FORSTER, C., CARLONE, L., DELLAERT, F. & SCARAMUZZA, D., 2015: IMU preintegration on manifold for efficient visual-inertial maximum-a-posteriori estimation. In *Robotics: Science and Systems*, <http://www.roboticsproceedings.org/rss11/p06.pdf>.
- GEIGER, A., LENZ, P. & URTASUN, R., 2012: Are we ready for autonomous driving? the KITTI vision benchmark suite. *IEEE Conference on Computer Vision and Pattern Recognition*, 3354-3361.
- GRISSETTI, G., KUMMERLE, R., STACHNISS, C. & BURGARD, W., 2010: A tutorial on graph-based SLAM. *IEEE Intelligent Transportation Systems Magazine* **2**(4), 31-43.
- GROVES, P. D., 2008: Principles of GNSS Inertial and Multi-Sensor Integrated Navigation Systems. In *GNSS Technology and Applications*. 2008.
- HESS, W., KOHLER, D., RAPP, H. & ANDOR, D., 2016: Real-time loop closure in 2D LIDAR SLAM. *IEEE International Conference on Robotics and Automation (ICRA)*, 1271-1278.
- IWASZCZUK, D., KOPPANYI, Z., PFRANG, J., TOTH, C., 2019: Evaluation of a Mobile Multi-Sensor System for Seamless Outdoor and Indoor Mapping. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **42**(1/W2), 31-35.
- MADYASTHA, V. K., RAVINDRAY, V. C., MALLIKARJUNAN, S. & GOYAL, A., 2011: Extended Kalman filter vs. error state Kalman filter for aircraft attitude estimation. *AIAA Guidance, Navigation, and Control Conference*, 6615.
- MAYBECK, P. S. & SIOURIS, G. M., 1980: Stochastic Models, Estimation, and Control. Volume I, *IEEE Transactions on Systems, Man and Cybernetics*.
- MOURIKIS, A. I. & ROUMELIOTIS, S. I., 2007: A multi-state constraint Kalman filter for vision-aided inertial navigation. *IEEE International Conference on Robotics and Automation*, 3565-3572.
- SOLÀ, J., 2017: Quaternion kinematics for the error-state Kalman filter. *arXiv:1711.02508*.
- YANG C. & MEDIONI, G., 1992: Object modelling by registration of multiple range images. *Image Vision Comput.* **10**(3), 145-155
- YE, H., & LIU, M., 2017: LiDAR and Inertial Fusion for Pose Estimation by Non-linear Optimization. *arXiv preprint arXiv:1710.07104*.
- ZHEN, W., ZENG, S. & SOBERER, S., 2017: Robust localization and localizability estimation with a rotating laser scanner. *IEEE International Conference on Robotics and Automation (ICRA)*, 6240-6245.
- ZHOU, Q.-Y., PARK, J. & KOLTUN, V., 2018: Open3D: A Modern Library for 3D Data Processing. *arXiv preprint arXiv:1801.09847*.

# Unsicherheitsauswertung von semantischer Segmentierung mittels Neuronaler Netze

LINA E. BUDDE<sup>1,2</sup>, STEFAN SCHMOHL<sup>1</sup> & UWE SÖRCEL<sup>1</sup>

*Zusammenfassung: Künstliche Neuronale Netze haben sich in weiten Teilen der maschinellen Dateninterpretation als Stand der Technik etabliert. Jedoch ist die Qualitätseinschätzung ihrer Klassifikationsergebnisse ohne gegebene Ground Truth noch ein offenes Forschungsfeld. Von zunehmendem Interesse ist dabei das sogenannte Monte-Carlo Dropout, mit welchem eine Verteilung der prädizierten Pseudo-Wahrscheinlichkeiten geschätzt werden kann. In dieser Arbeit werden verschiedene daraus abgeleitete Unsicherheitsmaße sowohl quantitativ als auch visuell miteinander verglichen. Als Studienobjekt dient dafür der ISPRS Potsdam 2D Semantic Labeling Benchmark Datensatz zur Luftbildklassifikation. Es zeigt sich, dass insbesondere mittels der Shannon Entropie der Pseudo-Wahrscheinlichkeiten unsichere bzw. fehlerhafte Pixelklassifikationen identifiziert werden können. Diese identifizierten Pixel befinden sich in erster Linie an Objektändern und ermöglichen beispielsweise durch die Entfernung der unsichersten 10 % der Pixel eine Steigerung der Gesamtgenauigkeit von 82,6 % auf 86,5 %.*

## 1 Einleitung

Derzeit genießt maschinelles Lernen, im Speziellen Deep Learning, große Aufmerksamkeit in der Wissenschaft und zunehmend auch in der praktischen Anwendung. Dabei gibt es keine Beschränkung auf eine spezielle Disziplin, sondern eine Vielzahl von Anwendungsmöglichkeiten, zum Beispiel in den Bereichen Physik, Biologie und Fertigung. Eine häufige Aufgabe in der Bildanalyse im Allgemeinen und in der Fernerkundung im Besonderen ist die sogenannte semantische Segmentierung von Bildern, bei der jedem Pixel eine Klasse zugeordnet wird. Aus Fernerkundungsdaten können so qualitativ hochwertige und aktuelle Landbedeckungskarten erstellt werden, die zum Beispiel für die Land- und Forstwirtschaft, Stadtentwicklung und Umweltüberwachung genutzt werden.

Die damit gewonnene, pixelweise Klassenzugehörigkeit ist jedoch als Information nicht immer ausreichend. Solche automatisch erzeugten Klassifikationen sind grundsätzlich fehlerbehaftet. Zusätzliche Informationen zu ihrer Qualität sind daher ebenfalls wünschenswert.

Die Bedeutung solcher Qualitätsmaße im Bereich der Fernerkundung ergibt sich dadurch, dass gerade die Landbedeckungskarten eine wichtige Grundlage sowohl politischer als auch wirtschaftlicher Entscheidungen bilden. Indikatoren für mögliche Fehlzuordnungen können außerdem verwendet werden, um effizient manuell nachzubearbeitende Bereiche auszuwählen, sollte eine sehr hohe Klassifikationsgüte erforderlich sein.

---

<sup>1</sup> Universität Stuttgart, Institut für Photogrammetrie, Geschwister-Scholl-Str. 24D, D-70174 Stuttgart, E-Mail: st115986@stud.uni-stuttgart.de, [stefan.schmohl, uwe.soergel]@ifp.uni-stuttgart.de

<sup>2</sup> Technische Universität Darmstadt, Institut für Geodäsie, Franziska-Braun-Straße 7, D-64287 Darmstadt, E-Mail: lina.budde@tu-darmstadt.de

Die bei Klassifikationen üblichen Genauigkeitsmaße zur Beurteilung der Qualität benötigen allerdings Referenzdaten als direkte Vergleichswerte, im Fall der Fernerkundung z.B. ausgewählte Referenzflächen im Testgebiet. Solche Referenzdaten mit bekannten Klassenzuordnungen sind aufwendig in der Erstellung und daher leider nur begrenzt verfügbar. Eine Möglichkeit diesem Problem entgegenzuwirken, wird von KÖLLE et. al. (2020) aufgezeigt. Dort wird die sogenannte Vote Entropie eingesetzt, um schrittweise mittels Active Learning die Verbesserung eines Random Forest Klassifikators zu erzielen. Die Quantifizierung der Klassifikations-Unsicherheit stellt somit eine Alternative dar, bei der Unsicherheiten unmittelbar vom Klassifikationsmodell mitgeliefert werden. Jedoch können die meisten Deep Learning Modelle solche Unsicherheiten nicht direkt ausgeben. Stattdessen werden häufig die prädierten Softmax-Scores als Wahrscheinlichkeiten und damit fälschlich als Modellkonfidenz interpretiert (GAL 2016).

Um dennoch ein fundiertes Maß für die Unsicherheit zu erhalten, wird in der vorliegenden Arbeit das von GAL & GHARAMANI (2015) entwickelte Monte-Carlo Dropout eingesetzt, um Klassifikations-Ergebnisse von Luftbildern zu beurteilen. Der Einsatz von Monte-Carlo Dropout im Kontext semantischer Segmentierung wurde erstmals von KENDALL et al. (2015) vorgeschlagen. KAMPPMEYER et al. (2016) zeigten, dass aus Monte-Carlo Dropout erzeugten Standardabweichungen als Unsicherheitsmaße per Schwellwert Fehlklassifikationen in Luftbildern herausgefiltert werden können. In dieser Arbeit präsentieren wir eine detaillierte Analyse dreier aus Monte-Carlo Dropout abgeleiteter Unsicherheitsmaße (Standardabweichung, Shannon Entropie und Mutual Information) und untersuchen diese hinsichtlich ihrer Tauglichkeit, falsch klassifizierte Pixel zu identifizieren.

## 2 Methodik

### 2.1 Monte-Carlo Dropout

Beim Dropout handelt es sich um eine übliche Regularisierungstechnik für neuronale Netze. Es basiert auf der Idee des Ensemble-Lernens, wobei mehrere schwache Klassifikatoren zusammen einen starken ergeben. Statt mehrere Netze separat zu trainieren, werden beim Dropout Subnetze innerhalb des eigentlichen Netzes simuliert. Hierfür wird beim Training pro Iteration ein anderer Teil der Neuronen deaktiviert, üblicherweise in einer der letzten neuronalen Schichten. Bei der Inferenz sind dagegen wieder alle Neuronen aktiv.

Die spezielle Variante des Monte-Carlo Dropouts ermöglicht die effiziente Simulation Bayesscher Netze (GAL, 2016). Statt Dropout bei der Inferenz zu deaktivieren, wird es nun genutzt, um bei mehreren Klassifikationsdurchläufen der Daten durch das neuronale Netz variierende Pseudo-Wahrscheinlichkeiten pro Punkt und Klasse zu erzielen, deren Verteilungsparameter hier als Unsicherheitsparameter dienen. Als prädiert gilt jene Klasse mit den im mittel höchsten Pseudo-Wahrscheinlichkeiten. Gegenüber der Verwendung von Bayesschen Netzen zur Unsicherheitsauswertung besteht der wesentliche Vorteil dieses Verfahrens darin, dass es in jedem neuronalen Netz mit Dropout eingesetzt werden kann, ohne zusätzliche Parameter aufwendig im Optimierungsprozess bestimmen zu müssen. Auch wird kein a-priori Vorwissen über Randverteilungen benötigt.

## 2.2 Unsicherheitsmaße

Aus den variierenden Pseudo-Wahrscheinlichkeiten beim Monte-Carlo Dropout können pro Punkt verschiedene Werte abgeleitet werden, die als Unsicherheitsmaße verwendet werden. Unter anderem ergibt sich über die Verteilungen der Pseudo-Wahrscheinlichkeiten pro Klasse eine Standardabweichung  $\sigma_c$ , welche über alle Klassen gemittelt werden (KAMPFFMEYER et al., 2016):

$$\sigma_c = \sqrt{\frac{1}{T} \sum_{t=1}^T (p_c^t - \bar{p}_c)^2} \quad (1)$$

Dabei steht  $p_c$  für den Softmax-Output für die Klasse  $c$  am jeweiligen Punkt und  $T$  für die Anzahl an Stichproben. Die gemittelte Standardabweichung über alle Klassen stellt ein Maß für die gesamte Modellunsicherheit dar, welches unabhängig von den einzelnen Klassen ist. Als weiteres Maß dient die normierte Shannon Entropie  $H$  aus den gemittelten Stichproben pro Punkt:

$$H = -\frac{1}{\log_2(C)} \sum_{c=1}^C \bar{p}_c \cdot \log_2(\bar{p}_c) \quad (2)$$

Die Shannon Entropie ermöglicht Aussagen zum Informationsgehalt pro Punkt. Dabei ist der Informationsgehalt dann größer, wenn die Unsicherheit größer ist. Das dritte Unsicherheitsmaß ist schließlich die Mutual Information (GAL 2016):

$$I_{MC} = \frac{1}{\log_2(C)} \left[ -\sum_{c=1}^C \left( \sum_{t=1}^T p_c^t \cdot \log_2 \sum_{t=1}^T p_c^t \right) + \frac{1}{T} \sum_{c,t=1}^{c,T} p_c^t \cdot \log_2(p_c^t) \right] \quad (3)$$

Über die Mutual Information wird der Unterschied zwischen der Entropie der Prädiktion und der A-posteriori Wahrscheinlichkeit der Parameter angegeben.

## 3 Versuchsaufbau

### 3.1 Semantische Segmentierung

Für die Untersuchung der Klassifikations-Unsicherheiten wird in dieser Studie der ISPRS Potsdam 2D Semantic Labeling Benchmark<sup>3</sup> verwendet, bestehend aus 38 6000 × 6000 Pixel großen Multispektralbildern mit 5 cm GSD (GERKE et al. 2014). Für die semantische Segmentierung stehen somit zum einen True Orthophotos der Kanäle RGB & IR, zum anderen ein dazugehöriges nDSM zur Verfügung. Das nDSM leitet sich aus einem DSM ab, welches mittels dense image matching erzeugt wurde.

Abweichend zum ursprünglichen Benchmark wurde in dieser Arbeit eine andere Verteilung von Test- und Trainingsbildern verwendet. Der neue Testsatz besteht aus nur noch acht statt vierzehn

<sup>3</sup> <http://www2.isprs.org/commissions/comm3/wg4/2d-sem-label-potsdam.html>

Bildern und ist über das gesamte Gebiet verteilt. Auf diese Weise konnte u.a. der bisher ausschließlich in den Trainingsdaten vorhandene See berücksichtigt werden. Auch wurden so leicht höhere Genauigkeiten erzielt.

Aufgrund der begrenzten Rechenleistung der verwendeten Hardware wird eine Auflösungsreduktion um den Faktor zwei für diese Daten durchgeführt. Die semantische Segmentierung erfolgt mittels eines einfachen U-Net (RONNEBERGER et al. 2015) und folgt somit einer Encoder-Decoder Struktur. Jedoch werden zu dem von RONNEBERGER et al. (2015) veröffentlichten U-Net Ergänzungen, wie Zero-Padding, Batch-Normalisierung und Dropout hinzugefügt. Das so aufgebaute Netz wird mit Bildausschnitten von jeweils  $300 \times 300$  Pixel unter der Verwendung der fünf zu Verfügung stehenden Kanäle RGB, IR & nDSM trainiert. Im Gegensatz dazu wird für die Testphase ein Bild in vier Teile zerteilt, die aus  $1.500 \times 1.500$  Pixeln ( $\cong 150 \text{ m} \times 150 \text{ m}$  auf dem Boden) bestehen.

Die für das Training notwendigen Referenzdaten beinhalten sechs verschiedene Klassen: „versiegelte Oberfläche“, „Gebäude“, „niedrige Vegetation“, „Bäume“, „Autos“ und alle weiteren Objekte zusammengefasst in der Klasse „Sonstiges“. Für das Training selbst wird zusätzlich eine künstliche Datenerweiterung durchgeführt. Durch Rotation, Spiegelung und additivem Rauschen ergibt sich eine größere Variation der Trainingsdaten, sodass die Gefahr der Überanpassung verringert werden kann. Als Verlustfunktion wird eine gewichtete Kreuzentropie mit weight decay verwendet. Dadurch kann die ungleiche Auftrittshäufigkeit der verschiedenen Klassen berücksichtigt werden. Insgesamt beträgt die Dauer der Trainingsphase 50 Epochen. Durch das Aufspalten des Datensatzes in Trainings-, Validierungs- und Testdaten, kann bereits während der Trainingsphase der Trainingserfolg mit den Validierungsdaten überprüft werden. Nach der Trainingsphase werden für die Auswertung, die bisher für das Netz unbekannt, Testdaten verwendet.

### 3.2 Unsicherheitsauswertung

Aus den Pseudo-Wahrscheinlichkeiten von insgesamt 20 Monte-Carlo Durchläufen je Bild werden die drei Unsicherheitsmaße Standardabweichung, Shannon Entropie und Mutual Information gebildet. Dazu werden die in den Gleichungen (1), (2) und (3) angegebenen Formeln genutzt. Zusätzlich wird die Shannon Entropie auch ohne Monte-Carlo Dropout bestimmt. Dadurch kann verglichen werden, inwieweit das Monte-Carlo Dropout für eine Unsicherheitsauswertung notwendig ist.

Mithilfe von Schwellwerten können aus den so erhaltenen Unsicherheitsmaßen anschließend die unsichersten Klassifikationen herausgefiltert werden, um einen möglichst großen Teil der Fehlklassifikationen zu entfernen. Dabei können sowohl absolute als auch relative Schwellwerte vorgegeben werden. Bei einem absoluten Schwellwert darf das jeweilige Pixel diesen Schwellwert nicht überschreiten, um als sicher eingestuft zu werden. Wird dagegen ein relativer Schwellwert angegeben, wird dadurch die Anzahl der unsicheren beziehungsweise sicheren Pixel festgelegt, die auf Grundlage ihres Unsicherheitswertes aussortiert beziehungsweise beibehalten werden.

## 4 Ergebnisse

### 4.1 Semantische Segmentierung

Aus den 20 Inferenzdurchläufen mit aktiviertem Monte-Carlo-Dropout gilt pro Pixel diejenige Klasse als prädiziert, welche den höchsten gemittelten Softmax-Score besitzt. Das Ergebnis einer solchen semantischen Segmentierung zeigt beispielhaft Abb. 1 für einen kleineren Testbildausschnitt.

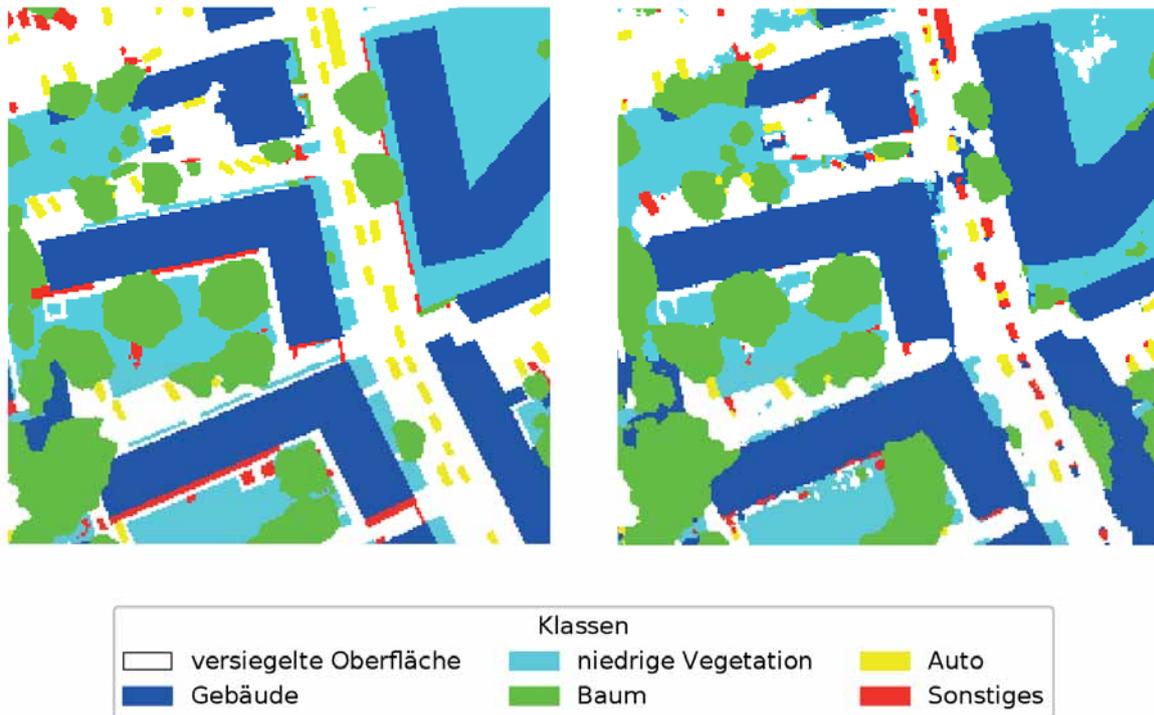


Abb. 1: Testbildausschnitt. Links: Referenz, rechts: prädizierte Label

Mit Hilfe der Referenzdaten ergeben sich aus den acht Testbildern die in Abb. 2 dargestellte Confusion Matrix und die in Tab. 1 enthaltenen F1-Werte. Es zeigt sich, dass besonders die Klasse „Gebäude“ sehr gut erkannt wird. Dagegen kann für die Klasse „Sonstiges“ nur eine geringe Genauigkeit erreicht werden. Dies liegt wahrscheinlich hauptsächlich in der Inhomogenität der Klasse „Sonstiges“. Starke Verwechslungen existieren zwischen den beiden Vegetationsklassen „Baum“ und „niedrige Vegetation“. Zudem konnte eine große Abhängigkeit der Klassifikation von der Qualität des gegebenen nDSM festgestellt werden, da in Bereichen von fehlerhaften nDSM-Daten die Klassifikation ebenfalls fehlerhaft war. Insgesamt ergibt sich mit dem verwendeten Netzwerk über alle Testbilder eine Gesamtgenauigkeit von 82,6 %.

Neben den Fehlklassifikationen bei den Klassen „Sonstiges“ und „Auto“ werden in Abb. 1 die teils ausgefranzten Objektkanten sichtbar. Die Identifikation solcher Fehlklassifikationen ist bei der nachfolgenden Unsicherheitsauswertung wünschenswert.

Tab. 1: F1-Werte der einzelnen Klassen aus den Testdaten

	Versiegelte Oberfläche	Gebäude	Niedrige Vegetation	Baum	Auto	Sonstiges
F1 [%]	87,7	90,3	79,6	72,2	65,7	23,7

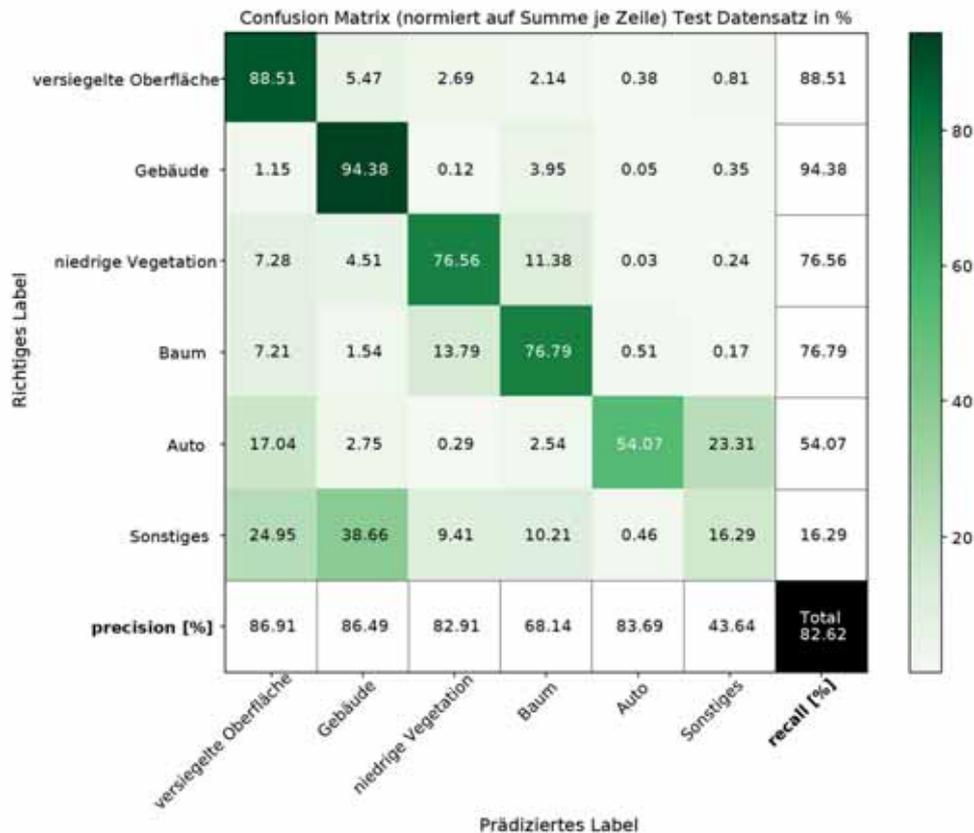


Abb. 2: Normierte Confusion Matrix in % aus den Testdaten

## 4.2 Unsicherheitsmaße

Zusätzlich zu Abb. 1 enthalten die Abb. 3 bis Abb. 6 die gemittelten Softmax-Scores sowie die drei Unsicherheitsmaße aus Monte-Carlo-Dropout für denselben Testbildausschnitt. Allen Abbildungen ist gemeinsam, dass Objektstrukturen eine homogene innere Färbung besitzen. Deutlich ist ein sich durchziehender senkrechter Bruch zu erkennen. Dies hängt mit der Zerteilung der einzelnen Testbilder in jeweils vier Bildausschnitte für die Prozessierung zusammen. Durch die iterative Klassifikation der einzelnen Bildausschnitte durch das Netz basieren die einzelnen Pseudowahrscheinlichkeiten bei jedem Bildausschnitt durch das Monte-Carlo Dropout auf unterschiedlichen Sub-Netzen. Dies betrifft nicht nur die gemittelten Softmax-Scores der prädizierten Klasse, sondern auch alle aus dem Monte-Carlo Dropout abgeleiteten Größen. Diese Variationen der einzelnen Bildausschnitte sind ebenfalls ein Zeichen für die Unsicherheit des Modells, da ohne Unsicherheiten jedes der Sub-Netze ähnliche Ergebnisse liefern sollte. Jedoch haben diese Variationen hier keinen Einfluss auf die prädizierte Klasse. Wie in Abb. 1 zu sehen ist.

Die mittlere Standardabweichung und Mutual Information erzeugen bei der optischen Betrachtung ähnliche Ergebnisse (Abb. 3b & Abb. 4a). Die Unsicherheitswerte sind bei Objekten innerhalb einer Klasse konsistent. So weisen Bäume im Großteil höhere Werte auf als niedrigere Objekte. Die geringsten Werte besitzen Gebäude. Im Ausschnitt rechts unten sind die Verhältnisse teilweise invertiert. Die Unsicherheiten scheinen hier daher in erster Linie mit der Klassenart zu korrelieren und weniger mit Fehlklassifikationen.

Im Gegensatz dazu treten bei der Shannon Entropie die höchsten Werte klar als linienförmige Strukturen auf (Abb. 4b). In diesen Bereichen sind die Unsicherheitswerte am stärksten und befinden sich entlang der Klassen- bzw. Objektgrenzen, was auch für einen großen Teil der Fehlklassifikationen zutrifft (vgl. Abb. 6). Ein ähnliches Muster zeigt sich bei den oft als Unsicherheitsmaß herangezogenen, hier mittleren, Softmax-Scores aus Abb. 3a.

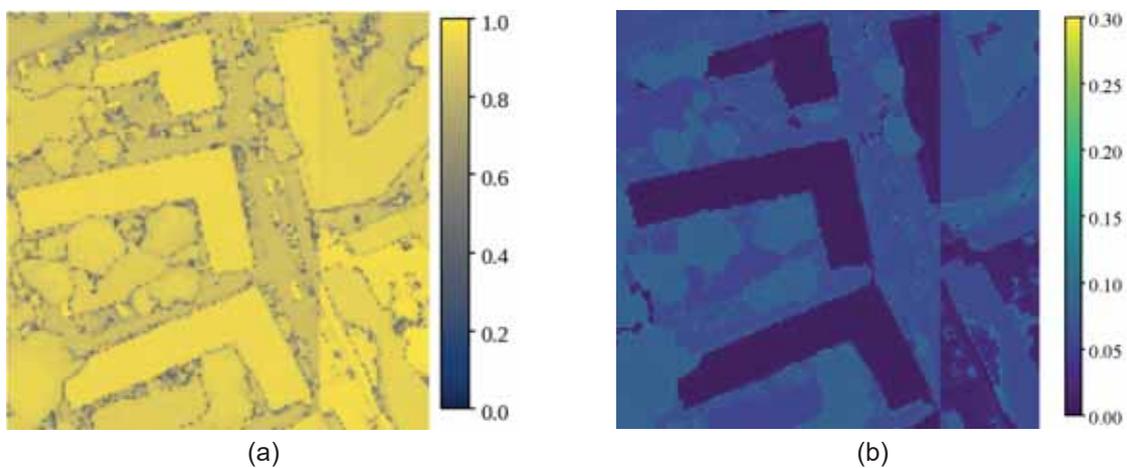


Abb. 1: Testbildausschnitt. (a): mittlere Softmax-Scores (Pseudo-Wahrscheinlichkeiten) der jeweils stärksten Klasse, (b): über alle Klassen gemittelte Standardabweichung

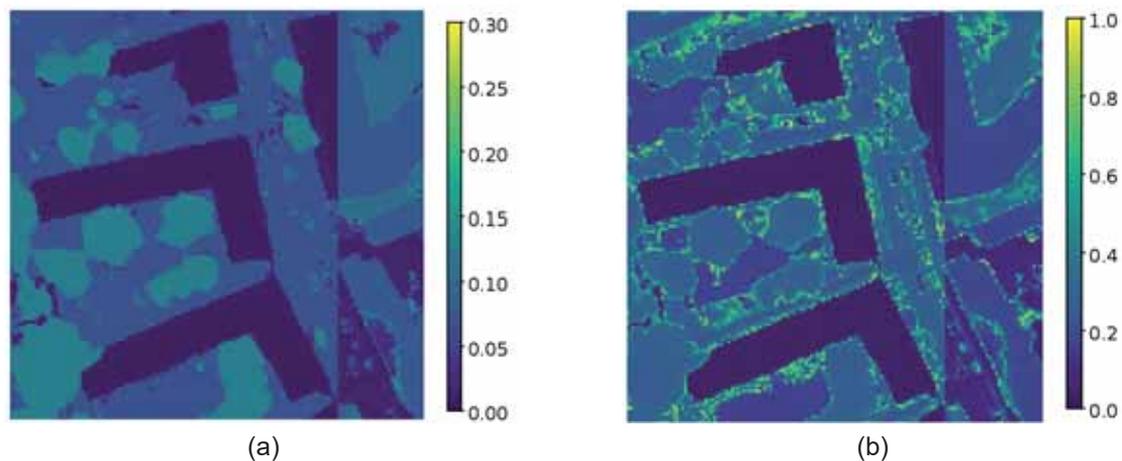


Abb. 2: Testbildausschnitt. (a): Mutual Information, (b): Shannon Entropie mit Monte-Carlo Dropout

Um die Unsicherheitsmaße besser vergleichen zu können, werden anhand verschiedener Schwellwerte Precision-Recall Kurven erstellt (Abb. 5). Dafür werden die Pixel mit Unsicherheitswerten,

die größer als der jeweilige Schwellwert sind, verworfen. Es zeigt sich, dass die Shannon Entropie mit Monte-Carlo Dropout das geeignetste Verfahren zur Bestimmung der Zuverlässigkeit ist. Dicht gefolgt von der Precision-Recall Kurve der gemittelten Softmax-Scores der prädierten Klasse. Deutlich schlechter schneiden die Standardabweichung und die Mutual Information ab. Da sich die Shannon-Entropie auch ohne Monte-Carlo-Dropout berechnen ließe, ist zum Vergleich auch solch eine Kurve dargestellt. Diese ist besonders auffällig, da die maximal erreichte Precision im Bereich von 84 % liegt.

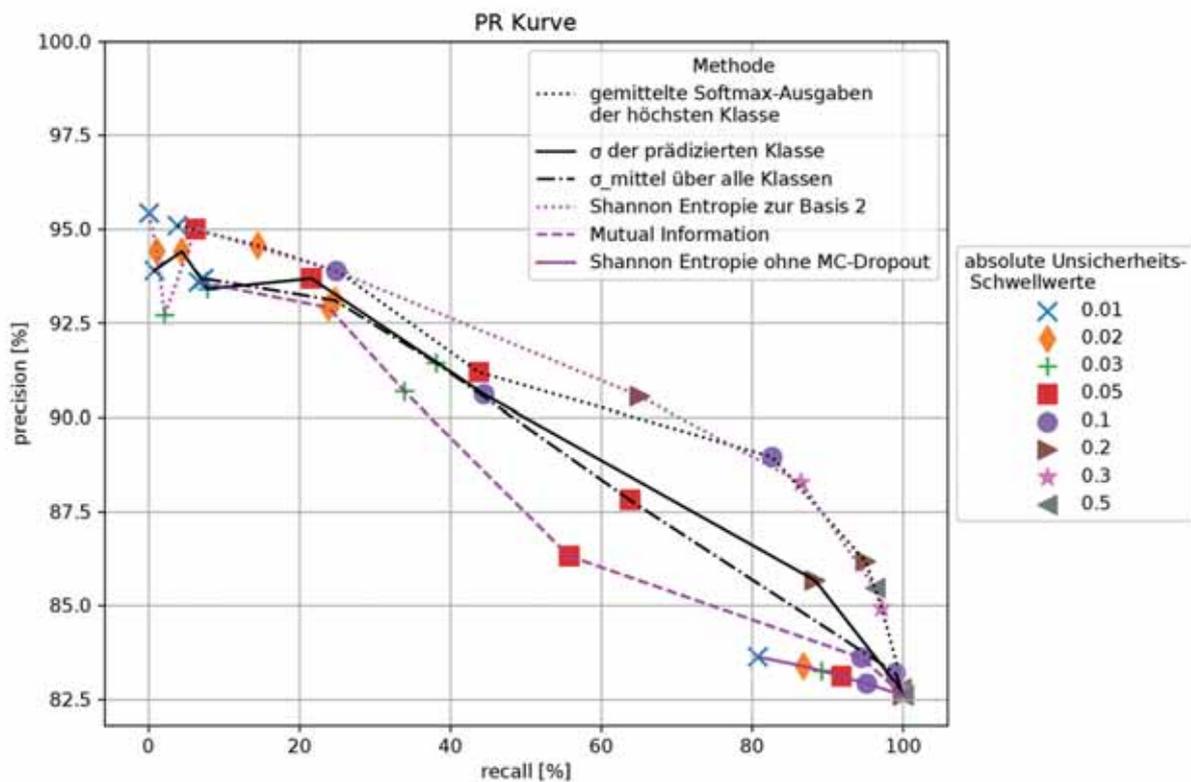


Abb. 5: Precision-Recall Kurven zum Vergleich verschiedener Unsicherheitsmaße mittels unterschiedlicher absoluter Schwellwerte

In der Praxis kann es praktischer sein, statt fester Unsicherheitsschwellwerte, relative Schwellwert anhand eines prozentualen Anteils an zu verwerfenden Pixeln festzulegen. Nach dem Herausfiltern von 10 % der Pixel mit den größten Entropiewerten, verbessert sich die Gesamtgenauigkeit von 82,6 % auf 86,5 %. Nach der Entfernung von 20 % liegt die Genauigkeit bereits bei 88,2 %. Die Ergebnisse solcher Filterungen sind in Abb. 6 & 7 zu sehen. Dabei kann unterschieden werden, ob ein Pixel richtig klassifiziert und gleichzeitig als unsicher identifiziert wurde (weiße Pixel) oder sowohl als falsch klassifiziert sowie zudem als unsicher (schwarze Pixel) anzusehen ist. In erster Linie werden die Pixel an den Objekträndern herausgefiltert. Der Schwellwert, nach dem die Pixel als sicher beziehungsweise unsicher kategorisiert werden, kann entsprechend den gewünschten Anforderungen festgelegt werden.

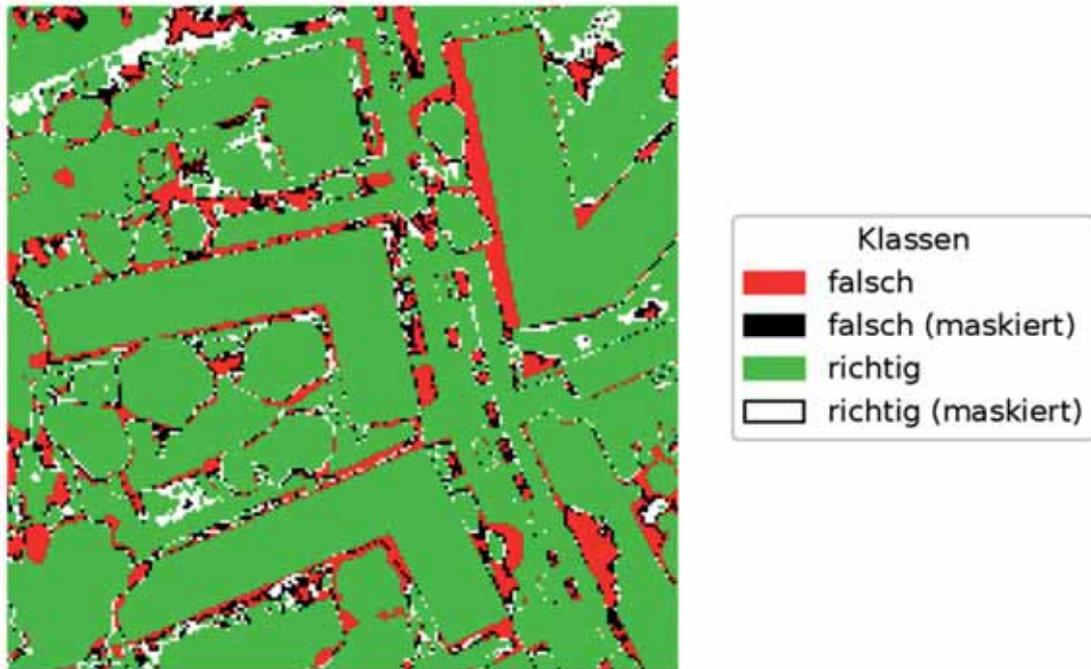


Abb. 6: Testbildausschnitt, bei dem die unsichersten 10 % der Pixel aus der Shannon Entropie maskiert wurden.

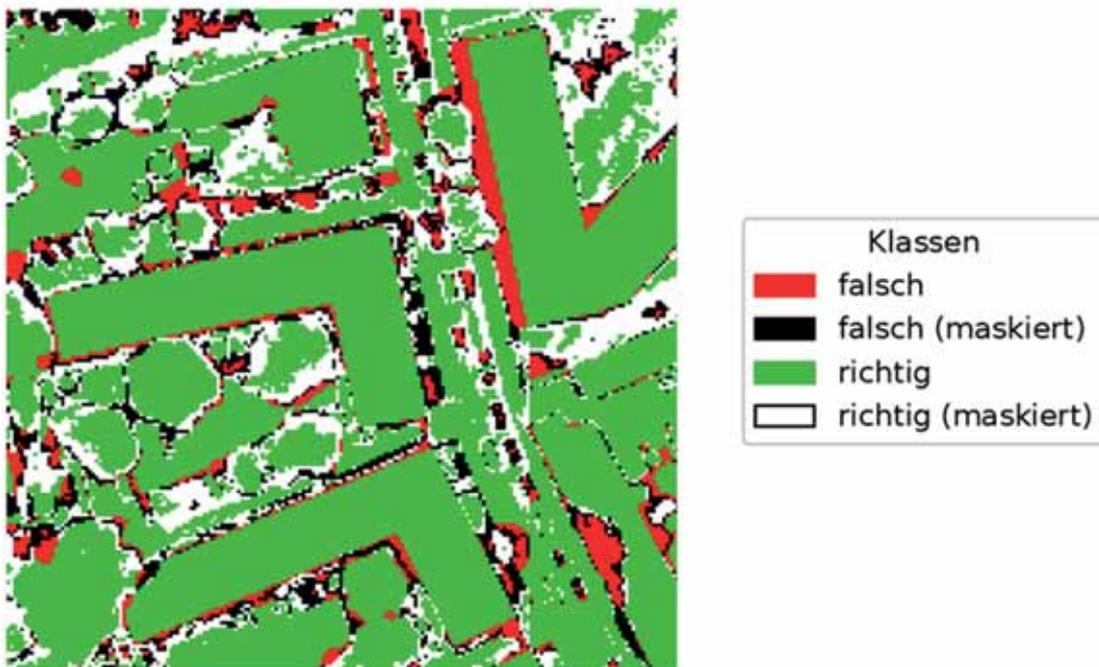


Abb. 7: Testbildausschnitt, bei dem die unsichersten 20 % der Pixel aus der Shannon Entropie maskiert wurden

## 5 Fazit & Ausblick

In dieser Arbeit konnte gezeigt werden, dass Monte-Carlo-Dropout robustere Maße für die Modellunsicherheit liefert als die Auswertung der Pseudo-Wahrscheinlichkeiten einer einzelnen Inferenz.

Die Precision-Recall Kurve aus Abb. 5 unterstützt den optischen Eindruck, dass die Shannon Entropie dabei das geeignetste Verfahren zur Bestimmung der Unsicherheit ist. Die Bereiche an den Objektgrenzen werden dabei als am unsichersten identifiziert. Dort sind auch viele der Fehlklassifikationen zu finden, was die Genauigkeitssteigerung durch das Herausfiltern von Pixeln mit besonders hohen Entropiewerten erlaubt. Je höher die Anforderung an die Precision ist, desto mehr Pixel müssen gegebenenfalls nachbearbeitet werden. Es ist allerdings mit diesem Ansatz nicht möglich Fehlklassifikationen zu erkennen, die sich aufgrund mangelnder Qualität der Eingabedaten auf ganze Objekte beziehen. Daher wäre es interessant ein gegebenenfalls vorhandenes Vorwissen zur Qualität der Daten mitberücksichtigen zu können.

Ein zusätzlicher Anwendungsbereich bildet das Active Learning. Mittels der zusätzlichen Informationen zu den Unsicherheiten jedes Pixels kann beispielsweise anstelle eines Random Forest wie bei KÖLLE et. al. auch ein neuronales Netz eingesetzt werden. Dabei können, durch gezieltes Labeln unsicherer, bisher nicht mit ground truth versehener Pixel, mit einem geringeren Aufwand neue und für die Klassifikation wertvolle Trainingsdaten geschaffen werden, anstatt größere Datenmengen vollständig manuell zu labeln.

## 6 Literaturverzeichnis

- GAL, Y., 2016: Uncertainty in Deep Learning. Dissertation. University of Cambridge.
- GAL, Y. & GHAHRAMANI, Z., 2015: Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. CoRR (arXiv: 1506.02142).
- GERKE, M., ROTTENSTEINER, F., WEGNER, J. & SOHN, G., 2014: ISPRS Semantic Labeling Contest. PCV - Photogrammetric Computer Vision.
- KAMPFMEYER, M. SALBERG, A. & JENSSEN, R., 2016: Semantic Segmentation of Small Objects and Modeling of Uncertainty in Urban Remote Sensing Images Using Deep Convolutional Neural Networks. 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 680-688.
- KENDALL, A. BADRINARAYANAN, V. & CIPOLLA, R., 2015: Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding. CoRR (arXiv: 1511.02680).
- KÖLLE, M., WALTER, V., SCHMOHL, S. & SÖRGEL, U., 2020: Evaluierung der Leistungsfähigkeit der Crowd für das Labeln von 3D-Punktwolken im Kontext von Active Learning. Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V., Band 29, 299-311.
- RONNEBERGER, O., FISCHER, P. & BROX, T., 2015: U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI), LNCS 9351, 234-241.

# Spatial Structure as an Element of Motivation in Location- Based Games

HEINRICH LOREI<sup>1</sup>, BERNHARD HÖFLE<sup>2</sup> & RENE WESTERHOLT<sup>3</sup>

*Abstract: Location-based games have emerged with the ubiquitous deployment of GPS-enabled devices. These games allow players to blend over the digital sphere with real-world whereabouts, implying that design choices made may affect the way players approach non-digital geographical spaces. This paper addresses scoring systems used to reward players for solving in-game tasks. We investigate how spatially structured scoring systems influence playing behaviours. The game we focus on is StreetComplete, an app that allows to collect and complement the OpenStreetMap database. We investigate indicators like walking speed, distance walked, and the game duration. Our results identify interesting effects spatial structure has on the game duration as well as the extent to which players are willing to explore an area geographically, hinting on interesting motivational elements of location-based games.*

## 1 Introduction

Location-based games are characterised by a conflation of physical and digital spaces. They allow players to enrich camera views with additional information, to switch attention between digital and physical environments, and to situate and contextualise themselves on a scale beyond the immediate context. Popular examples of location-based games include Pokémon Go, Ingress, and more traditional Geocaching applications. In some cases, the entertainment factor of such games is combined with scientific purposes like data collection, turning games into so-called serious games. This is the case with apps like Kort and StreetComplete that allow users to playfully collect OpenStreetMap data. This way, location-based games can be useful devices for the targeted collection of missing geodata, and to augment or update existing repositories. This paper focuses on the latter type of serious games, and on how the spatial arrangement of digital gaming elements influences the playing behaviour in a real-world setting.

The gaming element we investigate is scoring systems attached to in-game tasks a player may have to solve (e.g. answering questions about geographic features). Spatial structure refers to non-random geographic patterns like clustering or dispersion. We compare the influence of such spatially systematic deployments of scores to spatially random setups. The knowledge gained is not only important for research, but also for game designers and players themselves. Game designers get hints and detailed insights into the effects of spatial patterns in scoring systems. This knowledge can be used to design games more efficiently. For players of location-based games, our

---

<sup>1</sup> Metropolregion Rhein-Neckar GmbH, M1 4–5, D-68161 Mannheim, E-Mail: Heinrich.Lorei@m-r-n.com

<sup>2</sup> Heidelberg University, Institute of Geography, 3DGeo Research Group, Im Neuenheimer Feld 368, D-69120 Heidelberg, Germany, E-Mail: hoefle@uni-heidelberg.de

<sup>3</sup> TU Dortmund University, Faculty of Spatial Planning, Spatial Modelling Lab, August-Schmidt-Straße 10, D-44227 Dortmund, Germany, E-Mail: rene.westerholt@tu-dortmund.de

results can be useful, as appropriately designed games allow players to better explore and learn local areas, improve their physical activity and thus positively influence the gaming experience.

## 2 Motivation and Objective

Location-based games utilise real world geographies as playing fields. Their game experience is thus strongly influenced by contextual spatial and non-spatial motivational factors. The attractiveness of an area, for instance, can have an impact on how extensive players explore a playing field (SCHEIDER & KIEFER 2018; WEBER 2017; HARTEVELD 2011). Similarly, the integrability of a location-based game with everyday life can support gaining new perspectives on otherwise routine places and activities (MATYAS 2011; MÄYRÄ & LANKOSKI 2009). Another closely related aspect is the degree to which a game is adapted to local conditions. The strong alignment of gaming elements with local geographical features that players may be familiar with often improves the gaming experience through evoking a so-called ‘pride of place’ (SCHLIEDER 2014; WILL 2013; COLEMAN et al. 2009). The geography of a given playing field clearly influences the motivational ability of a location-based game.

In this study, we investigate the influence of spatial patterning on the playing experience of location-based games. The focus is on scoring systems used to reward players for completing individual tasks during the game, such as annotating geographical features with attribute information. It is known that location-based games are affected by the physical and mental capabilities of players like their physical endurance, spatial cognition, and navigational capabilities (SCHLIEDER 2014; WILL 2013; JACOB & COELHO 2011). Our underlying hypothesis is therefore that players may (consciously or unconsciously) be affected by systematic structure found in the spatial distribution of scores attached to in-game tasks. The game we use in our investigations is a modified version of the open access application StreetComplete, and the tasks we look at are attached to actual OpenStreetMap features.

## 3 Literature Review

Location-based games take place in real geographic environments. Therefore, the geography of an area becomes an integral part of such games, including the related contextual factors (SCHLIEDER et al. 2006). Several factors that influence game behaviour are of an ambient nature. For example, it has been found that the time of day is important not only because a player's attention varies with it, but also because other characteristics, such as the buzz of streets, are strongly correlated with time (CARRIGY et al. 2010). Similarly, the actual physical and perceived conditions during the game have an impact on the motivation of players. Weather conditions are one example, but stress factors such as noisy environments (KNÖLL et al. 2014) and traffic-related air pollution are also important in evaluating the gaming experience. Some contextual factors are directly related to the morphology of a playing field. There is evidence that complex urban street layouts are demotivating, as they not only make it difficult for players to find their way around, but also to understand a local morphology (OLIVEIRA 2016, BEDÖ 2017). For similar reasons, the availability of prominent landmarks makes a difference, too (BESTGEN et al. 2017; RICHTER & WINTER 2014). Players can use them to orient themselves in potentially complex urban areas. This has an

influence, consciously or unconsciously, on how comfortable players of location-based games feel with real-world playing fields.

Location-based games have the advantage that they can be played literally anywhere and 'on the go'. This makes it easy for players to integrate these types of games into their everyday lives. One aspect of this integration, which has proved to be particularly positive in terms of the geoliteracy of the players, is the re-experience of seemingly familiar places (WEBER 2017; MATYAS 2011; MÄRYÄ & LANKOSKI 2010). This gives the players a new perspective on their own everyday activity spaces, which in turn enriches their everyday life with new experiences that they would not have had without location-based games. This re-experiencing of familiar places is supported by games that offer a broad distribution of game elements over an area (WEBER 2017; FRÄNTI et al. 2017; SCHLIEDER 2014; Will 2013; CELINO et al. 2012). The wider the geographical distribution of the game elements, the stronger the positive influence a game can have on increasing the daily level of exercise, exploring an area and gaining new experiences. Furthermore, adapting the game elements to the local conditions of an area can further improve the game experience and thus motivate the players to play even more. Our research provides additional evidence for the importance of geography in location-based games. We shed light on the influence that a systematic geospatial pattern in scores can have on player behaviour.

## 4 Methodology

### 4.1 The game StreetComplete

The game we used for our research in this paper is StreetComplete. This game is developed by Tobias Zwick, a German software developer from Hamburg. StreetComplete is an open source application and therefore the code is freely available on Github<sup>1</sup>. The main goal of the app is to give also inexperienced OpenStreetMap users the opportunity to participate in the OpenStreetMap project. For this reason, the questions (called quests) that are asked during the game are relatively simple, such as dichotomous yes/no questions. New quest types are suggested by the wider OpenStreetMap and StreetComplete communities and are collected and discussed on Github. Once approved by the community, the quests are implemented by volunteers and automatically attached to current issues on the map, for example based on missing tags. StreetComplete is an open-ended game. This means that players do not work towards a specific goal, but rather altruistically collect or improve data. The only form of reward is a counter of already solved quests. This is an advantage for our study, as it ensures that none of our players (see Section 4.3) has previously used a more complex form of scoring in the context of StreetComplete. The fact that StreetComplete is open source has further allowed us to modify the game and tailor it to our needs. This has allowed us to introduce various forms of spatially structured scoring systems.

### 4.2 Spatially-Structured Scoring System

Our analysis is based on two trial groups of players. While the task locations were the same for both groups (uniformly distributed across the map to avoid visual clustering on the map), the scores attached to those locations differed with respect to their spatial patterning. One group of subjects played the game with a spatially randomised scoring system. That is, we generated scores on the range  $[0, 100]$  from a spatially autoregressive model with the spatial parameter adjusted to  $\rho=0.01$ .

Analogously, the spatially structured scores were generated from the same model but with  $\rho=0.99$  leading to a strongly spatially autocorrelated scoring system. Figure 1 visualises both outlined scoring systems and how the tasks are distributed across the investigation area. The scores generated were not made visible on the map interface of the game to avoid introducing visual confounding factors. This way, we have been able to isolate the effect of spatially structured scores and to compare two groups of players under different spatial scenarios.

### 4.3 Subjects and Playing Field

Both scoring systems were deployed under controlled conditions. Our subjects comprised 40 geography students. This choice may limit the scope of the results to a specific target population, but it homogenises with respect to demographics and educational level, as well as technical proficiency. Those 40 subjects were randomly assigned to the two trial groups (20 each) playing the two different spatial scoring systems. The playing field is comprised of an urban area of 1.3 km<sup>2</sup> in size, located close to the centre of the city of Heidelberg (Fig. 2). The size chosen is motivated by findings from a prior study recommending 1.5 km<sup>2</sup> to be optimal for playing times of 30 to 60 min (SCHLIEDER 2014), a duration we considered appropriate to test our hypothesis. The area is diverse on a small scale comprising quiet zones like backyards but also busy roads. This allowed us to diminish the effect of subjects being more likely to move to pleasant parts of the area only.

### 4.4 Indicators of the Playing Behaviour

We have tested a range of parameters of the players' behaviours. All of those are indicators of the players' engagement and how motivated players remain during the game. Our indicators assessed include playing time, distance walked (normalised by playing time), variety of road types explored, numbers of tasks per minute solved, standard deviational ellipses, and a detour factor (ratio of shortest and trajectory-derived path, see Fig. 3). We investigated these indicators for significant mean differences between the two trial groups. Following Shapiro-Wilk tests, mean testing was performed using the non-parametric Mann-Whitney U test to account for non-normality. The only approximately normal variable is the numbers of tasks, which we tested by means of a t-test.

## 5 Results

A range of parameters did not differ significantly between the two trial groups (Tab. 1), but we were able to disclose two systematic differences. The group exposed to spatially structured scores played the game significantly longer than the control group (39 min vs 30 min,  $p=0.02$ ). A second though slightly weaker result is that the group with spatially structured scores explored the area more extensively than the members of the control group did (detour factor of 5.02 vs 4.04,  $p=0.09$ , t-test). These results indicate that the ways in which scoring systems are laid out spatially could be an interesting way to tweak location-based games such that players remain motivated over time and space during the play.

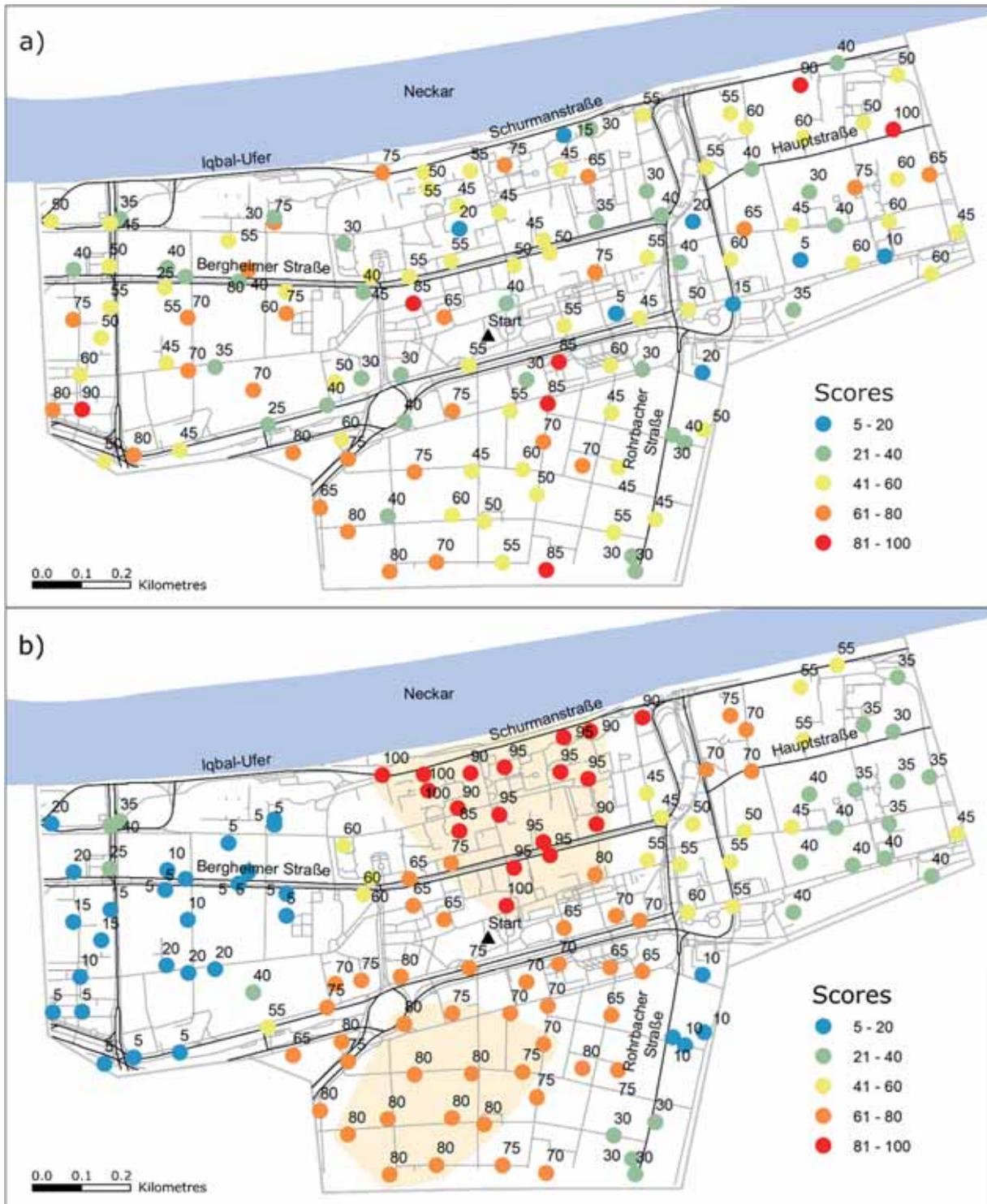


Fig. 1: Overview of the playing field. a) Endowed with a spatially random pattern of scores. b) Endowed with a spatially structured pattern of scores; clusters of the highest scores highlighted in orange. The background maps are based on OpenStreetMap data copyrighted by the OpenStreetMap contributors and available from <https://www.openstreetmap.org>.

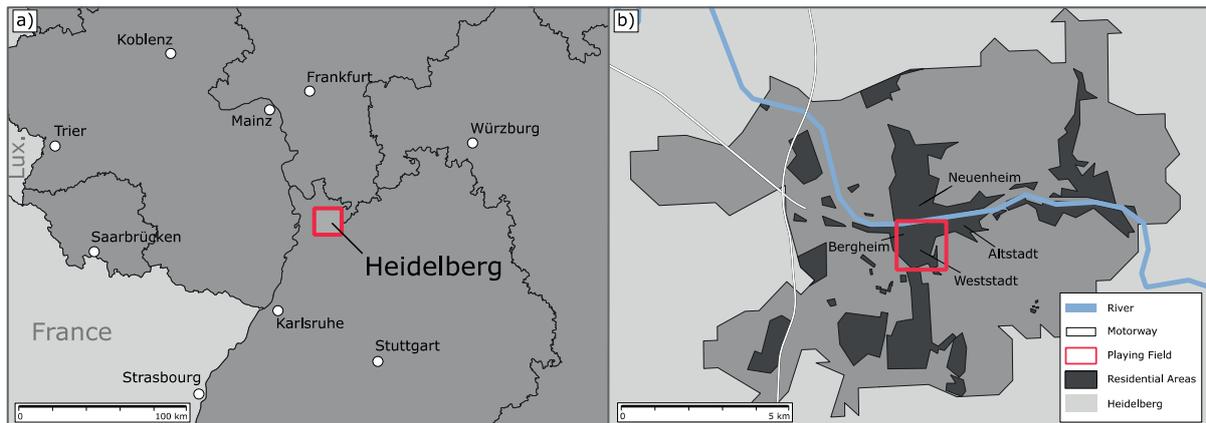


Fig. 2: Location and geographical context of the playing field. a) Location of Heidelberg in the wider context. b) Location of the playing field within the city of Heidelberg. The background maps are based on data from DIVA-GIS (HIJMANS et al. 2001) and OpenStreetMap data copyrighted by the OpenStreetMap contributors and available from <https://www.openstreetmap.org>.

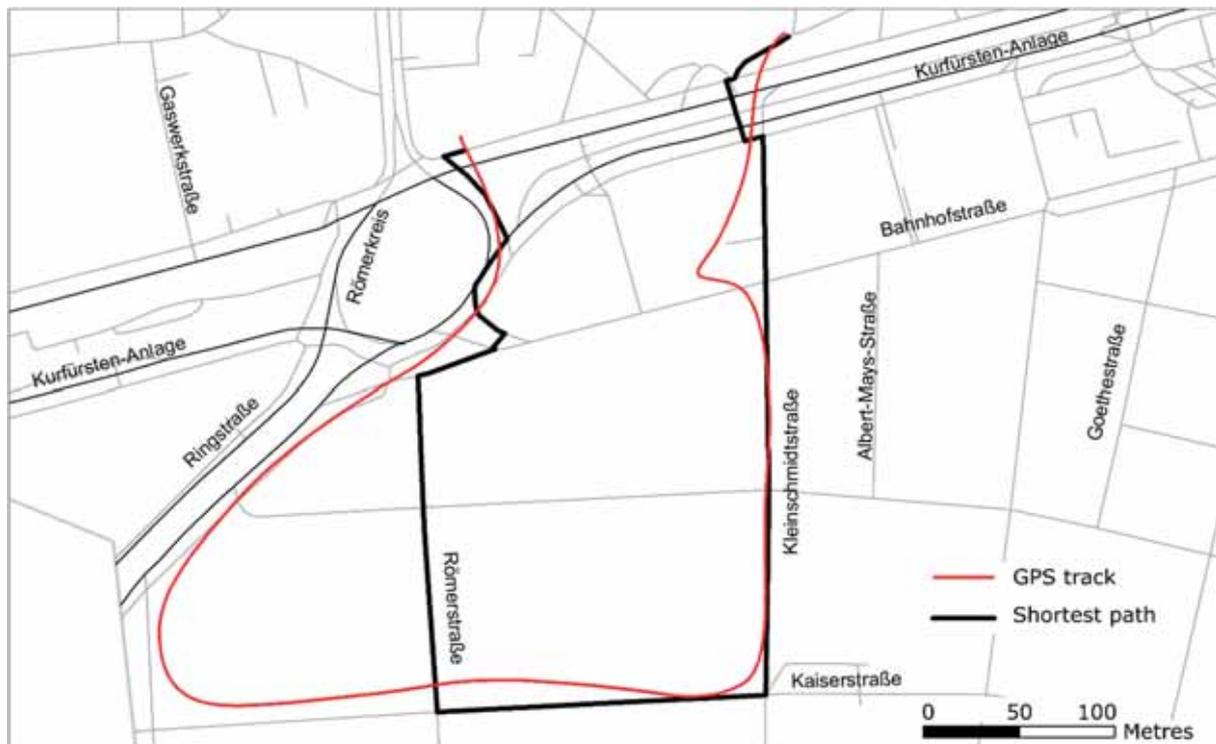


Fig. 3: Illustration of the detour factor as the ratio of the lengths of the GPS tracks and the shortest paths passing through all tasks visited. The background map is based on OpenStreetMap data copyrighted by the OpenStreetMap contributors and available from <https://www.openstreetmap.org>.

Interpreting our results in more detail reveals that some of them are not only statistically significant but also notable with respect to their effect strengths. The players exposed to spatially structured scores on average played the game longer than the control group. In addition, the same group added the whole length of a shortest path distance to their distance walked according to our

assessed detour factors. Considering also the other indicators that do not differ significantly in a statistical sense (which may be an effect of the limited numbers of participants in the groups) reveals that the players exposed to a spatially structured scoring system consistently solved more tasks, walked slower and longer distances on average, and traversed a higher diversity of different road types. The results obtained are thus highly indicative of an interesting relation between the players' motivation and the spatial layout of the game.

The parameter that differed most statistically and in terms of the absolute mean deviation is the duration of play. This is an important result because the game StreetComplete is unlimited in time and players could end the game at any time. The fact that the players from the experimental group played the game almost 33% longer than the players who were presented with the spatially random point system *ceteris paribus* is therefore a strong indication of a systematic influence of the spatial structure in the scores on the motivational aspects of the game. The random allocation of game versions to the players additionally supports this finding, since no obvious distortions result from the investigation structure. This result is significant beyond the case of location-based games. It supports the previous evidence that shows how important it is for players to be able to understand a game in order to be and keep motivated (LEE et al. 2017). Our results add a geographical dimension in the form of spatial structure.

Another very important parameter that was tested is the detour factor. As with the duration of the game, this parameter varied significantly and strongly between the two groups. While the longer game duration indicates a general tendency towards higher game motivation, the detour factor provides information about the motivation of the players not only to play the game but also to explore the field. Compared to the control group, the players in the trial group added the length of an entire "optimal" shortest path to their trajectories. This not only shows how important a comprehensible point system is for the effective design of location-based games, but also points to the relevance of geographically rewarding areas in general and beyond the present context, for example regarding quality of stay. Players from the experimental group have become more involved in the playing field, which, transferred to cases beyond playing, shows that they might also be more likely to explore areas if there are incentives to do so. The results obtained here are thus also of importance for urban planners and related researchers and practitioners.

Tab. 1: Mean values and their differences between the indicators calculated for the two trial groups. Statistical significance is flagged for confidence levels  $\alpha=0.10$  (\*) and  $\alpha=0.05$  (\*\*).

Parameter	Group with spatial pattern	Randomised control group	p-value
Playing time [min]	39.20	30.20	0.02**
Normalised distance walked [m]	2587	1954	0.46
Walking speed [m/s]	0.92	0.99	0.40
Area: standard deviational ellipse [km <sup>2</sup> ]	0.114	0.102	0.86
Variety of road types traversed [%]	79	72	0.29
Tasks solved [1/min]	17.80	12.55	0.38
Detour factor	5.02	4.04	0.09*

## 6 Conclusions

Based on our results obtained, we conclude that a spatially comprehensible layout is likely to support higher levels of motivation with location-based games. In this sense, our results support and add to prior results achieved in non-spatial settings demonstrating the importance of traceable scoring systems that players can make sense of (either consciously or subconsciously) (LEE et al. 2017). Future research should investigate other types of spatial structures in scoring systems to identify optimal layouts for game designs. Further, other gamification elements beyond scoring systems may be tested for spatial effects in similar ways. This way, it will be possible to optimise location-based games and to better utilise them for research purposes such as data collection. Also, research in the nexus of gamification and spatial analysis may contribute to the revealing of interesting, general psycho-geographic mechanisms.

### Notes

<sup>1</sup> <https://github.com/westnordost/StreetComplete>

## 7 References

- BEDÖ, V., 2017: Size and Shape of the Playing Field: Research Through Game Design Approach. Playable Cities, Nijholt, A. (ed.), Springer, Singapore, 67-86.
- BESTGEN, A. K., EDLER, D., KUCHINKE, L. & DICKMANN, F., 2017: Analyzing the effects of VGI-based landmarks on spatial memory and navigation performance. *KI-Künstliche Intelligenz* **31**(2), 179-183.
- CARRIGY, T., NALIUKA, K., PATERSON, N. & HAAHR, M., 2010: Design and evaluation of player experience of a location-based mobile game. *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, ACM, 92-101.
- CELINO, I., CERIZZA, D., CONTESSA, S., CORUBOLO, M., DELL'AGLIO, D., VALLE, E. D., FUMEO, S. & PICCININI, F., 2012: Urbanopoly: collection and quality assesment of geo-spatial linked data via a human computation game. *Proceedings of the 10th Semantic Web Challenge*, 148-163.
- COLEMAN, J.D., GEORGIADOU, Y. & LABONTE, J., 2009: Volunteered Geographic Information: The Nature and Motivation of Producers. *International Journal of Spatial Data Infrastructures Research* **4**(1), 332-358.
- FRÄNTI, P., MARIESCU-ISTODOR, R. & SENGUPTA, L., 2017: O-Mopsi: Mobile orienteering game for sightseeing, exercising, and education. *ACM Transactions on Multimedia Computing, Communications, and Applications* **13**(4), 56.
- HARTEVELD, C., 2011: *Triadic game design: Balancing reality, meaning and play*. Springer Science & Business Media, Heidelberg.
- HIJMANS, R. J., GUARINO, L., CRUZ, M. & ROJAS, E., 2001. Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. *Plant Genetic Resources Newsletter* **127**, 15-19.
- JACOB, J.T.P.N. & COELHO, A.F., 2011: Issues in the development of location-based games. *International Journal of Computer Games Technology*, **2011**, 495437.

- KNÖLL, M., NEUHEUSER, K., VOGT, J. & RUDOLPH-CLEFF, A., 2014: Einflussfaktoren der gebauten Umwelt auf wahrgenommene Aufenthaltsqualität während der Nutzung städtischer Räume. *Umweltpsychologie* **18**(2), 84-103.
- LEE, C.-I., CHEN, I.-P., HSIEH, C.-N. & LIAO, C.-N., 2017: Design aspects of scoring systems in game. *Art and Design Review* **5**(1), 26-43.
- MATYAS, S., 2011: Gemeinschaftliche qualitätsgesicherte Erhebung und semantische Integration von raumbezogenen Daten. University of Bamberg Press, Bamberg.
- MÄYRÄ, F. & LANKOSKI, P., 2010: Play in hybrid reality: Alternative approaches to game design. *Digital Cityscapes: Merging digital and urban playspaces*, De Souza e Silva, A. & Sutoko, D. (eds.), Peter Lang Publishers, New York, US, 129-147.
- OLIVEIRA, V., 2016: *Urban morphology: an introduction to the study of the physical form of cities*. Springer, Cham, Switzerland.
- RICHTER, K. F. & WINTER, S., 2014. Cognitive aspects: How people perceive, memorize, think and talk about landmarks. *Landmarks*, Richter, K. F., & Winter, S. (eds.), Springer, Cham, Switzerland, 41-108.
- SCHEIDER, S. & KIEFER, P., 2018: (Re-) Localization of location-based games. *Geogames and geoplay*, Ahlqvist, O. & Schlieder, C. (eds.), Springer, Cham, Switzerland, 131-159.
- SCHLIEDER, C., 2014: *Geogames Organizer's Guide v1.0*. Technical report, Universität Bamberg, Germany.
- SCHLIEDER, C., 2018: Geogames – Gestaltungsaufgaben und geoinformatische Lösungsansätze. *Informatik-Spektrum* **37**(6), 567-574.
- SCHLIEDER, C., KIEFER, P. & MATYAS, S. 2006. Geogames: Designing location-based games from classic board games. *IEEE Intelligent Systems* **21**(5), 40-46.
- WEBER, J., 2017: *Designing engaging experiences with location-based augmented reality games for urban tourism environments*. Dissertation, Bournemouth University, UK.
- WILL, C., 2013: *A Pattern Language for Designing Location-based Games*. Dissertation, RWTH Aachen, Germany.

# Evaluierung der Leistungsfähigkeit der Crowd für das Labeln von 3D-Punktwolken im Kontext von Active Learning

MICHAEL KÖLLE<sup>1</sup>, VOLKER WALTER<sup>1</sup>, STEFAN SCHMOHL<sup>1</sup> & UWE SÖRGEL<sup>1</sup>

*Zusammenfassung: Für die automatisierte semantische Interpretation von 3D-Punktwolken bedarf es typischerweise einer großen Menge gelabelter Trainingsdaten. Diese werden oft von Experten bereitgestellt. Da es sich hierbei allerdings um einen sehr zeit- und damit kostenintensiven Prozess handelt, wird in diesem Beitrag ein Ansatz zur signifikanten Reduktion des Labelaufwands mittels Active Learning (AL) in Kombination mit Crowdsourcing vorgestellt, welcher die Einbindung eines Experten in den Labelprozess gänzlich vermeidet. Die im Rahmen des Beitrags vorgestellte Pipeline wird auf den ISPRS Vaihingen 3D Semantic Labeling Benchmark Datensatz angewandt. Hierbei kann gezeigt werden, dass mit nur 0,4% durch die Crowd gelabelter Punkte sowohl der angewandte Random-Forest-Klassifikator als auch das Sparse 3D CNN eine Overall Accuracy (OA) erreicht, die sich im Vergleich zum Training mittels des vollständig gelabelten Datensatzes um weniger als 3 Prozentpunkte unterscheidet.*

## 1 Motivation

Gegenwärtig ist ein zunehmender Einsatz von Methoden des maschinellen Lernens, wie etwa Convolutional Neural Networks (CNNs), für verschiedene technische Anwendungen zu verzeichnen. Das Training solcher Netzwerke erfordert in der Regel allerdings enorme Mengen annotierter Daten, denen also semantische Bedeutung in Form eines sogenannten Labels zugeordnet worden ist. Im Fall von Geodaten, insbesondere im dreidimensionalen Raum, sind jedoch nur wenige solcher Datensätze vorhanden. Um entsprechende Klassifikationsansätze auch im Bereich der automatisierten Interpretation von Geodaten effizient nutzen zu können, gilt es, den Labelprozess zur Bildung des Trainingsdatensatzes für diese speziellen Daten zu optimieren.

Dies kann mittels Active Learning (AL) erreicht werden (KOVASHKA et al. 2016). In WORTMAN & VAUGHAN (2018) wird der AL-Prozess als hybrides Intelligenzsystem bezeichnet, da hier eine Maschine interaktiv mit einem Menschen zusammenarbeitet, um die Stärken beider Parteien zu kombinieren. Das Ziel ist es, aus der Fülle der bislang ungelabelten Daten nur jene Untermenge zu bearbeiten, die im Sinne einer besseren Trennung der Klassen besonders hilfreich ist. Dabei identifiziert die Maschine mit Hilfe eines Machine-Learning-Algorithmus informative Instanzen, welche dann vom Menschen gelabelt werden, wodurch die Leistungsfähigkeit der Maschine zunimmt. Daher sind solche „human-in-the-loop“ Systeme (BRANSON et al. 2010) ein effizientes Mittel zur signifikanten Reduktion des Labelaufwands. Dennoch müssen manuell Label bereitgestellt werden, was häufig durch Experten erfolgt. Da der Labelprozess sehr zeit- und damit kostenaufwendig ist, wächst zunehmend das Interesse, solche einfachen Labelaufgaben mittels Crowdsourcing zu lösen. Das bedeutet, der Labelprozess wird über das Internet an eine undefinierte Menge von Arbeitskräften ausgelagert.

---

<sup>1</sup> Universität Stuttgart, Institut für Photogrammetrie, Geschwister-Scholl-Str. 24D, D-70174 Stuttgart, E-Mail: [michael.koelle, volker.walter, stefan.schmohl, uwe.soergel]@ifp.uni-stuttgart.de

Mit eben dieser Methodik wurde auch bereits erfolgreich der *ImageNet*-Datensatz (DENG et al. 2009) abgeleitet. Dieser enthält über 14 Millionen annotierte 2D-Bilddaten, welche alltägliche Szenen darstellen und somit den Crowdworkern vertraut sind. Für das Labeln von 3D-Daten werden jedoch deutlich höhere Anforderungen an die Crowd gestellt, da die Interpretation von 3D-Daten, die auf einem 2-dimensionalen Bildschirm betrachtet werden, ein ausgeprägtes räumliches Vorstellungsvermögen erfordert. Die Arbeit von HERFORT et al. (2018) ist nach unserem Wissen die einzige, die sich mit 3D-Geodaten im Kontext mit Crowdsourcing auseinandersetzt.

In diesem Beitrag wird ein Ansatz zur (aus Sicht des Auftraggebers) vollautomatischen effizienten Auswahl und Annotation informativer Punkte aus einer gegebenen Punktwolke vorgestellt, wobei Crowdsourcing mit AL kombiniert wird. In dem Beitrag wird sowohl die Leistungsfähigkeit der Crowd bei der Interpretation von 3D-Daten untersucht als auch die Effizienz des AL-Prozesses aufgezeigt.

## 2 Methodik

### 2.1 Crowd-basierte Active Learning Pipeline

Die Grundidee dieser Arbeit basiert auf dem Prinzip des AL (SETTLES 2009), das auf eine Verringerung des Aufwands zur Generierung von annotierten Trainingsdaten abzielt. Dabei steht im Gegensatz zum passiven Lernen, bei dem einem Klassifikator ein bereits gelabelter Datensatz zur Verfügung gestellt wird, die Interaktion des Klassifikators mit einem Operateur während des Labeling-Prozesses im Fokus. Konkret selektiert der Klassifikator aktiv, ausgehend von einem ersten Trainingsdatensatz geringen Umfangs, welche Punkte aufgrund ihres hohen Informationsgehalts dem Trainingsdatensatz zuzuführen sind. Da diese selektierten Punkte manuell zu bearbeiten sind, ist die Maschine auf den Input eines Operateurs angewiesen. Um den kostspieligen Einsatz eines Experten nicht nur zu reduzieren, sondern gänzlich zu vermeiden, wird im Rahmen dieses Beitrags der Ansatz eines rein crowd-basierten Erfassungsprozesses verfolgt.

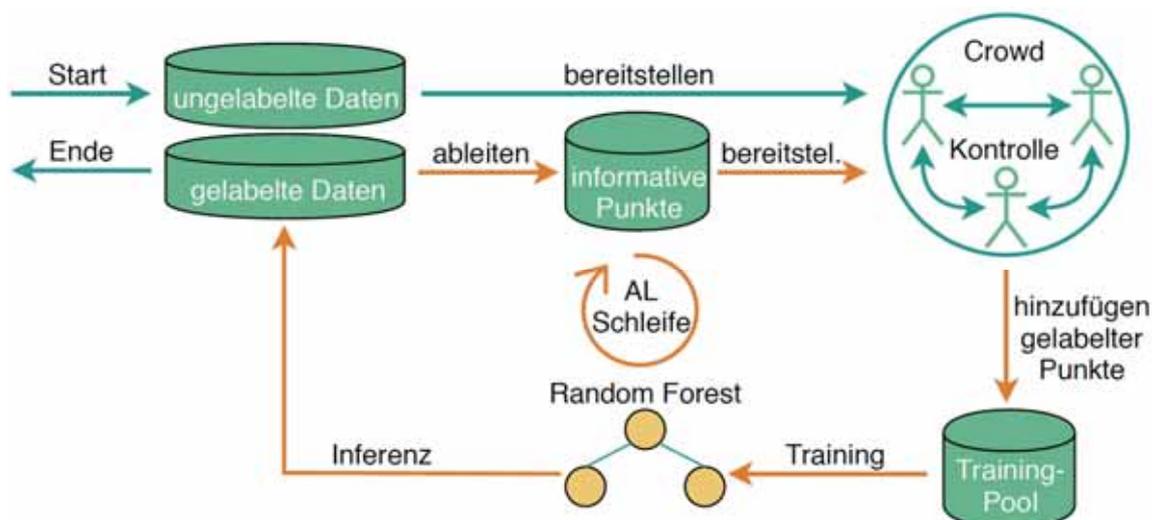


Abb. 1: Crowd-basierte AL Pipeline

Die Besonderheit liegt darin, dass, obwohl in diesem Prozess menschliche Interaktion notwendig ist, der gesamte Ablauf vollständig automatisierbar ist, da informative Instanzen automatisch durch einen Machine-Learning-Algorithmus detektiert werden. Das eigentliche Labeln durch einen Crowdworker ist aus Sicht des Auftraggebers vergleichbar mit einem Funktionsaufruf innerhalb eines Programms. Das Einstellen der Aufgabe auf einer Crowdsourcing-Plattform erfolgt automatisch unter Nutzung einer Programmierschnittstelle (API). Dasselbe gilt für die Übernahme der Ergebnisse und die Durchführung der Bezahlung. Dadurch lassen sich vollautomatisierte Prozesse realisieren, bei denen ein Teil der Aufgabe durch Menschen durchgeführt wird. Somit hat ein Auftraggeber lediglich eine Punktwolke, aus welcher der Trainingsdatensatz extrahiert werden soll, sowie das Budget zur Entlohnung der Crowdworker zur Verfügung zu stellen. Der grundsätzliche Ablaufprozess ist in Abb. 1 visualisiert.



Abb. 2: Implementiertes Webtool zur Angabe repräsentativer Vertreter der einzelnen Klassen in der voll-interaktionsfähigen Punktwolke

Die erste Aufgabe eines jeden Crowdworkers besteht darin, aus einer gegebenen Punktwolke für jede vorgegebene Klasse je einen repräsentativen Punkt zu labeln (siehe Abb. 2). Um ggf. falsch annotierte Punkte zu eliminieren, erfolgt im Rahmen einer zweiten Aufgabe eine unabhängige Kontrolle der erfassten Label. Das bedeutet, den Crowdworkern der zweiten Stufe werden jeweils selektierte 3D-Punkte präsentiert, wobei die Crowdworker zu entscheiden haben, ob diese Punkte

der korrekten Klasse zugeordnet sind. Die damit als Ergebnis erhaltenen Label tragen jedoch typischerweise einen eher geringen Informationsgehalt. Dies ist dadurch begründet, dass informative Punkte sowohl im Merkmalsraum als auch im Objektraum vornehmlich an den Entscheidungsgrenzen liegen (ERTEKIN et al. 2007). Crowdworker tendieren jedoch häufig dazu, von den Klassengrenzen weiter entfernte Punkte auszuwählen, da diese einfach und schnell einer Klasse zugeordnet werden können. Die damit erhaltenen wenigen gelabelten Punkte werden anschließend verwendet, um einen Klassifikator, in diesem Fall einen Random Forest (RF) (BREIMAN 2001), zu trainieren. Damit werden nun alle verbleibenden Punkte der Punktwolke automatisch klassifiziert, wodurch eine erste vollständige Klassifikation, allerdings von noch geringer Güte, gegeben ist. Das zentrale Element des AL-Prozesses ist, ausgehend vom Ergebnis des RF, diejenigen Punkte zu identifizieren, die sich bis dato nur mit geringer Zuverlässigkeit klassifizieren lassen. Hierbei handelt es sich um eben jene Punkte mit hohem Informationsgehalt, welche dem Trainingspool zugefügt werden sollen. Nach SETTLES (2009) kann zur Identifikation dieser Punkte zwischen *Uncertainty Sampling* und *Query-by-Committee* unterschieden werden. Beim *Uncertainty Sampling* basiert die Wahl der informativen Instanzen auf der prädizierten a-posteriori-Wahrscheinlichkeit  $P(x|c)$ , dass Punkt  $x$  zu Klasse  $c$  gehört, wobei die Unsicherheit mittels Entropie bestimmt wird. Das Prinzip *Query-by-Committee* dagegen basiert auf der Evaluierung des Widerspruchs einzelner Modelle eines Ensemble-Klassifikators. Die im Rahmen dieses Beitrags genutzte *Query Function* gehört der zweiten Klasse an und wurde von ARGAMON-ENGELSON & DAGAN (1999) als Vote Entropy ( $VE$ ) vorgestellt (Gleichung 1).

$$VE = - \sum_c \frac{\sum_e D(\mathbf{P}_e, c)}{N_E} \cdot \log \frac{\sum_e D(\mathbf{P}_e, c)}{N_E} \quad (1)$$

$$\text{wobei } D(a, c) = \begin{cases} 1, & \text{wenn } \operatorname{argmax}(a) = c \\ 0, & \text{sonst} \end{cases}$$

Dabei prädiziert jeder Klassifikator  $e$  des Ensembles  $E$  eine a-posteriori-Wahrscheinlichkeit für die Klassenzugehörigkeit einer jeden Punktinstanz, welche in  $\mathbf{P}_e$  eingeht. Basierend auf  $\mathbf{P}_e$  wird von jedem Modell des Ensembles eine Klasse favorisiert, sodass jedes Modell für eine bestimmte Klasse stimmt. Anschließend werden die Stimmen aller Modelle aufaddiert und mit der Anzahl der Mitglieder des Ensembles  $N_E$  normiert. Für die dadurch erhaltenen Verteilungen kann die Entropie nach SHANNON (1948) bestimmt werden.

Der wesentliche Vorteil bei der Verwendung von  $VE$  als Suchfunktion gegenüber der Berechnung der gemittelten Entropie aller Klassifikatoren des Ensembles liegt darin, dass Punkte, deren maximale a-posteriori-Wahrscheinlichkeit gering ausfällt, nicht notwendigerweise selektiert werden, solange die Ensemble-Mitglieder für dieselbe Klasse stimmen. Im Rahmen dieser Arbeit ist der Ensemble-Klassifikator durch den in KÖLLE et al. (2019) vorgestellten RF-Klassifikator gegeben, wobei sowohl geometrische als auch radiometrische Merkmale genutzt werden.

Auf Grundlage der bestimmten  $VE$  für jeden 3D-Punkt, werden dann diejenigen Punkte mit der höchsten  $VE$  ausgewählt und den Crowdworkern im Rahmen einer dritten Crowdaufgabe zum Labeln angeboten und per Mehrheitsentscheid überprüft. Nach entsprechender Ergänzung des Trainingsdatensatzes um diese nun gelabelten Punkte kann jeweils ein erneutes Trainieren des RF

und ein automatisches Klassifizieren der verbleibenden Punktwolke erfolgen. Es handelt sich also um einen iterativen Prozess, in welchem der Trainingsdatensatz schrittweise aus Punkten mit hohem Informationsgehalt aufgebaut wird mit dem Ziel, die Performance des Klassifikators inkrementell zu erhöhen. Die Iteration wird solange fortgesetzt, bis das Ergebnis der gewünschten Qualität entspricht oder das Label-Budget erschöpft ist.

## 2.2 Datengrundlage

Die in Kapitel 2.1 vorgestellte Pipeline wird auf den *ISPRS Vaihingen 3D Semantic Labeling Benchmark* Datensatz (V3D) (NIEMEYER et al. 2014) angewandt (Klassen: *Powerline, Low Vegetation, Impervious Surface, Car, Fence/Hedge, Roof, Façade, Shrub, Tree*). Diese mittels Airborne Laserscanning (ALS) erfasste Punktwolke bildet ein für westliche Länder typisches Vorstadtgebiet ab. Die Punktdichte dieses Datensatzes liegt bei 4-8 Punkten/m<sup>2</sup>. Um den Crowdworkern das Erkennen bekannter Objekte und Strukturen zu erleichtern, wird die Punktwolke zusätzlich mithilfe des in CRAMER (2010) abgeleiteten Orthophotos durch orthogonale Projektion koloriert (siehe Abb. 3). Da jedoch die Bilddaten (6. August 2008) und die ALS-Daten (21. August 2008) nicht zeitgleich erfasst wurden, ergeben sich insbesondere bei dynamischen Objekten wie Fahrzeugen Artefakte.



Abb. 3: Mittels Orthophoto kolorierter V3D-Trainingsdatensatz

## 3 Ergebnisse

Im Rahmen dieses Kapitels werden die Ergebnisse bei Anwendung der vorgestellten Pipeline auf den V3D-Datensatz diskutiert. Um abschätzen zu können, inwieweit das manuelle Labeln von 3D-Punkten crowd-basiert erfolgen kann und um generalisierbare Ergebnisse zu erhalten, ist es entscheidend, dass alle Experimente mit unabhängigen Crowdworkern durchgeführt werden, die nicht explizit instruiert werden. Auf eine solche unabhängige Crowd kann unter Nutzung der Plattform

*Microworkers* zurückgegriffen werden, welche in HIRTH et al. (2011) detailliert analysiert wird. Laut WEBLABCENTER INC. (2019) kann über diese Plattform auf ca. 1,4 Millionen Crowdworker zurückgegriffen werden, welche in verschiedenen Kategorien, wie zum Beispiel *Top Performers*, *All EU Workers* oder *Top EU Workers* organisiert sind.

Um eine Crowd-Kampagne auf *Microworkers* zu starten, müssen neben der exakten Aufgabenbeschreibung die anfallenden Gebühren übermittelt werden. Anschließend wird diese Kampagne seitens *Microworkers* freigeschaltet, sodass alle Crowdworker der spezifizierten Gruppe an dieser Kampagne mitarbeiten können. Im Rahmen der Anleitung auf der Plattform wird die URL zu den implementierten Webtools bereitgestellt. Nach Beendigung einer Aufgabe kann dem jeweiligen Crowdworker ein Code ausgestellt werden, welcher diesem als Nachweis zum Erhalt des Honorars dient.

Bei der Evaluation der Ergebnisse wird zum einen die erreichte Labelgenauigkeit der Crowdworker bei der manuellen Annotation von 3D-Punktwolken untersucht. Zum anderen wird die Performance des RF-Klassifikators über die gesamte Pipeline hinweg analysiert, wobei diese maßgeblich von der Labelgüte der Crowd bestimmt wird, da hier die Maschine, repräsentiert durch den RF, vom Menschen, genauer der Crowd, lernt.

### 3.1 Labelgenauigkeit der Crowd

Im Gesamten wurden 10 Crowd-Kampagnen (2 für die Initialisierung und 8 AL-Iterationsschritte), bestehend aus 1016 Aufgaben (100 freie Labelaufgaben, 100 Kontrollaufgaben und 102 Labelaufgaben pro Iterationsschritt) für das Labeln bzw. Kontrollieren von 3348 3D-Punkten (100·9 Klassen + Batchgröße von 306·8 Iterationsschritten) durchgeführt. Jede Aufgabe wurde an die *Microworkers*-Gruppe *Top Performers*, welcher mehr als 2000 aktive Worker angehören, gestellt und, je nach Aufgabentyp, mit 0,10\$ bis 0,15\$ vergütet.

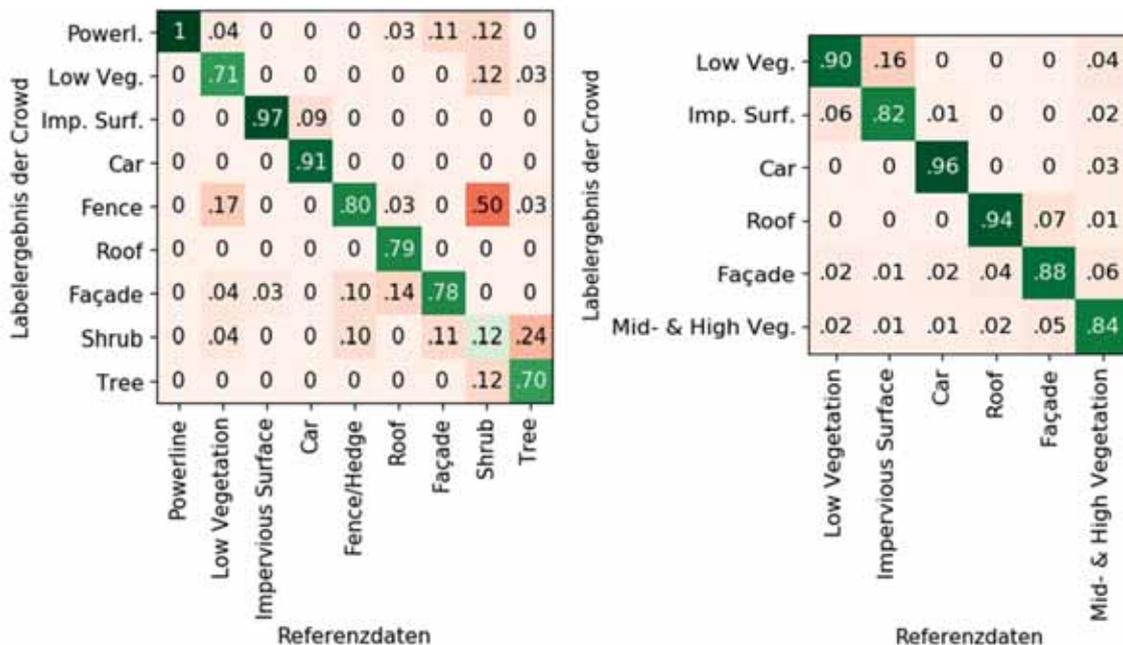


Abb. 4: Konfusionsmatrizen für die Labelgenauigkeit der Crowd für die Initialisierung (*links*) und über alle Iterationsschritte hinweg (*rechts*)

Die Analyse der Ergebnisse der von der Crowd im Rahmen der ersten Aufgabe frei gewählten und gelabelten Punkte ergibt erwartungsgemäß zunächst eine geringe Overall Accuracy (OA) von 56,44%, da hierbei die unkontrollierten Ergebnisse einzelner Crowdworker eingehen. Darüber hinaus ist zu berücksichtigen, dass im Zusammenhang mit dem im Rahmen dieses Beitrags angewandten bezahlten Crowdsourcings der hauptsächliche Anreiz der Worker im Honorar liegt und daher Crowdworker oft versuchen, entsprechende Aufgaben so schnell wie möglich abzuschließen, um ihren Gewinn zu maximieren (WORTMAN VAUGHAN 2018). Daher sind effiziente Kontrollmechanismen unerlässlich, um qualitativ hochwertige Ergebnisse zu erhalten. Mittels der in Kapitel 2.1 diskutierten Kontrollaufgabe, welche der Crowd gestellt wird, kann die OA bereits auf 71.79% erhöht werden. Die zugehörige Konfusionsmatrix ist in Abb. 4 (*links*) dargestellt.

Es zeigt sich, dass trotz der unabhängigen Kontrolle durch die zweite Gruppe von Crowdworkern insbesondere Vegetationsklassen (*Fence/Hedge, Shrub, Tree*) von den Crowdworkern verwechselt werden. Da die subjektiv beeinflusste definitionsabhängige Trennung dieser Klassen selbst für Experten eine Herausforderung darstellt, werden diese zur Klasse *Mid- & High Vegetation* fusioniert. Die fälschliche Zuordnung von Punkten zur Klasse *Powerline* ist durch den typischerweise geringen Anteil von Vertretern dieser Klasse in ALS-Punktwolken begründet. Daher werden von den Crowdworkern oft zufällig Punkte ausgewählt, um die Aufgabe möglichst schnell beenden zu können und die Bezahlung zu erhalten. Auch kann beobachtet werden, dass andere linienförmig angeordnete Punkte, wie sie etwa im V3D-Datensatz entlang einiger Fassaden auftreten (siehe Abb. 5, *rechts*), fälschlicherweise als *Powerline* gelabelt werden. Aufgrund der diskutierten Schwierigkeiten wird die Klasse *Powerline* für die weitere Auswertung mit der Klasse *Roof* vereinigt. Mittels dieser reduzierten Klassenanzahl kann bereits eine OA von 88,59% erreicht werden, sodass ein geeigneter initialer Datensatz gestellt werden kann. Dementsprechend liegt der Fokus der Pipeline auf Klassen, welche sich für eine Erfassung durch die Crowd als geeignet herausgestellt haben.



Abb. 5: Fehlklassifikationen durch die Crowdworker (gelb). Diese sind teils auf Verdeckungen (*links*) zurückzuführen, teilweise sind diese aber auch definitionsabhängig (*rechts*)

Werden die von der Crowd erhaltenen Label über alle Iterationsschritte hinweg betrachtet, wird eine sehr ähnliche OA von 88,87% erhalten (Konfusionsmatrix in Abb. 4, *rechts*). Beispielhaft werden für weiterhin auftretende Klassenverwechslungen in Abb. 5 zwei Szenarien dargestellt, anhand derer die fehlerhaften Zuordnungen erklärbar sind. Beide Beispiele zeigen, dass die in Kapitel 2.1 vorgestellte *Query Function* zielführend Punkte, die sowohl im Merkmalsraum als

auch im Objektraum jeweils auf der Entscheidungsgrenze liegen, selektiert. Jedoch stellt die korrekte Klassenzuweisung dieser Punkte für Crowdworker offensichtlich oft eine Herausforderung dar. Im ersten Fall wurde ein der Klasse *Impervious Surface* angehörender Punkt, welcher von der benachbarten Hecke leicht verdeckt wird, fälschlicherweise der Klasse *Low Vegetation* zugeordnet. Dies zeigt, dass für die Crowd solche Verdeckungen nur schwerlich zu interpretieren sind. Das zweite Beispiel veranschaulicht, dass Klassenzuordnungen oft interpretationsabhängig sind, was anhand der im Datensatz enthaltenen Wohnblöcke mit Flachdächern veranschaulicht werden kann. Punkte, welche direkt auf der Kante zwischen den Klassen *Roof* und *Facade* liegen, werden hier häufig, je nach eigener Interpretation, der jeweils anderen Klasse zugeordnet.

### 3.2 Zeit- und Kostenaufwand der Crowd-Kampagnen

Neben den von der Crowd erhaltenen Ergebnissen wird zusätzlich der Zeitaufwand für die Bearbeitung einer einzelnen Aufgabe sowie einer vollständigen Crowd-Kampagne diskutiert, wobei dieser nach WALTER & SÖRGEL (2018) von der Bezahlung abhängt. Anhand Tab. 1 wird bezüglich dieser beiden Aspekte eine Übersicht für die im Rahmen dieses Beitrags durchgeführten Crowd-Kampagnen gegeben.

Tab. 1: Zeit- und Kostenaufwand für die durchgeführten Crowd-Kampagnen.

Kampagne	mittlere Dauer/ Aufgabe [min]	Dauer/ Kam- pagne [h]	Bezahlung/ Auf- gabe [\$]
Freie Punktselektion	6,42	23,60	0,10
Kontrolle	4,70	17,20	0,12
Iteration Ø	2,53	14,80	0,15

Ausgehend von der mittleren Dauer, die für die Bearbeitung der einzelnen Aufgaben anfällt, zeigt sich, dass die Selektion beliebiger Punkte als Vertreter der jeweiligen Klasse erwartungsgemäß die meiste Zeit in Anspruch nimmt, da hierfür eine Navigation innerhalb der Punktwolke notwendig ist. Des Weiteren benötigen die Crowdworker mehr Zeit für das Kontrollieren gegebener Label als für das Labeln bestimmter Punkte innerhalb der AL-Iterationsschritte. Zwar erfordern beide Aufgaben eine eigenständige Interpretation der markierten Punkte, jedoch ist bei der Kontrollaufgabe zudem ein Vergleich mit der angegebenen Klasse erforderlich. In Bezug auf das Labeln informativer Instanzen im Rahmen der AL-Iteration ergibt sich für die einzelnen Aufgaben eine Standardabweichung von 0,2 min, was darauf hinweist, dass bezüglich des Zeitaufwands die tatsächliche Klassenzugehörigkeit der einzelnen Punkte keinen Einfluss hat. Im Mittel entsteht pro Iterationsschritt für das manuelle Labeln ein Zeitaufwand von 14,80h.

Im Rahmen der Crowd-Kampagnen wurden an die Crowdworker insgesamt 144,40\$ ( $100 \cdot 0,10\$ + 100 \cdot 0,12\$ + 102 \cdot 8 \cdot 0,15\$$ ) für das Bereitstellen der Label zum Aufbau des Trainingsdatensatzes ausbezahlt. Dabei ist zu berücksichtigen, dass die Anzahl der notwendigen Label mit zunehmender räumlicher Ausdehnung der Punktwolken nicht zwangsweise ansteigt. Zwar kann davon ausgegangen werden, dass eine Punktwolke, welche einen größeren räumlichen Bereich abdeckt, auch Punkte enthält, deren Merkmale sich von denjenigen, wie sie in einem enger begrenzten Gebiet auftreten, unterscheiden, sodass weitere informative Punkte im Datensatz enthalten sind. Jedoch

wird vor allem die Redundanz von Punkten mit ähnlichen Merkmalen erhöht. Beispielsweise entsteht bei der Klasse *Roof* ein zusätzlicher Labelaufwand lediglich für Punkte, welche auf Dachformen liegen, wie sie evtl. nur in einem größeren Gebiet auftreten. Jedoch überwiegen auch hier, analog zu einem kleineren Ausschnitt, insbesondere Flach- und Giebeldächer.

### 3.3 Performance der Active Learning Pipeline

Auf Grundlage der von der Crowd generierten Label (vgl. Kapitel 3.1) kann der RF-Klassifikator im Rahmen des AL-Prozesses pro Iterationsschritt neu trainiert (vgl. Kapitel 2.1) und dessen Ergebnis analysiert werden. Hierfür wird das bisher nicht verwendete Testgebiet des V3D-Datensatzes herangezogen. In Tab. 2 werden anhand der OA die Genauigkeiten für die Initialisierung der AL Pipeline sowie die 8 durchgeführten Iterationsschritte dargestellt.

Tab. 2: Erzielte OA des auf den Testdatensatz von V3D angewandten RF in Abhängigkeit des durch jeden Iterationsschritt erweiterten Trainingspools.

	Iterationsschritt								
	0	1	2	3	4	5	6	7	8
OA [%]	52,19	61,38	72,32	71,79	77,93	81,21	84,48	85,62	85,82

Das schrittweise Hinzunehmen informativer Punkte führt anfangs zu einem rapiden Anstieg der OA, welcher dann bis zu Iterationsschritt 8 mehr und mehr abflacht. Insgesamt konnte diese, beginnend bei einem geringen Wert von 52,19% nach der Initialisierung, durch das selektive Erweitern des Trainingsdatensatzes um informative Instanzen auf 85,82% gesteigert werden.

Um neben der OA, welche wesentlich durch die Prädiktionsgenauigkeit in Bezug auf stark repräsentierte Klassen bestimmt wird, auch die Fähigkeit des AL-Prozesses zur Selektion und daraus resultierender Prädiktionsgüte für unterrepräsentierte Klassen zu bewerten, wird in Abb. 6 die relative Klassenhäufigkeit aller Punkte des V3D-Trainingsdatensatzes (siehe Abb. 6 *links*) mit derjenigen der im Rahmen der AL Pipeline selektierten Punkte (siehe Abb. 6 *rechts*) verglichen. Es zeigt sich, dass trotz des signifikant geringeren Punktanteils der Klassen *Car* und *Façade* im V3D-Datensatz, mittels des vorgestellten Ansatzes nahezu eine Gleichverteilung der Klassen im schrittweise aufgebauten Trainingsdatensatz erreicht werden kann, wobei pro Iterationsschritt 306 Punkte (Anzahl experimentell bestimmt) ausgewählt werden. Die erreichte Gleichverteilung der Klassen ist darauf zurückzuführen, dass die *Query Function*, ausgehend vom initialen Trainingsdatensatz, jeweils aus dem gesamten V3D-Trainingsdatensatz diejenigen Punkte selektiert, deren

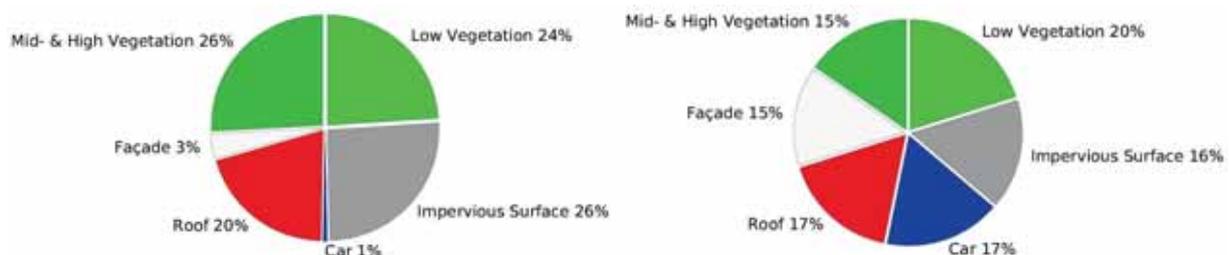


Abb. 6: Relative Klassenhäufigkeit aller Punkte des V3D Trainingsdatensatzes (*links*) und der im Rahmen der crowd-basierten AL Pipeline ausgewählten Punkte (*rechts*).

Merkmale für den Klassifikator bis dato unbekannt sind und damit den größten Widerspruch/Informationsgehalt aufweisen. Deshalb werden auch Punkte unterrepräsentierter Klassen in den Trainingsdatensatz aufgenommen, jedoch nur solange deren Merkmalsvektoren ausreichend unterschiedlich sind. Dadurch kann ein Quasi-Oversampling vermieden werden. Diese Möglichkeit, AL zur Erzeugung von Datensätzen nahezu gleichverteilter Klassenzugehörigkeit zu nutzen, wurde auch bereits von ATTENBERG & ERTEKIN (2013) beobachtet. Dadurch erübrigt sich die explizite Anpassung der Klassifikatoren an die jeweilige Klassenverteilung, beispielsweise über die Wahl einer Kostenfunktion (HE & GARCIA 2009), um weniger stark vertretene Klassen ausreichend berücksichtigen zu können. Dies stellt insbesondere für Anwendungen im Bereich der Fernerkundung einen entscheidenden Vorteil dar, da den meisten ALS-Punktwolken eine unausgewogene Klassenverteilung inhärent ist.

Um die Effizienz des vorgestellten Verfahrens aufzuzeigen, wird die Prädiktionsgenauigkeit des RF für die einzelnen Klassen bei Verwendung des in diesem Beitrag abgeleiteten Trainingsdatensatzes anhand der F1-Scores (GOUTTE & GAUSSIER 2005) jeweils derjenigen gegenübergestellt, welche sich bei Einsatz des gesamten V3D-Trainingsdatensatzes (reduzierte Klassenanzahl) ergibt (siehe Tab. 3). Es zeigt sich, dass der verwendete Ansatz nach nur 8 Iterationsschritten eine um weniger als 3 Prozentpunkte geringere OA hervorbringt.

Da der RF-Klassifikator bereits bei der Ableitung des Trainingsdatensatzes involviert war, wird zur Demonstration der allgemeinen Anwendbarkeit des erzeugten Trainingsdatensatzes zusätzlich gezeigt, dass dieser auch für das Training der typischerweise auf einen großen Trainingspool angewiesenen CNNs geeignet ist. Hierfür wird das in SCHMOHL & SÖRGEL (2019) genutzte 3D-Submanifold Sparse Convolutional Network (SSCN) anhand des abgeleiteten Trainingsdatensatzes trainiert. Tab. 3 zeigt auf, dass mit diesem Klassifikator im Vergleich zum RF sehr ähnliche Ergebnisse erzielt werden. Dadurch wird deutlich, dass ein kleinerer, mannigfaltig aufgebauter Trainingsdatensatz, welcher sich aus informativen Instanzen zusammensetzt, einem großen redundanten Trainingsdatensatz vorzuziehen ist.

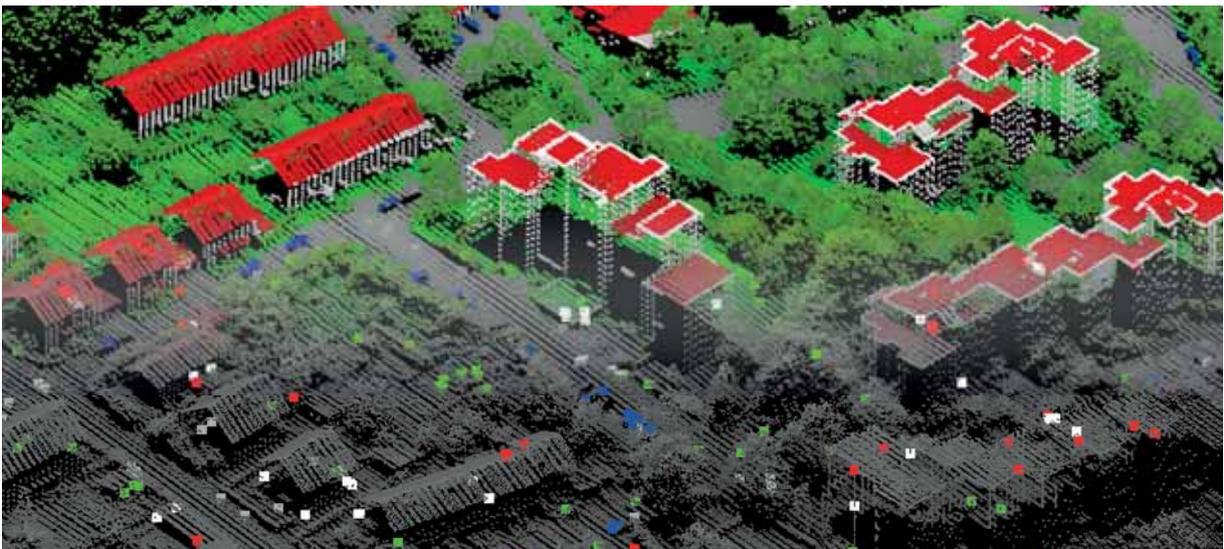


Abb. 7: Mittels der vorgestellten Pipeline kann die durch einen Experten vollständig gelabelte Trainingspunktwolke (*obere Bildhälfte*) durch das Labeln weniger informativer Punkte durch die Crowd ersetzt werden (*untere Bildhälfte*).

Hervorzuheben ist, dass diese Genauigkeiten ohne die Verwendung der im Rahmen des V3D-Datensatzes bereitgestellten Trainingsdaten erreicht wurden, indem lediglich 0,4% der Punkte des verfügbaren Trainingsdatensatzes durch die Crowd gelabelt wurden (siehe Abb. 7). Um dieses Klassifikationsergebnis zu erreichen, musste der Crowd lediglich die ungelabelte Trainingspunkt-wolke und ein finanzielles Budget von ca. 145\$ für die Entlohnung der Worker zur Verfügung gestellt werden.

Tab. 3: Vergleich der Klassifikationsergebnisse für den Testdatensatz von V3D bei Verwendung des V3D-Trainingsdatensatzes vs. Verwendung der durch die crowd-basierte AL Pipeline abgeleiteten Trainingsdaten.

Klassifikator	Trainings-datensatz	F1-Score [%]						OA [%]
		Low Ve- getation	Impervious Surface	Car	Roof	Façade	Mid-and High Vegetation	
RF	V3D	82,55	91,84	69,85	95,01	62,67	86,15	88,42
	CB-AL	79,91	86,69	68,23	93,86	61,57	85,77	85,82
SSCN	V3D	82,48	91,16	75,15	94,89	61,53	87,29	88,39
	CB-AL	80,00	88,14	75,20	91,08	57,34	84,72	85,43

#### 4 Zusammenfassung und Ausblick

Im Rahmen dieses Beitrags wurde gezeigt, dass ein kompakter, aus informativen Instanzen aufgebauter Trainingsdatensatz, mittels eines vollautomatisierbaren Frameworks abgeleitet werden kann, indem Crowdsourcing mit Machine-Learning-Techniken kombiniert wird. Dieser hybride Prozess umfasst die automatische Selektion eben dieser informativen Punkte, welche dann manuell durch Crowdworker gelabelt werden. Durch die Nutzung eines solchen Datensatzes für das Training der Klassifikatoren konnte der notwendige Labelaufwand, bei annähernd gleicher Prädiktionsgenauigkeit im Vergleich zum vollständig gelabelten V3D-Trainingsdatensatz, signifikant reduziert werden.

Des Weiteren haben die Experimente ergeben, dass die Crowd nicht in der Lage ist, beliebige Anforderungen eines Auftraggebers zu erfüllen. Insbesondere die subjektiv beeinflusste Trennung von Vegetationsklassen, welche selbst für Experten eine Herausforderung darstellt, ist für die Crowdworker nur schwer zu bewältigen. Während die Punktdichte des genutzten V3D- Datensatzes im Bereich der Landesvermessung häufig realisiert wird, kann die räumliche Auflösung in UAV-Befliegungen erheblich gesteigert werden (CRAMER et al. 2018). Es ist davon auszugehen, dass solche hochaufgelösten Punktwolken für Crowdworker deutlich einfacher zu labeln sind und die Möglichkeit bieten, detaillierte Strukturen, wie etwa Fassaden oder Dachinformationen, zu extrahieren.

## 5 Literaturverzeichnis

- ARGAMON-ENGELSON, S. & DAGAN, I., 1999: Committee-Based Sample Selection For Probabilistic Classifiers. *Artif. Intell. Res.* **11**, 335-360.
- ATTENBERG, J. & ERTEKIN, S., 2013: Class Imbalance and Active Learning. *Imbalanced Learning*, John Wiley & Sons Ltd., 101-149.
- BRANSON, S., WAH, C., SCHROFF, F., BABENKO, B., WELINDER, P., PERONA, P. & BELONGIE, S., 2010: Visual Recognition with Humans in the Loop. *ECCV 2010*, 438-451.
- BREIMAN, L., 2001: Random Forests. *Mach. Learn.* **45**(1), 5-32.
- CRAMER, M., 2010: The DGPF-Test on Digital Airborne Camera Evaluation – Overview and Test Design. *Photogrammetrie-Fernerkundung-Geoinformation 2010*(2), 73-82.
- CRAMER, M., HAALA, N., LAUPHEIMER, D., MANDLBURGER, G. & HAVEL, P., 2018: Ultra-High Precision UAV-Based LiDAR and Dense Image Matching. *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-1*, 115-120.
- DENG, J., DONG, W., SOCHER, R., LI, L.-J., LI, K. & LI, F. F., 2009: ImageNet: A Large-Scale Hierarchical Image Database. *CVPR 2009*, 248-255.
- ERTEKIN, S., HUANG, J., BOTTOU, L. & GILES, L., 2007: Learning on the Border: Active Learning in Imbalanced Data Classification. *Proc. 16th ACM Conf. inform. knowledge management*, 127-136.
- GOUTTE, C. & GAUSSIER, E., 2005: A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 345-359.
- HE, H. & GARCIA, E., 2009: Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering* **21**(9), 1263-1284.
- HERFORT, B., HÖFLE, B. & KLONNER, C., 2018: 3D micro-mapping: Towards assessing the quality of crowdsourcing to support 3D point cloud analysis. *ISPRS Jour. Photogramm. Remote Sens.* **137**, 73-83.
- HIRTH, M., HOßFELD, T. & TRAN-GIA, P., 2011: Anatomy of a Crowdsourcing Platform – Using the Example of Microworkers.com. *Proc. IMIS'11, IEEE Computer Society, Washington D.C., USA*, 322-329.
- KÖLLE, M., LAUPHEIMER, D. & HAALA, N., 2019: Klassifikation hochaufgelöster LiDAR- und MVS-Punktwolken zu Monitoringzwecken. *Dreiländertagung der DGPF, der OVG und der SGPF in Wien, Österreich – Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V., Band 28*, 692-701.
- KOVASHKA, A., RUSSAKOVSKY, O., FEI-FEI, L. & GRAUMAN, K., 2016: Crowdsourcing in Computer Vision. *Found. Trends Comput. Graph. Vis.* **10**(3), 177-243.
- NIEMEYER, J., ROTTENSTEINER, F. & SOERGEL, U., 2014: Contextual classification of LiDAR data and building object detection in urban areas. *ISPRS Jour. Photogramm. Remote Sens.*, **87**, 152-165.
- SCHMOHL, S. & SÖRGEL, U., 2019: Submanifold Sparse Convolutional Networks for Semantic Segmentation of Large-Scale ALS Point Clouds. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **IV-2/W5**, 77-84.

SETTLES, B., 2009: Active Learning Literature Survey. Computer Sciences Technical Report, 1648, University of Wisconsin-Madison.

SHANNON, C., 1948: A Mathematical Theory of Communication. The Bell System Technical Journal **27**(3), 379-423.

WALTER, V. & SÖRGEL, U., 2018: Implementation, Results, and Problems of Paid Crowd-Based Geospatial Data Collection. Journal of Photogrammetry, Remote Sensing and Geoinformation Science **86**(3-4), 187-197.

WEBLABCENTER INC., 2019: Microworkers. URL: <https://www.microworkers.com> (letzter Zugriff 20.10.2019).

WORTMAN VAUGHAN, J., 2018: Making Better Use of the Crowd: How Crowdsourcing Can Advance Machine Learning Research. Mach. Learn. Res. **18**(193), 1-46.

# Einfluss der zeitlichen Auflösung auf die raumzeitliche Segmentierung geomorphologischer Änderungsprozesse in 3D-Punktwolken

KATHARINA ANDERS<sup>1,2</sup>, LUKAS WINIWARTER<sup>1</sup>, HUBERT MARA<sup>2</sup>,  
RODERIK LINDENBERGH<sup>3</sup>, SANDER E. VOS<sup>4</sup> & BERNHARD HÖFLE<sup>1,2</sup>

*Zusammenfassung: Zeitserien von 3D-Punktwolken werden zunehmend für die Beobachtung geomorphologischer Phänomene genutzt. Dieser Beitrag untersucht, wie eine veränderte zeitliche Auflösung die raumzeitliche Abgrenzung von Oberflächenprozessen beeinflusst. Dazu wird eine Zeitserien-basierte Region Growing Segmentierung verwendet. Die Untersuchung erfolgt am Beispiel einer Sandbank an einem Strand in den Niederlanden, der über fünf Monate stündlich mit terrestrischem Laserscanning erfasst wurde. Die Ergebnisse zeigen, dass die zeitliche Prozessabgrenzung maßgeblich vom gewählten Zeitintervall abhängt. Auf Basis dieser Prozessabgrenzung kann die anschließende raumzeitliche Segmentierung auf ausgedünnten Zeitserien und somit mit reduziertem Berechnungsaufwand erfolgen.*

## 1 Motivation und Ziel

Geomorphologische Prozesse unterschiedlicher Magnitude, Häufigkeit und Geschwindigkeit charakterisieren die Topographie einer Landschaft. Für die Analyse von Oberflächenänderungen auf verschiedenen räumlichen und zeitlichen Skalen können Zeitserien von 3D-Punktwolken durch permanentes terrestrisches Laserscanning (TLS) erfasst werden (EITEL et al. 2016). Die zeitliche Auflösung der Datenaufnahme (z.B. stündlich, wöchentlich, jährlich) wird derzeit vor allem hinsichtlich der Bewegungsraten der erwarteten Prozesse gewählt (z.B. KROMER et al. 2017). Oftmals wird diese dabei hinsichtlich i) der praktischen Durchführbarkeit einzelner Messkampagnen oder ii) des minimalen zeitlichen Abstandes einzelner aufeinanderfolgender Aufnahmen höchstmöglich gesetzt. Anhand des Beispiels eines stündlich über mehrere Monate aufgenommenen Sandstrands wurde die Relevanz einer TLS-basierten Beobachtungsserie in hoher zeitlicher Auflösung und über lange Perioden gegenüber weniger, einzelner Aufnahmezeitpunkte in Bezug auf eine Änderungsanalyse gezeigt (ANDERS et al. 2019).

In diesem Beitrag untersuchen wir den Einfluss einer veränderten zeitlichen Auflösung auf die räumliche und zeitliche Abgrenzung eines Änderungsprozesses. Dazu verwenden wir eine neue Methode der raumzeitlichen Segmentierung (ANDERS et al. 2020) für das Beispiel einer Sandbank

---

<sup>1</sup> Ruprecht-Karls-Universität Heidelberg, 3DGeo Research Group, Im Neuenheimer Feld 368, D-69120 Heidelberg, E-Mail: [katharina.anders, lukas.winiwarter, hoefle]@uni-heidelberg.de

<sup>2</sup> Ruprecht-Karls-Universität Heidelberg, Interdisziplinäres Zentrum für Wissenschaftliches Rechnen, Im Neuenheimer Feld 205, D-69120 Heidelberg, E-Mail: hubert.mara@iwr.uni-heidelberg.de

<sup>3</sup> Delft University of Technology, Department of Geoscience & Remote Sensing, Niederlande, E-Mail: R.C.Lindenbergh@tudelft.nl

<sup>4</sup> Delft University of Technology, Department of Hydraulic Engineering, Niederlande, E-Mail: S.E.Vos@tudelft.nl

als typisches Änderungsobjekt an einem Strand. Dieser wurde in stündlicher Auflösung mit terrestrischen Laserscanning erfasst. Die zeitliche Auflösung der Daten wird für die Analyse systematisch reduziert.

## 2 Methode

### 2.1 Datengrundlage

Als Datengrundlage dient eine TLS-Zeitserie eines Sandstrands mit Datensätzen in stündlicher Auflösung, die über einen Zeitraum von fünf Monaten in Kijkduin, in den Niederlanden, aufgenommen wurde (VOS et al. 2017). Die Vorprozessierung der Daten erfolgte nach der Beschreibung in ANDERS et al. (2019). Die Oberflächenänderung wurde für jede Epoche als vertikale Punktwolkendistanz der Geländeoberfläche zur ersten Punktwolke der Zeitserie berechnet. Die Berechnung erfolgt für Punkte in einem regulären 2D-Gitter mit einem Abstand von 0.5 m. Damit ergibt sich als Eingangsdatensatz für die Änderungsanalyse ein 3D-Tensor, in dem jeder Zeitschnitt die Änderungswerte der Szene an allen räumlichen Positionen repräsentiert. Messungenauigkeiten werden durch Anwendung eines gleitenden Median in der temporalen Domäne mit einem Zeitfenster von einer Woche verringert.

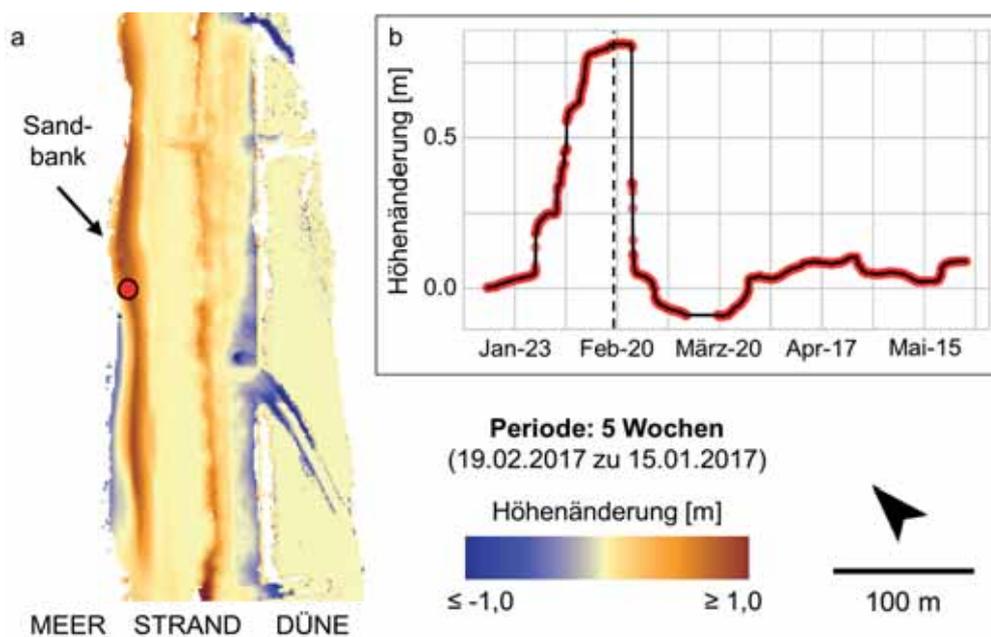


Abb. 1: (a) Höhenänderung des Sandstrands über eine Periode von fünf Wochen und (b) Verlauf der Höhenänderung an einer Position auf der Sandbank (roter Punkt in a). Die gestrichelte Linie markiert den Zeitpunkt der dargestellten Höhenänderungskarte. Höhenänderungen beziehen sich auf die erste Aufnahme der Zeitserie.

Als Beispiel wählen wir in diesem Beitrag eine Sandbank, die sich über mehrere Wochen am Strand bildet und dann abrupt zerstört wird. Abb. 1 zeigt die Höhenänderungen der Strandszene

zu einem Zeitpunkt maximaler Ausprägung der Sandbank und die komplette Zeitserie einer zentralen Position in der Sandbank. Diese Position wird als Ausgangspunkt für die raumzeitliche Segmentierung verwendet.

## 2.2 Raumzeitliche Segmentierung von Änderungsprozessen

Zur Detektion und räumlichen Abgrenzung verwenden wir eine Zeitserien-basierte Methode der Änderungsanalyse von Oberflächenprozessen (ANDERS et al. 2020). Die Methode berücksichtigt die komplette Historie der Oberflächenänderung durch eine *Region Growing* Segmentierung basierend auf der Ähnlichkeit benachbarter Zeitserien der Höhenänderungen. Dadurch ist im Unterschied zu bisherigen Methoden keine festgelegte Auswahl der Analyseperiode notwendig.

Die Methode detektiert Änderungen zunächst in der zeitlichen Domäne einer 2D-Position als Verschiebungen im Median der Werteverteilung in einem gleitenden Zeitfenster (TRUONG et al. 2019). Danach werden abgeschlossene Prozesse, bei denen die ursprüngliche Oberflächenhöhe wiederhergestellt wurde, zeitlich abgegrenzt. Ausgehend von einem detektierten Änderungspunkt als Startzeitpunkt wird dazu das normalisierte Volumen der Änderungswerte entlang der Zeitserie maximiert (in Anlehnung an PILTZ et al. 2016). Die daraus abgeleitete Periode des zeitlichen Änderungsprozesses an der gewählten initialen Position dient als Ausgangspunkt für das *Region Growing*. *Dynamic Time Warping* (DTW)-Distanzen (BERNDT & CLIFFORD 1994) als Ähnlichkeitsmetrik zwischen Zeitserien werden als Homogenitätskriterium für die Segmentierung verwendet. Die Berechnung der DTW-Distanzen erfolgt mittels des *Fast DTW* (SALVADOR & CHAN 2007). Diese Weiterentwicklung reduziert den Berechnungsaufwand des ursprünglichen Algorithmus von quadratischer zu linearer Komplexität. Der Schwellenwert für die Segmentierung eines Objekts wird als Mittelwert der DTW-Distanzen zur Zeitserie an der initialen Position in einer Nachbarschaft von 10 x 10 m ermittelt (ANDERS et al. 2020).

## 2.3 Untersuchung des Einflusses der zeitlichen Auflösung auf die Objektsegmentierung

Der Einfluss der zeitlichen Auflösung auf die Segmentierung eines raumzeitlichen Objekts wird untersucht, indem der Datensatz in zunehmend größere Zeitintervalle ausgedünnt wird. Dazu wird nur jede  $n$ -te Szene des 3D-Tensor mit stündlichen Oberflächenänderungen verwendet, mit  $n$  als Vielfaches von 2 bis 168 h. Daraus ergeben sich insgesamt 84 Varianten einer reduzierten zeitlichen Auflösung. Als Zielmetrik unserer Untersuchung dienen die Dauer des detektierten Änderungsprozesses und die Fläche des segmentierten Objekts. Als Referenz wird das Ergebnis der stündlichen Auflösung verwendet, in der Annahme, dass der Prozess im höchsten zeitlichen Detailgrad am besten abgebildet ist.

## 3 Ergebnis

Die zeitliche Abgrenzung der Sandbank hängt davon ab, zu welchem Zeitpunkt der Beginn des Prozesses als Änderungspunkt detektiert wird. Bei einzelnen Zeitintervallen wird die Höhenzunahme der Sandbank nicht detektiert (Abb. 2a). In diesen Fällen wäre eine an das Zeitintervall

angepasste Parametrisierung der Detektion von Änderungspunkten erforderlich. Tendenziell erfolgt die Detektion des Prozessbeginns bei niedrigerer zeitlicher Auflösung an einem späteren Zeitpunkt in der Zeitserie. Dadurch wird die Gesamtdauer des Prozesses zunehmend unterschätzt.

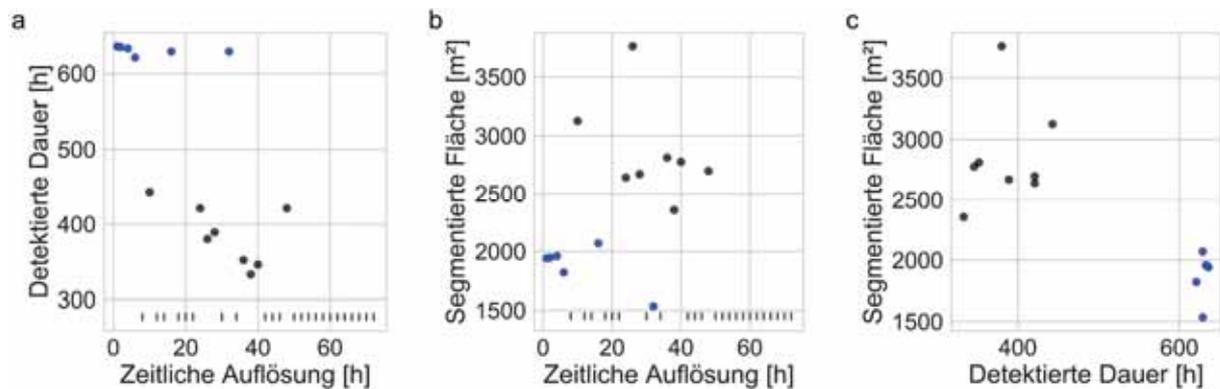


Abb. 2: Einfluss der zeitlichen Auflösung auf die (a) detektierte Dauer und (b) segmentierte Fläche einer Sandbank, sowie (c) Zusammenhang zwischen detektierter Dauer und segmentierter Fläche. Werte in blau haben eine hohe Übereinstimmung mit der Referenz. Strichmarken kennzeichnen zeitliche Auflösungen, bei denen keine Änderungspunkte detektiert wurden

Da die räumliche Auflösung der Daten und die initiale Position für die Segmentierung für alle zeitlichen Auflösungen gleich sind, ist die räumliche Abgrenzung in dieser Untersuchung einzig davon abhängig, wie der Prozess in den Zeitserien der Eingangsdaten, also dem 3D-Tensor des jeweiligen Intervalls, repräsentiert ist. Die Übereinstimmung der segmentierten Fläche mit der Referenz nimmt mit niedrigerer zeitlicher Auflösung nicht konstant ab (Abb. 2b). Vielmehr zeigt sich ein Zusammenhang mit der detektierten Dauer des Prozesses (Abb. 2c). Dies deutet darauf hin, dass die Übereinstimmung der raumzeitlichen Segmentierung mit der Referenz bei unterschiedlichen Auflösungen maßgeblich von der Prozessabgrenzung in der Zeitserie abhängt, nicht jedoch von der Dichte der Zeitserie. Je geringer demnach die Abdeckung der Prozessdauer, desto stärkere Abweichungen ergeben sich in der segmentierten Fläche des Änderungsobjekts zur Referenz.

## 4 Fazit & Ausblick

Für die raumzeitliche Analyse topographischer Zeitserien bedeutet das Ergebnis, dass die Prozessdetektion und -abgrenzung in der zeitlichen Domäne mit höchstmöglicher Auflösung erfolgen sollte. Für die anschließende raumzeitliche Segmentierung kann die Zeitserie in der detektierten Periode ausgedünnt werden. So lässt sich der Berechnungsaufwand reduzieren, welcher für die DTW-Distanz mit zunehmender Anzahl an Zeitpunkten linear zunimmt. Für die Interpretation der Änderungsprozesse steht dann wiederum die volle zeitliche Auflösung zur Verfügung. Hierbei ist das geeignete Aufnahmeintervall abhängig von der Bewegungsrate der zu beobachtenden Prozesse.

## 5 Danksagung

Dieser Forschungsbeitrag wurde teilweise von der Heidelberger Graduiertenschule HGS-MathComp finanziert, gegründet im Rahmen der Exzellenzinitiative (DFG: GSC 220). Die Datenaufnahme erfolgte im Projekt CoastScan und wurde vom Europäischen Forschungsrat finanziert (ERC Advanced grant 291206, NEMO).

## 6 Literaturverzeichnis

- ANDERS, K., LINDENBERGH, R. C., VOS, S. E., MARA, H., DE VRIES, S. & HÖFLE, B., 2019: High-Frequency 3D Geomorphic Observation Using Hourly Terrestrial Laser Scanning Data Of A Sandy Beach. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Science* **IV-2/W5**, 317-324, doi: 10.5194/isprs-annals-IV-2-W5-317-2019.
- ANDERS, K., WINIWARTER, L., RODERIK, R. C., WILLIAMS, J. G., VOS, S. E. & HÖFLE, B., 2020: 4D Objects-By-Change: Spatiotemporal Segmentation of Geomorphic Surface Change from LiDAR Time Series. *ISPRS Journal of Photogrammetry and Remote Sensing* **159**, 352-363, doi: 10.1016/j.isprsjprs.2019.11.025.
- BERNDT, D. J. & CLIFFORD, J., 1994: Using Dynamic Time Warping to Find Patterns in Time Series. *AAAI-94 Workshop on Knowledge Discovery in Databases* **10(16)**, 359-370.
- EITEL, J. U. H., HÖFLE, B., VIÉRLING, L. A., ABELLÁN, A., ASNER, G. P., DEEMS, J. S., GLENNIE, C. L., JOERG, P. C., LEWINTER, A. L., MAGNEY, T. S., MANDLBURGER, G., MORTON, D. C., MÜLLER, J. & VIÉRLING, K. T., 2016: Beyond 3-D: The new spectrum of lidar applications for earth and ecological sciences. *Remote Sensing of Environment* **186**, 372-392, doi: 10.1016/j.rse.2016.08.018.
- KROMER, R. A., ABELLÁN, A., HUTCHINSON, D. J., LATO, M., CHANUT, M.-A., DUBOIS, L. & JABOYEDOFF, M., 2017: Automated Terrestrial Laser Scanning with Near Real-Time Change Detection - Monitoring of the Séchillenne Landslide. *Earth Surface Dynamics* **5**, 293-310, doi: 10.5194/esurf-5-293-2017.
- PILTZ, B., BAYER, S. & POZNANSKA, A. M., 2016: Volume Based DTM Generation From Very High Resolution Photogrammetric DSMs. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **XLI-B3**, 83-90, doi: 10.5194/isprs-archives-XLI-B3-83-2016.
- SALVADOR, S. & CHAN, P., 2007: Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis* **11(5)**, 561-580, doi: 10.3233/IDA-2007-11508.
- TRUONG, C., OUDRE, L. & VAYATIS, N., 2020: Selective review of offline change point detection methods. *Signal Processing* **167**, doi: 10.1016/j.sigpro.2019.107299.
- VOS, S., LINDENBERGH, R. & DE VRIES, S., 2017: CoastScan: Continuous Monitoring of Coastal Change using Terrestrial Laser Scanning. *Proc. of Coastal Dynamics* **233**, 1518-1528.

# Assessing automatically-detected changes in the post-classification change-detection of Sentinel-2 data with Visual Analytics

MIKE SIPS<sup>1</sup>, VALERY LAVRENTIEV<sup>1</sup>, DANIEL EGGERT<sup>1</sup> & JULIA NEELMEIJER<sup>1</sup>

*Abstract: Satellite remote sensing offers the possibility to monitor the Earth's surface at high temporal and spatial resolutions. An important methodological field is the detection and interpretation of changes on the Earth's surface. A robust and widely utilized family of approaches is post-classification change-detection (PCCD). In our research, we address an important challenge to using PCCD from a user's perspective. Users often face difficulties finding changes in the result sets of PCCD that are relevant to their application scenarios. We propose a Visual Analytics approach that supports users in terms of exploring the temporal dynamics and the spatial distribution of automatically-detected changes generated via PCCD.*

## 1 Introduction

Satellite remote sensing offers the possibility of monitoring the Earth's surface at high temporal and spatial resolutions. Mission operators increasingly publish multispectral observations of the Earth's surface in freely accessible archives, e.g., the U.S. Geological Survey (U.S. GEOLOGICAL SURVEY 2020) or the Copernicus Open Access Hub (COPERNICUS OPEN ACCESS HUB 2020). The availability of multispectral images opens opportunities for users to develop novel approaches that help human communities to address challenges such as climate change, population growth, water scarcity, and the loss of biodiversity.

An important methodological field in this context is the detection and interpretation of changes on the Earth's surface. A robust and widely utilized family of approaches for detecting changes is post-classification change-detection (PCCD) (see e.g., MAS 1992). The input for PCCD is usually a set of multispectral images describing the Earth's surface at specific points in time. PCCD then utilizes these multispectral images to construct a classified map, e.g., land-cover maps, at every time step using supervised learning methods. The utilized supervised learning methods assign a categorical label (target label), e.g. land-cover labels such as forest, urban settlement, etc. to each pixel in a multispectral image to construct these change maps. To assign target labels to pixels, these methods need some data containing ground-truth target labels. PCCD approaches assume changes at pixels where transitions in target labels are observed across two or more time steps. The output of PCCD approaches is a set of sequences representing these transitions.

---

<sup>1</sup> German Research Center for GeoSciences GFZ, Telegrafenberg, D-14473 Potsdam, E-Mail: [sips, vlavr, daniel.eggert, neelmeijer]@gfz-potsdam.de

## 2 Challenge

An important challenge in using PCCD approaches is that the multispectral characteristic of a predicted target can vary from pixel to pixel, e.g., between two pixels representing a road in Berlin and Munich, or between time steps, e.g. a pixel can contain a multispectral profile of a road in winter time but can also represent a multispectral profile of a dense forest in the summer due to a dense canopy covering the road. We can attribute much of this variability in the multispectral characteristics of target labels to seasonal cycles or other long-term intra-annual changes. With the presence of the variability in multispectral characteristics, supervised learning methods often have difficulties in deriving classification models that can predict target labels precisely across space and time. Hence, the output for PCCD is usually a set of sequences representing automatically-detected changes. Users often face difficulties in finding the changes that a) describe meaningful changes, i.e. users can interpret some linear sequences as known processes at the Earth's surface, and b) are relevant for their application scenario. In our recent research project SEVA (SEVA 2020), we address this crucial challenge for PCCD. We utilize random forest classification in our Visual Analytics approach (BREIMAN 2001).

## 3 Visual Analytics Approach

Our Visual Analytics approach supports users in assessing sequences of transitioning target labels and deciding whether sequences are relevant for their application scenarios through interactive exploration. To decide whether sequences are relevant for their application scenarios, users need to investigate the following: a) occurrence of particular target labels within sequences in time, b) occurrence of specific transitions of target labels in sequences over time, c) distribution of target labels of particular sequences in space. To support users in their investigations of (a) - (c) through interactive exploration, we provide an overview and assessment component to users. Users can select time steps or target labels in the overview visualization for further inspection in the assessment component.

Our overview component aggregates the occurrence of target labels into a frequency distribution for each time step. We depict the corresponding frequency distribution in a bar plot (see Fig. 1(a)). The bar plot supports users in terms of understanding what target labels are involved in sequences and how frequently target labels occur at particular time steps (see (a) - occurrence of target labels in time). Thus, the bar plot supports users in selecting interesting time steps for further assessment. In addition, our approach depicts the sequence of target labels and their frequency of occurrence in an icicle plot (KRUSKAL & LANDWEHR 1983). In contrast to the original approach, we represent each time step as a circle (see Fig. 1(b)). In our icicle plot, the starting point of all sequences is the small circle in the middle of the visualization (bull's eye), and the surrounding individual rings depict the frequency of target labels in the consecutive time steps. We depict the frequency of the target labels proportional to the size available within each circle (see (b) - occurrence of specific transitions of target labels in time). Our icicle plot supports users in obtaining an overview of the major transitions and selecting potentially interesting target labels.

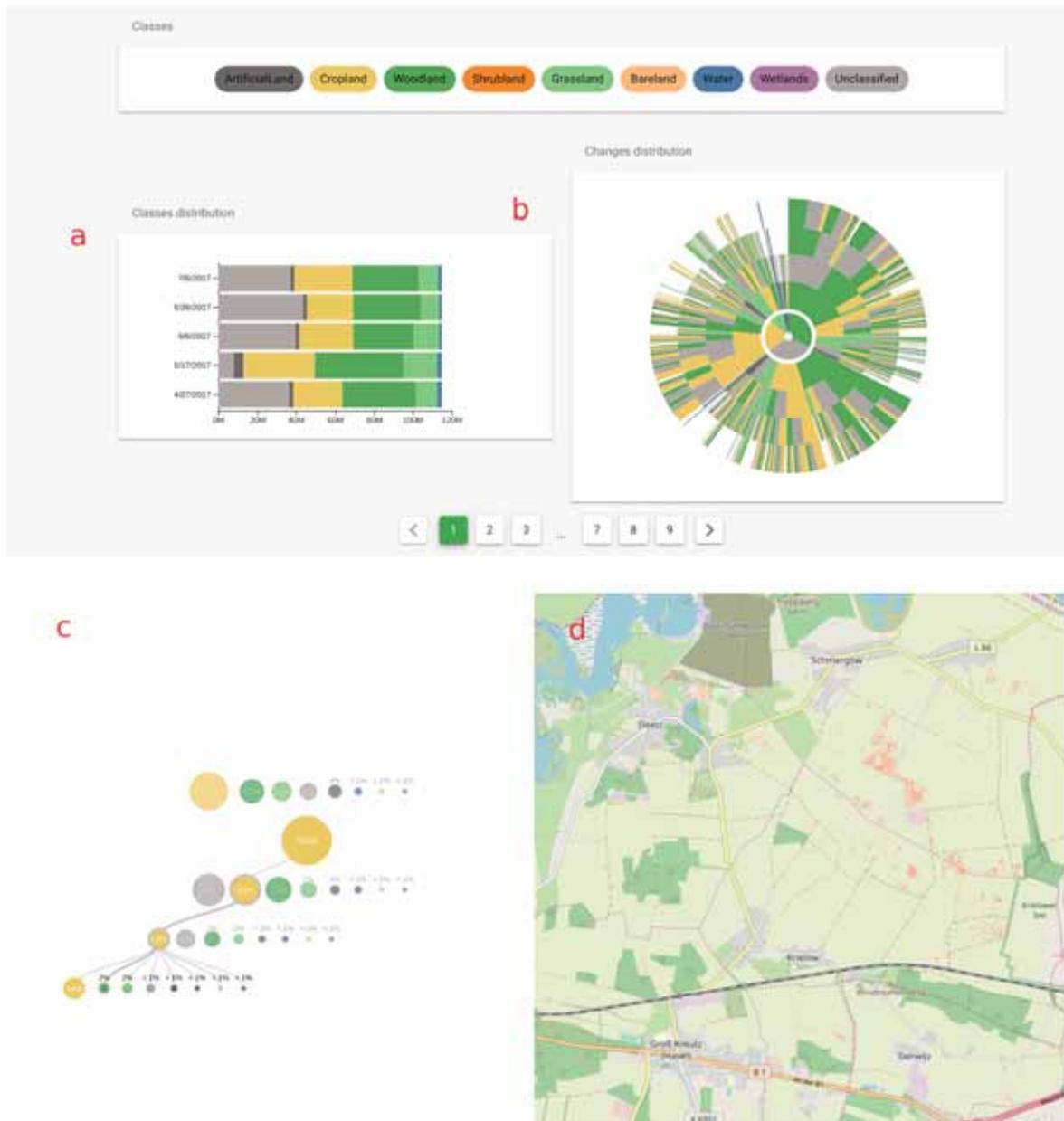


Fig. 1: Our Visual Analytics approach provides two main components for users: overview and assessment. The overview component depicts the frequency of target labels (called classes) for all time steps in a bar plot (see a) and the transition of target labels using an icicle plot (see b). The assessment component allows users to grow a sequence from a selected time step and target label into a tree visualization across time (see c). It also represents the spatial distribution of the depicted time steps and target labels of the tree in c) in the map visualization (see d).

Our assessment component depicts the selected time steps and transitions of target labels in detail. The default visualization represents the selected time step and target label as a single circle (see Fig. 1(c)). Users can explore transitions starting with the selected target label back transitions and occurrences between target labels. In addition, our assessment component depicts the

spatial distributions of all target labels depicted in the tree visualization (see Fig. 1(d) and (c) - distribution of target labels of particular sequences in space).

## 4 Preliminary User Feedback and Future Work

The developed approach has been presented to selected experts in remote sensing for feedback regarding the usability of the tool and the usefulness for performing efficient change detection analyses. These experts confirm that our Visual Analytics approach would allow users to explore potentially interesting change sequences. They further highlight the easy-to-perform analytical interaction method, the fast response to user interaction, and the ability to explore sequences in different views in their feedback to us. In addition, the experts predicted that the ability to interactively refine individual sequences would help to reduce their daily routine work effort. In the current version, users are able to use our approach for exploring land-cover changes in Central Europe generated by PCCD on Sentinel-2 multispectral images.

The experts also provided recommendations how the Visual Analytics approach should be further extended in future projects. The three most important recommendations were: a) adding the possibility to define individual areas of interest (e.g. in form of polygons) to restrict the spatial distribution of changes, b) enabling users to assess the accuracy of the classification step in the PCCD and c) allowing users to explore changes generated by other change detection methods. These extensions would open the door to a wide range of applications in e.g., agriculture, forest and water management; landscape, habitat and urban monitoring and planning; which are all especially of interest to authorities having a variety of reporting duties on land cover changes in general.

## 5 References

- BREIMAN, L., 2001: Random Forests. *Machine Learning* **45**, 5-32.
- COPERNICUS OPEN ACCESS HUB, 2020: Sentinel-2 multispectral observations. <https://scihub.copernicus.eu>.
- KRUSKAL, J. B. & LANDWEHR, J. M., 1983: Icicle Plots: Better Displays for Hierarchical Clustering. *The American Statistician* **37**, 162-168.
- MAS, J.-F., 1999: Monitoring land-cover changes: A comparison of change detection techniques. *Journal of Remote Sensing* **20**, 139-152.
- SEVA, 2020: Skalierbare Explorative Veränderungsanalyse für große Sentinel-2 Daten. <https://www.gfz-potsdam.de/sektion/fernerkundung-und-geoinformatik/projekte/seva/>
- U.S. GEOLOGICAL SURVEY, 2020: Landsat multispectral observations. <https://www.usgs.gov>.

# Genauigkeits- und Performanceuntersuchung eines neuen bathymetrischen UAV-Laserscanners

GOTTFRIED MANDLBURGER<sup>1,2</sup>, MARTIN PFENNIGBAUER<sup>3</sup> & ROLAND SCHWARZ<sup>3</sup>

*Zusammenfassung: Durch fortschreitende Sensorminiaturisierung sind mittlerweile bathymetrische Scanner mit einem Gesamtgewicht von etwa 10 kg verfügbar, welche auf wendigen bemannten Plattformen (Helikopter, Gyrokopter) oder auf unbemannten Flugsystemen betrieben werden können. Einer der ersten UAV-tauglichen topo-bathymetrischen Laserscanner ist der RIEGL VQ-840-G. In diesem Beitrag stellen wir das Sensorkonzept des Scanners vor und präsentieren eine erste, objektive Genauigkeits- und Performanceuntersuchung. Dazu wurde Ende August 2019 ein 500 m langer Abschnitt des Pielach Flusses sowie zwei angrenzende Schotterteiche erfasst. Die Datenerhebung erfolgte aus einer Flughöhe von 50-75 m mit Puls-wiederholraten von 50-100 kHz. Anhand terrestrischer Kontrollmessungen konnte eine absolute Höhengenaugigkeit von <10cm und eine Tiefenmessperformance von mehr als der 2-fachen Secchi-Tiefe nachgewiesen werden. UAV-basierte bathymetrische Laserscanner eignen sich daher für die präzise Erfassung kleinerer Gewässer und ermöglichen vielfältige Anwendungen wie etwa Gefahrenzonenplanung, Habitatmodellierung und Monitoring von Morphodynamik.*

## 1 Einleitung

Während sich UAV-basiertes Laserscanning für die Ableitung von sehr dichten und präzisen 3D Punktwolken im Topografiebereich bereits seit einigen Jahren etabliert hat (RESOP et al. 2019), sind entsprechend kompakte Scanner für bathymetrische Zwecke bis dato kaum verfügbar. Erste Schritte hin zur Erfassung von Gewässertiefen mittels LiDAR (Light Detection and Ranging) konnten in letzter Zeit durch die Einführung eines bathymetrischen Laserdistanzmessers mit konstanter Strahlachse beobachtet werden, welcher Gewässervermessung mittels Querprofilen erlaubt (MANDLBURGER et al. 2016). Eine effiziente und flächenhafte Erfassung erfordert aber ein scannendes System. Durch fortschreitende Sensorminiaturisierung sind mittlerweile auch bathymetrische Scanner mit einem Gesamtgewicht von etwa 10 kg verfügbar, welche auf wendigen bemannten Plattformen (Helikopter, Gyrokopter) oder auch auf unbemannten Flugsystemen (Unmanned Aerial Vehicle, UAV) betrieben werden können (ASTRALiTE 2020; FUGRO 2020).

Bei allen bathymetrischen LiDAR Sensoren wird Laserstrahlung im grünen Bereich des elektromagnetischen Spektrums eingesetzt (GUENTHER et al. 2000). Bei Laserstrahlung im sichtbaren grünen Bereich ist beim Betrieb des Lasers auf Augensicherheit zu achten. Aus diesem Grund weisen bathymetrischen Laserscanner meist eine relativ große Strahldivergenz von 1-5 mrad auf,

---

<sup>1</sup> TU Wien, Department für Geodäsie and Geoinformation, Wiedner Hauptstr. 8-10, A-1040 Wien, Österreich, E-Mail: gottfried.mandlbürger@geo.tuwien.ac.at

<sup>2</sup> Universität Stuttgart, Institut für Photogrammetrie, Geschwister-Scholl-Str. 24D, D-70174 Stuttgart, E-Mail: gottfried.mandlbürger@ifp.uni-stuttgart.de

<sup>3</sup> Riegl Research Forschungsgesellschaft mbH, A-3580 Horn, Österreich, E-Mail: [martin.pfennigbauer, roland.schwarz@riegl.com]

was bei typischen Flughöhen von 500 m über Grund einem Durchmesser des Abtastflecks am Boden von 0.5-2.5 m entspricht. Folglich ist die räumliche Auflösung auch bei modernen topo-bathymetrischen Sensoren mit hoher Pulswiederholrate und geringem Messpunktabstand durch den verhältnismäßig großen Laserfootprint beschränkt.

Ein wesentlicher Vorteil der Integration Bathymetrie-tauglicher, grüner Laserscanner auf UAV ist die geringere Messdistanz. Bei einer Flughöhe von 50 m und einer Strahldivergenz von 2 mrad beträgt der Durchmesser des Laserfootprints lediglich 10 cm. Dies bringt deutliche Vorteile für die Erfassung von Detailstrukturen wie etwa von Blöcken in Fließgewässern. Als Anwendungsbereiche sind detaillierte Überflutungssimulationen, Monitoring von Geschiebetransport, Modellierung aquatischer Habitate, etc. zu nennen. Da der Einsatz von Laserbathymetrie durch die Gewässertrübung eingeschränkt ist, eignet sich diese Technik bei Binnengewässern vor allem bei klaren, alpinen Flüssen. Diese weisen i.d.R. moderate Gewässerbreite und oftmals einen mäanderförmigen Verlauf auf, weshalb gerade für diese Anwendungsfälle der Einsatz UAV-basierter Messplattformen besonders vorteilhaft ist.

Einer der ersten UAV-tauglichen topo-bathymetrischen Laserscanner ist der *RIEGL VQ-840-G* (RIEGL 2020). Zur Überprüfung der Performance des Messsystems mit Hinblick auf erzielbare Eindringtiefe, räumliches Auflösungsvermögen und Genauigkeitspotenzial wurde im August 2019 eine Messkampagne an der Pielach, einem voralpinen Fluss in Niederösterreich, durchgeführt (MANDLBURGER et al. 2015; MANDLBURGER et al. 2016). Neben der UAV-basierten Datenerfassung erfolgte dabei auch die geodätische Einmessung von im Fluss und einem angrenzenden Schotterteich versenkten Zieltafeln mittels RTK-GNSS und Tachymetrie. Auf Basis dieser Referenzmessungen können objektive Aussagen zur Genauigkeit und Performance des Sensorsystems gemacht werden.

In diesem Beitrag stellen wir das Sensorkonzept und die Ergebnisse der Performanceuntersuchung vor. Der weitere Beitrag ist wie folgt strukturiert: In Abschnitt 2 werden das Sensorkonzept und die technischen Daten des kompakten Laserbathymetriescanners beschrieben. Abschnitt 3 behandelt das Untersuchungsgebiet und die zugrundeliegenden Auswertemethoden. Die Präsentation und Diskussion der Ergebnisse der Datenevaluierung erfolgt im Abschnitt 4. Abschnitt 5 beschließt den Beitrag mit einer Zusammenfassung der wesentlichen Erkenntnisse.

## 2 Sensorkonzept

Der VQ-840-G ist ein voll integriertes Laserscannersystem zur luftgestützten Erfassung von Topographie und Bathymetrie. Der Sensor besteht aus dem Laserscanner, einer integrierten Navigationseinheit, sowie optional einer RGB Kamera. Die Distanzmessung basiert auf dem Impulslaufzeitverfahren unter Verwendung von vergleichsweise kurzen Laserpulsen im sichtbaren, grünen Bereich des elektromagnetischen Spektrums ( $\lambda=532$  nm). Der Scanner weist eine benutzerdefinierbare Strahldivergenz von 1-6 mrad auf, was bei einer nominellen Flughöhe von 75 m einem Durchmesser des Abtastflecks am Boden von 7.5-45 cm entspricht. Für die 6 mrad Strahldivergenz beträgt die nominelle und erweiterte augensichere Distanz (NOHD/ENOHD) des Klasse 3B Lasers 15 m bzw. 75 m. Die Pulsrepetitionrate kann zwischen 50-200 kHz gewählt werden, was für UAV-basierten Betrieb einer typischen Punktdichte von mehr als 100 Punkten/m<sup>2</sup> entspricht. Das Scanmuster am Boden ist nahezu elliptisch (Palmer Scanner) mit einem Winkel zwischen

Laserstrahl und Nadir-Richtung von  $\pm 20^\circ$  quer zur Flugrichtung sowie  $\pm 14^\circ$  entlang der Trajektorie. Bei einer Flughöhe über Grund von 75 m führt dies zu einer Streifenbreite von 55 m. Laut Herstellerangabe beträgt die maximale Eindringtiefe bei einer Pulswiederholrate von 50 kHz die zweifache Secchi-Tiefe. Die absolute Genauigkeit ist mit 20 mm angegeben. Damit ist der Scanner vor allem für die Erfassung von mittelgroßen Flüssen mit klarem Wasser geeignet.

Neben der Strahldivergenz ist auch das Gesichtsfeld (engl. instantaneous Field of View, iFoV) des Empfängers veränderbar. Bei konventionellen LiDAR Sensoren entspricht iFoV i. d. R. etwa der 5-fachen Strahldivergenz, sodass die gesamte, diffus in Richtung des Sensors rückgestreute Energie vom Empfänger erfasst wird. In der Laserbathymetrie kommt es durch Volumenstreuung in der Wassersäule allerdings zu einer weiteren Aufweitung des Abtastflecks, welche nicht nur von der Strahldivergenz, sondern von den optischen Eigenschaften des Wassers (Trübe) abhängt. Bei einem relativ engen iFoV besteht daher die Gefahr, dass Anteile des rückgestreuten Gewässerbodensignals nicht vom Detektor erfasst werden. Durch die konfigurierbare Größe des iFoV besteht für den Anwender die Möglichkeit, zwischen besserer räumlicher Auflösung (kleines iFoV) und potenziell größerer Erfassungstiefe (großes iFoV) zu wählen.

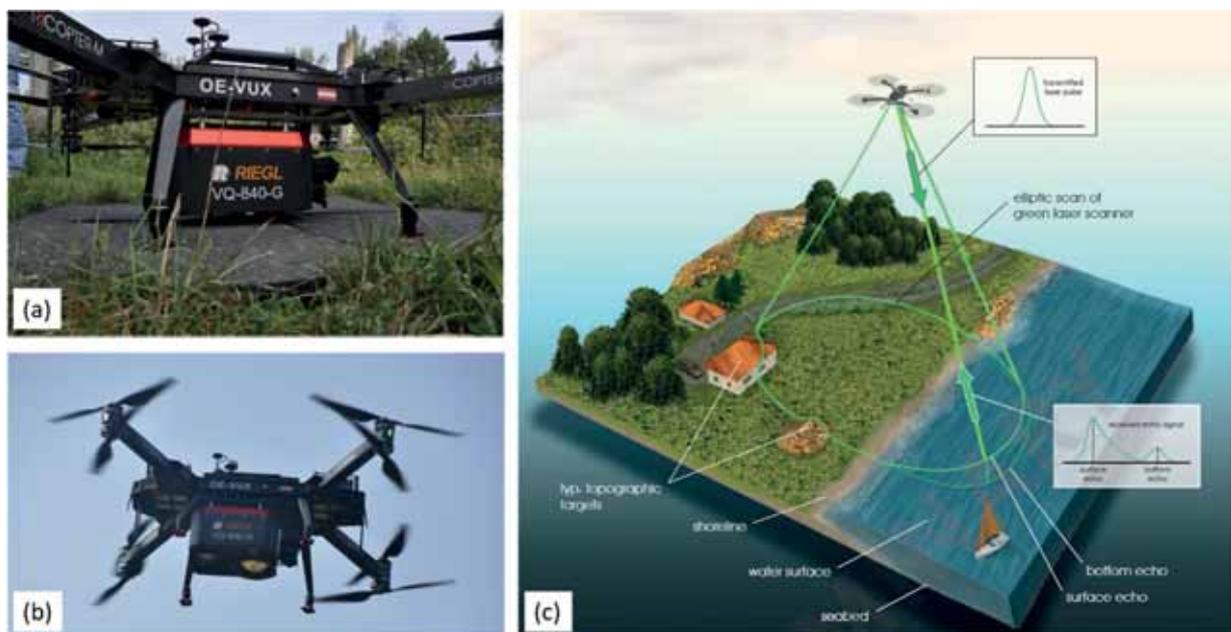


Abb. 1: UAV-basiertes topo-bathymetrisches Laserscansystem (RiCOPTER+VQ-840-G) vor dem Start am Boden (a) und in der Luft (b); (c) Prinzipskizze der UAV-basiert, topo-bathymetrische Datenerfassung

Abb. 1a und b zeigen das Sensorsystem montiert auf einer Oktokopter UAV Plattform (RiCOPTER-M), wobei bei der gezeigten Installation eine Sony 7R RGB Kamera neben dem topo-bathymetrischen Laserscanner zum Einsatz kam. Abb. 1c veranschaulicht in Form einer Prinzipskizze das elliptische Scanmuster (Palmer Scanner) sowie die simultane Erfassung von Topographie und Bathymetrie basierend auf Online-Wellenformprozessierung (PFENNIGBAUER et al. 2014). Die wesentlichen Sensorparameter sind in Tab. 1 zusammengefasst.

Tab. 1: Wesentlichen Sensorparameter des topo-bathymetrischen Scanners

Pulsrepetitionsrate (PRR)	≤ 200 kHz
Scanrate (Scankreisfrequenz)	10-100 Hz
Scanwinkel quer/längs zum Flugpfad	± 20° / ±14°
Strahldivergenz	1-6 mrad
Größe des Abtastflecks bei Flughöhe 50/100 m	5-30 cm /10-60 cm
Empfängergesichtsfeld (iFOV)	3-18 mrad
Größe (Länge x Breite x Höhe)	360 mm x 285 mm x 200 mm
Masse	12 kg
Leistungsaufnahme	160 W
Eindringtiefe in Vielfachen der Secchi-Tiefe	1.7@200 kHz, 2.0@50 kHz, 2.5@0.5 kHz
Präzision/Genauigkeit (Scanner)	15 mm / 20 mm

### 3 Untersuchungsgebiet und Auswertemethoden

#### 3.1 Datenerfassung

Zur Evaluierung der Genauigkeit und Tiefenmessperformance des bathymetrischen UAV-Laserscanners, wurde am 27.08.2019 eine Datenerfassung im Bereich der Neubacher Au, Niederösterreich, durchgeführt (siehe Abb. 2). Dabei wurden mehrere Grundwasser-gespeiste Teiche und ein Abschnitt des Pielach Flusses aus unterschiedlichen Flughöhen und mit variierenden Scanner-Einstellungen aufgenommen. Die Pielach ist ein vor-alpiner Schotterfluss mit einem mittleren Durchfluss von 7 m<sup>3</sup>/s. Das Untersuchungsgebiet ist Teil des Natur2000 Schutzgebietes „Niederösterreichische Voralpenflüsse“ und ist in MANDLBURGER et al. (2015) sowie MANDLBURGER et al. (2016) detailliert beschrieben.

Als Flugplattform diente ein Oktokopter mit maximalem Startgewicht von 35 kg (RICOPTER-M). Zeitgleich zu den Laserscans wurden RGB Nadir-Luftbilder mit einer Sony 7R Kamera aufgenommen. Zur Georeferenzierung des Bildblocks wurden Schachbretttafeln (Abb. 3a) installiert und mittels RTK GNSS eingemessen. Zur Überprüfung der maximal erreichbaren Messtiefe wurden vier 1x1 m<sup>2</sup> große Metallplatten in 1-3 m Tiefe in einem der Teiche versenkt (Secchi Tiefe: 1.35 m; Abb. 3b). Die Evaluierung der Unterwasser-Messgenauigkeit erfolgte anhand von 20 kreisförmigen Scheiben mit Schachbrettmuster, die in einer Tiefe von 0.5-2 m im Flussgrund entlang eines quer über den Fluss gespannten Seils verankert wurden (Abb. 3c und d). In Abb. 3d sind darüber hinaus zwei Satteldach-förmige Schalttafeln zu sehen, die zur absoluten Orientierung des Laserblocks eingesetzt wurden. Die Einmessung der Passpunkte, Passflächen und Kontrollpunkte erfolgte mit einer Leica TS16 Totalstation (Abb. 3e). Die beiden Detailuntersuchungsgebiete zur Überprüfung der Messgenauigkeit und der Tiefenmessperformance sind in Abb. 2 mit halbtransparenten weißen Rechtecken markiert.

Zum Zeitpunkt der Datenaufnahme war die Trübung in den erfassten Gewässern sehr hoch. Während die Trübung in den stehenden Gewässern durch das Aufwühlen der Schlammschicht am Boden der Teiche zurückzuführen ist, wurde diese im Flusslauf durch gelöste Sedimente infolge eines Starkregenereignisses in den Tagen vor der Messkampagne verursacht. Die Secchi-Tiefe betrug demnach in den Teichen 1.35 m und im Fluss ca. 1.0 m. Wie im Abschnitt 4 gezeigt werden wird, hat die Trübung nicht nur Auswirkungen auf die maximale Messtiefe, sondern auch auf die erzielbare

Messgenauigkeit. Während in einem Produktionsumfeld bei derartigem hydrologischen Zustand auf eine Datenerfassung mit optischen Methoden verzichtet werden würde, stellen die ungünstigen Bedingungen für das vorliegende Experiment gerade eine interessante Grundlage dar.

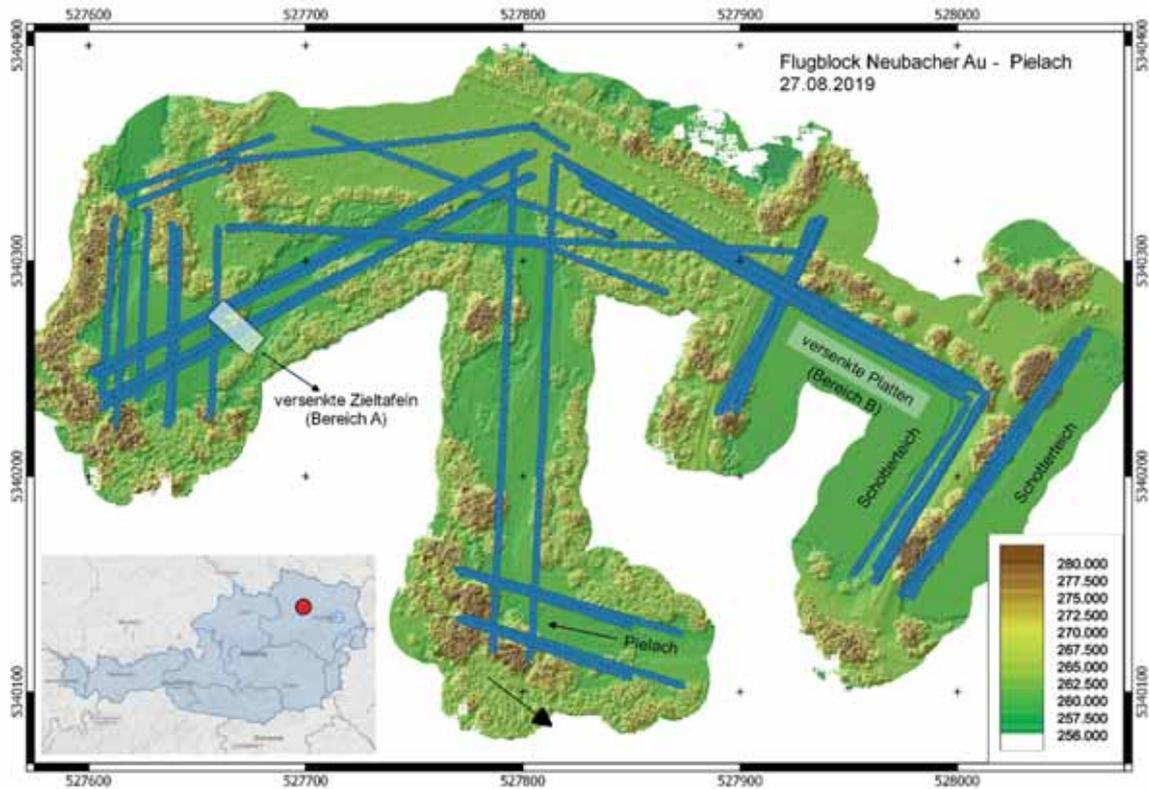


Abb. 2: Untersuchungsgebiet Neubacher Au – Pielach, Bildhintergrund: Laser Höhenmodell (Schummierung überlagert mit farbkodierter Höhenkarte), Flugpfade (blaue Kreise), weiße Rechtecke: Detailgebiete für Untersuchung der Messgenauigkeit (A) und Eindringtiefe (B); links unten: Übersichtskarte Österreich mit markierter Lage des Untersuchungsgebiets (Daten: openstreetmap und geoland.at)

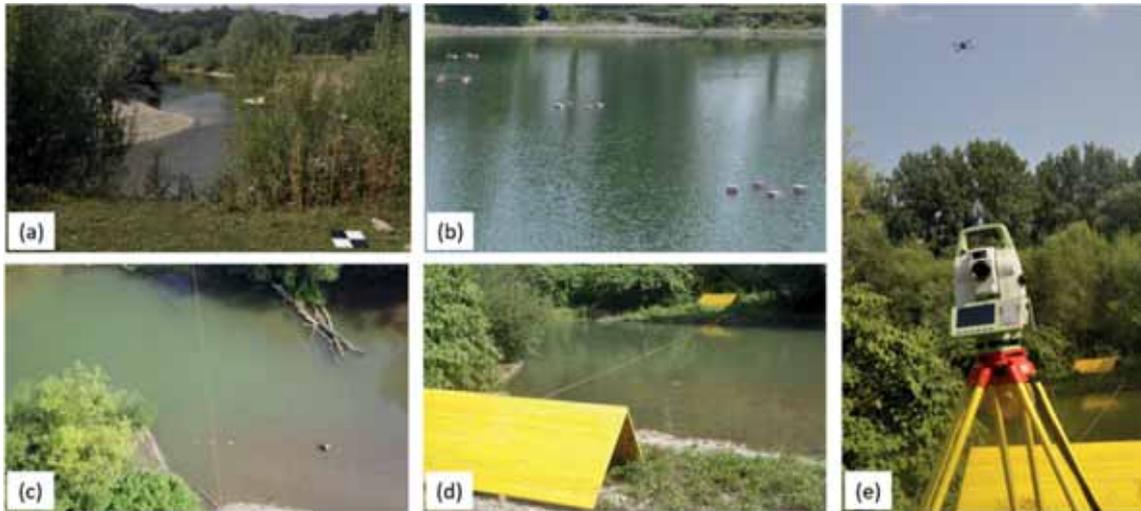


Abb. 3: Photodokumentation der Messkampagne Neubacher Au am 27.08.2019: (a) Pielach Flusslauf mit Passpunkt (Schachbretttafel) im Vordergrund, (b) Versenkte Metallplatten (Styropor Schwimmkörper) an der Wasseroberfläche sichtbar, (c) Detailgebiet für Genauigkeitsuntersuchung (UAV-Photo), versenkte kreisrunde Zieltafeln in drei Reihen entlang eines quer über den Fluss gespannten Seils angeordnet, (d) Satteldach Passflächen (an beiden) Flussufern zur absoluten Orientierung des Laserblocks, (e) Tachymeter für Eimessung der Referenzpunkte während der UAV-Datenerfassung

### 3.2 Auswertestrategie

Die Auswertung der UAV-basierten Laserdatenerfassung erfolgte weitgehend mit der wissenschaftlichen Laserscanning Software OPALS (PFEIFER et al. 2015) und umfasste folgende Schritte:

- Auswertung der RTK-GNSS-basierten Passpunktmessung und der tachymetrischen Aufnahme der Referenzpunkte (LEICA 2020).
- Bündelblockausgleichung aller aufgezeichneten Luftbilder und Berechnung eines digitalen Orthophotos in Pix4DMapper (PIX4D 2020). Absolute Orientierung des Bildblocks basierend auf photogrammetrischen Schachbrett Passpunkten.
- Post-processing der Flugtrajektorie (APPLANIX 2020) unter Zuhilfenahme der im Messgebiet stationierten permanenten GNSS Basisstation.
- Auswertung der Laserechos (Wellenformanalyse) und direkte Georeferenzierung der Messdaten (RiProcess).
- Feingeoreferenzierung der UAV-Flugstreifen mittels rigoroser Streifenausgleichung (GLIRA et al. 2015; GLIRA et al. 2016). Absolute Orientierung der Flugstreifen basierend auf Schachbrett Passpunkten und Passflächen
- Bestimmung der Wasseroberfläche: Näherungsweise Bestimmung der Wasserspiegellage über interaktives Editiertool in qpals (QGIS plugin für OPALS), Feinmodellierung durch Segmentierung der potenziellen Wasseroberflächenpunkte und robuste Interpolation eines digitalen Wasseroberflächenmodells (DWM).
- Refraktionskorrektur der rohen Laserscandaten durch Anbringung der Winkel- und Laufzeitkorrektur für Laserechos unter der Wasseroberfläche gemäß dem Brechungsgesetz von Snellius (MANDLBURGER et al. 2015).

Darauf aufbauend wurden Genauigkeit, Tiefenmessperformance und räumliche Auflösung wie folgt evaluiert:

- Ermittlung der relativen Streifenhöhendifferenzen zwischen den einzelnen Flugstreifen getrennt für Land- und Wasserbereiche anhand von streifenweisen digitalen Höhenmodellen (DHM) interpoliert als 0.1 m Raster aus der Laserpunktwolke (letzte Echos) mittels gleitender Schrägebeneninterpolation (RESSL et al. 2008).
- Ermittlung der absoluten Abweichungen zwischen Laserpunktwolke und Referenzpunkten an Land und Unterwasser. Dazu wurden in einer sphärischen Suchumgebung rund um jeden Referenzpunkt alle Laserpunkte extrahiert und eine best anpassende Ebene ermittelt. Die Abweichung wurde schließlich als Vorzeichen-behafteter Normalabstand zwischen Referenzpunkt und Laserpunkt-Ebene berechnet.
- Abschätzung der räumlichen Auflösung der Laserpunktwolke. Neben der theoretischen Abschätzung der räumlichen Auflösung basierend auf Strahldivergenz und Flughöhe sowie dem Messpunktstand erfolgt auch eine visuell-empirische Überprüfung anhand linearer Objekte (über den Fluss gespanntes Seil) und künstlicher rechter Winkel (Satteldach-Passfläche)
- Ermittlung der effektiven Eindringtiefe anhand der vier, in unterschiedlichen Tiefen versenkten  $1 \times 1 \text{ m}^2$  Platten. Dies erfolgt durch Segmentierung der 3D Punkt wolke, Selektion aller Laserechos auf den versenkten Platten und Bestimmung einer repräsentativen Höhe/Wassertiefe pro erfolgreich detektierter Messplatte und Gegenüberstellung mit den Referenztiefenmessungen bzw. der vor Ort gemessenen Sichttiefe.

## 4 Ergebnisse und Diskussion

Im Folgenden werden die Ergebnisse der Evaluierung des UAV-basierten topo-bathymetrischen Laserscanners in Form von Abbildungen und Tabellen präsentiert und die daraus gewonnenen Erkenntnisse diskutiert.

### 4.1 Messgenauigkeit

Abb. 4 Streifendifferenzen sind ein Maß für die relative Passgenauigkeit der Flugstreifen. Die vorherrschenden Weißtöne belegen die hohe Relativgenauigkeit des Flugblocks, welche auch aus dem Histogramm der Höhenabweichungen (Abb. 4, links unten) abgelesen werden kann. Die Verteilung ist symmetrisch um den Erwartungswert Null (Mittelwert: +1 mm) und fällt nach links und rechts in Richtung der positiven und negativen Abweichungen steil ab. Die Randquantile Q02/Q98 betragen -24/+25 mm und die robust abgeschätzte Standardabweichung ( $\sigma_{\text{MAD}}$ ) 9 mm. Wenige im Datensatz verbliebenen Ausreißer (Vegetation) beeinflussen die klassisch aus der Abweichungsquadratsumme aller Höhendifferenzen berechnete Standardabweichung, sind aber für den Flugblock nicht repräsentativ.

Abb. 5 zeigt für das in Abb. 4 mit einem weißen Rechteck gekennzeichnete Detailuntersuchungsgebiet (i) die Streifendifferenzen des trockenen und überströmten Bereichs sowie (ii) die farbkontrollierten Punkt-zu-Ebene-Abweichungen zwischen den tachymetrisch bestimmten Pass- und Kontrollpunkten und den jeweils benachbarten 3D-Laserpunkten. Die absoluten Abweichungen sind darüber hinaus auch in Tab. 2 für ausgewählte Referenzpunkte quantitativ angegeben. Die Punkte

an den im Uferbereich positionierten Satteldach-förmigen Holzplatten wurden für die absolute Orientierung des Flugblocks verwendet und belegen mit einer mittleren Abweichung im mm-Bereich die hohe Qualität des gesamten Sensorsystems (Laserscanner, GNN, IMU), für dessen geometrische Kalibrierung lediglich ein konservatives Modell zur Trajektorienkorrektur (konstanter Offset für X, Y, Z, roll, pitch, yaw pro Streifen) herangezogen wurde (GLIRA et al. 2016). Während die Abweichungen zum großen Teil geringer als  $\pm 1$  cm betragen, wird auf die Ursache für die systematisch negativen Differenzen im Bereich der Firstlinien der Satteldachpassflächen (z.B. Punkte 1022, 2010 in Tab. 2) im Abschnitt 4.3 näher eingegangen.

Größere Abweichungen treten hingegen bei den Unterwasser-Kontrollpunkten auf, wobei eine tiefenabhängige Tendenz zu erkennen ist (siehe Abb. 5). Während die Abweichungen für Punkte mit einer Wassertiefe von  $<1$  m nahe Null sind, steigen diese für Punkte in 2 m Wassertiefe auf etwa 7 cm an (vgl. Abb. 5 und Tab. 2). Die Laserpunkte sind dabei tendenziell zu tief.

Tab. 2: Absolute Höhenabweichungen zwischen Laserpunktwolke und Referenzmessungen für ausgewählte Pass- und Kontrollpunkte

Bereich	Punkt	$\Delta H$ [m]	Bereich	Punkt	Tiefe [m]	$\Delta H$ [m]
Passfläche (W)	1001	-0.006	Uferlinie	110	0.00	0.006
Passfläche (W)	1010	0.008	Fluss	201	0.41	0.004
Passfläche (W)	1019	-0.004	Fluss	202	0.93	-0.027
Passfläche (W)	1022	-0.034	Fluss	204	1.43	-0.066
Passfläche (O)	2002	0.006	Fluss	205	1.88	-0.069
Passfläche (O)	2010	-0.035	Fluss	206	2.10	-0.078

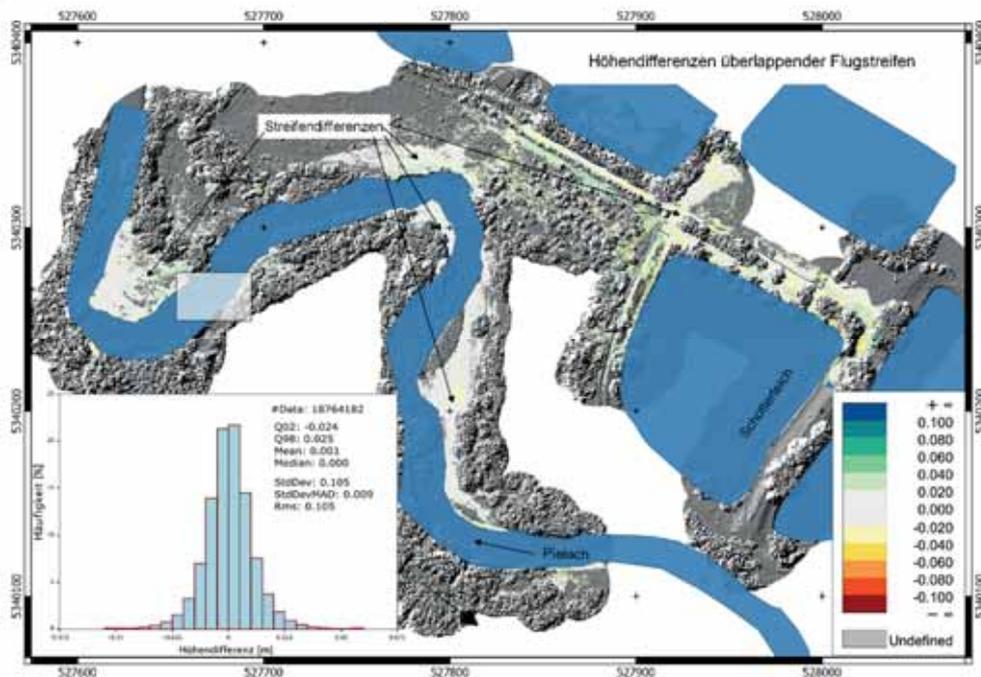


Abb. 4: Mosaik der farbkodierten Höhendifferenzen aller überlappenden Flugstreifen im trockenen Bereich, Bildhintergrund: Wasserflächen (blau) und Höhenmodell-Schummerung; links unten: Histogramm der Höhenabweichungen und statistische Kenngrößen; weißes Rechteck: Detailbereich der Abb. 5

Dieser Trend kann auch aus der in Abb. 6 dargestellten Profildarstellung der 3D-Punktwolke abgelesen werden, in die auch die Referenzpunkte (magenta) eingezeichnet sind. Die in Abb. 5 im südlichen Uferbereich ersichtlichen systematisch positiven Streifenhöhendifferenzen (dunkelgrüne und blaue Bereiche) sind ebenfalls auf den an den Referenzpunkten erkennbaren systematischen Effekt zurückzuführen. Für die Laufzeit- und Brechungskorrektur der rohen Lasermessungen kam der Wert  $n_w=1.36$  für den Refraktionsindex zur Anwendung. Bei hypothetischen Versuchen mit größeren Werten für  $n_w$  von 1.40-1.42 nehmen die Streifendifferenzen sowie die tiefenabhängige Systematik an den Referenzpunkten noch weiter ab. Der verwendete Wert von  $n_w=1.36$  wird durch zum Zeitpunkt der Verfassung des Manuskriptes noch laufende Untersuchungen nahegelegt. Größere Werte für  $n_w$  würden den berichteten absoluten Wertebereich von 1.329 bis 1.367 (MOBLEY 1994) aber in jedem Fall überschreiten.

Die Ursache für diese Systematik wird Gegenstand weiterer Untersuchungen sein. Eine mögliche Erklärung sind Mehrwegeeffekte innerhalb der Wassersäule, die durch komplexe Streuung an Schwebstoffteilchen zustande kommen, und ev. wegen der hohen Trübung zum Zeitpunkt der Messung besonders stark ausfallen. Eine systematische Fehlschätzung der Wasseroberfläche konnte durch Vergleich mit terrestrisch gemessenen Kontrollpunkten am Ufer ausgeschlossen werden.

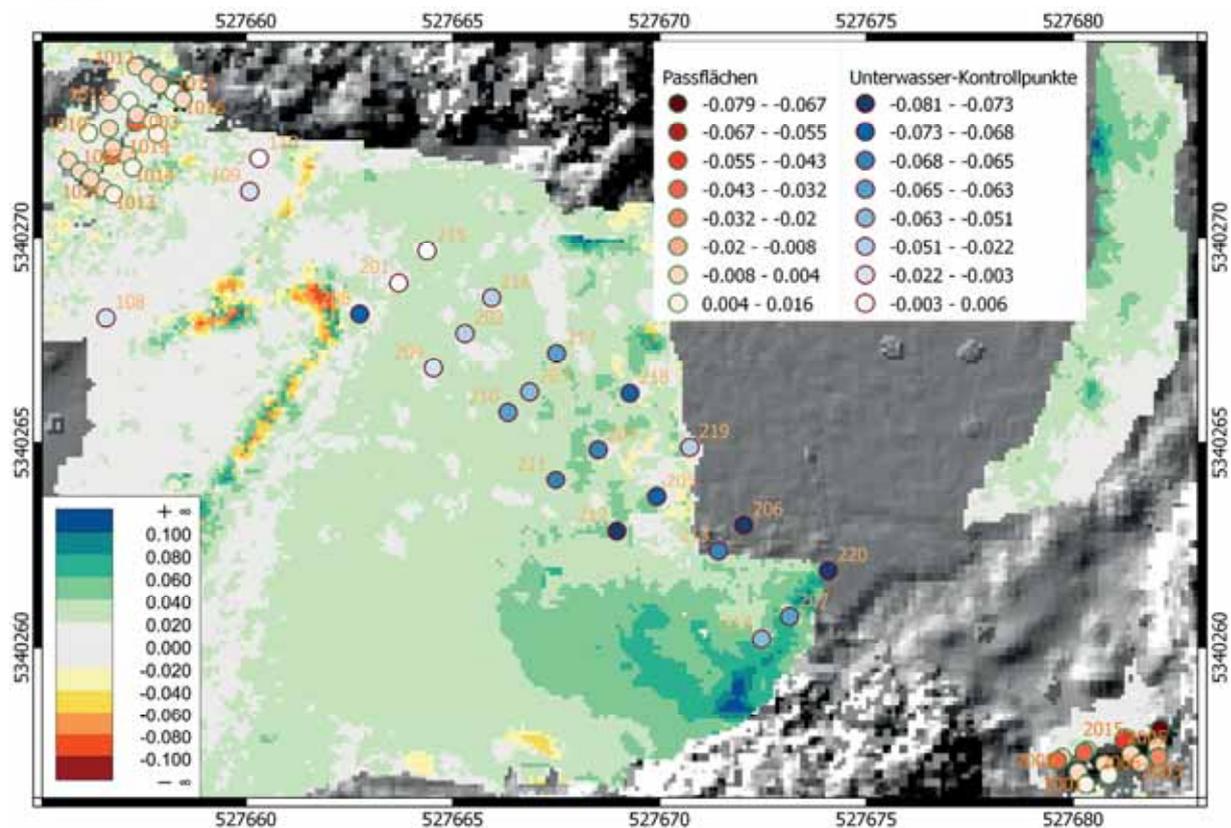


Abb. 5: Streifendifferenzen an Land und unter Wasser, Absolute Höhenabweichungen zwischen Laserpunkten und Passflächenpunkten (rot) bzw. Unterwasser-Kontrollpunkten (blau)

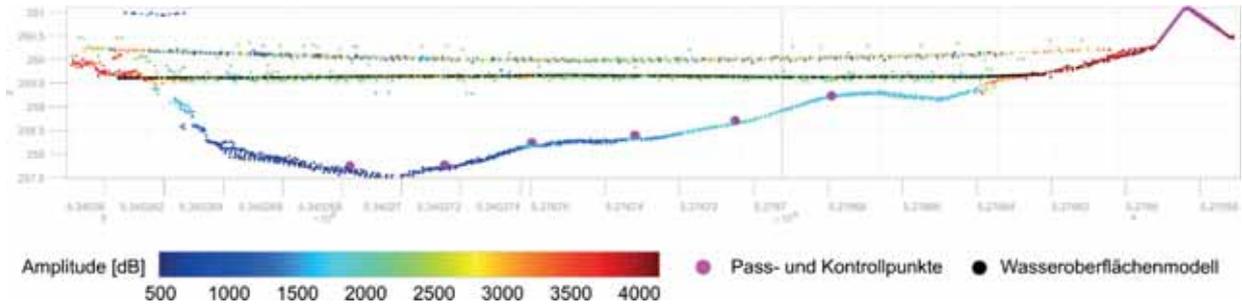


Abb. 6: Profilschnitt der refraktionskorrigierten 3D Punktwolke des Streifens 105314 eingefärbt nach Signalamplitude. Magenta: Pass- und Kontrollpunkte, Schwarz: modelliertes Wasseroberflächenmodell

## 4.2 Tiefenmessperformance

Zur Überprüfung der Tiefenmessperformance wurden in einem der flussnahen Schotterteiche 4 Metallplatten in Tiefen von 1.16 m, 1.99 m, 2.64 m und 3.11 m an Seilen abgesenkt. Abb. 7 zeigt eine Perspektivansicht der jeweiligen 3D Laserpunktwolken für einen ausgewählten Messstreifen, und die numerischen Ergebnisse im Vergleich zu den aus den Seillängen bekannten Wassertiefen sind in Tab. 3 zusammengefasst.

Es kann festgehalten werden, dass die Platten bis zu einer Tiefe von 2.64 m (P3) deutlich in der Punktwolke abgebildet sind (Amplitude P3: 6.8 dB) und lediglich die Platte P4 in 3.1 m Wassertiefe nicht erfasst ist. Bei der für den Teich gemessenen Secchi-Tiefe von 1.35 m beträgt die maximale Messtiefe daher die zweifache Secchi-Tiefe und entspricht damit dem Erwartungswert, der vom Hersteller für die Pulswiederholrate von 50 kHz angegeben ist. Die Abweichungen der aus den Laserpunkten bestimmten Tiefen folgen demselben tiefenabhängigen Trend, der bereits in Abschnitt 4.1 dokumentiert wurde. Bei dem auch hier für die Korrektur verwendeten Refraktionsindex von 1.36 beträgt die Überschätzung der Wassertiefe bei P3 9 cm oder umgerechnet ca. 3%.

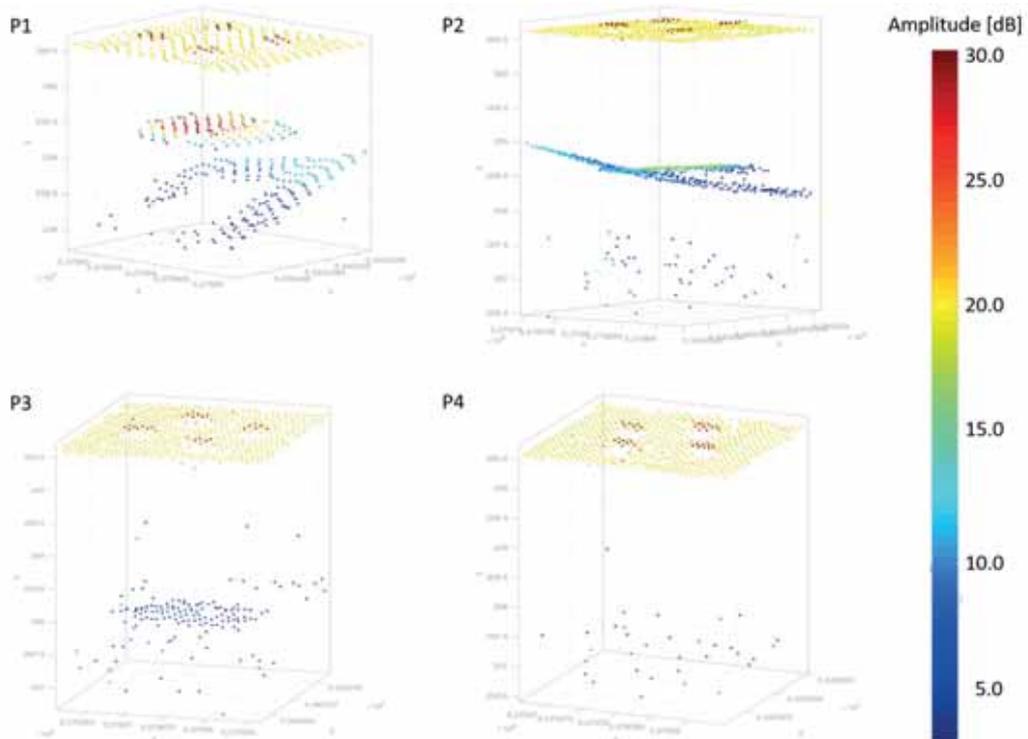


Abb. 7: Perspektivansicht der 3D Punktwolke im Bereich der versenkten Platten P1-P4 eingefärbt nach Signalamplitude

Tab. 3: Soll-Ist-Vergleich der vier abgesenkten Metallplatten

Platte	Soll-Tiefe [m]	Ist-Tiefe [m]	Abweichung [m]	Rel. Tiefe [x Secchi]	Amplitude [dB]
P1	1.16	1.18 ± 0.01	-0.02	0.86	26.8
P2	1.99	2.03 ± 0.02	-0.04	1.47	17.1
P3	2.64	2.73 ± 0.03	-0.09	1,96	6.8
P4	3.11	---	---	---	---

### 4.3 Räumliche Auflösung

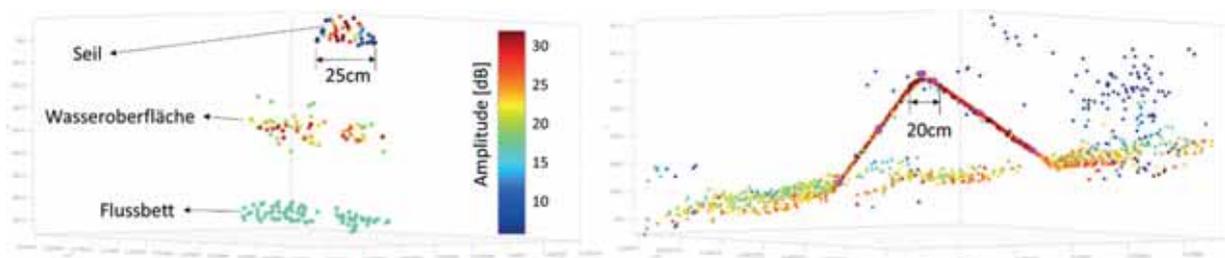


Abb. 8: 3D-Laserpunktwolke des Streifens 105314; (links) Detail im Bereich des gespannten Seils, (b) westliche Satteldach Passfläche, magenta: terrestrische Referenzpunkte

Wie im Abschnitt 3.1 ausgeführt, erfolgte die Datenerfassung aus einer Flughöhe von 50-75 m über Grund mit einer Strahldivergenz von 2 mrad. Die im Datenblatt des Herstellers angegebene Strahldivergenz bezieht sich dabei auf einen Abfall der Pulsamplitude auf  $1/e^2$  des Maximalwertes. Daraus errechnet sich ein theoretischer Durchmesser des Laserfootprints von 10-12.5 cm. Bei einer Pulswiederholrate von 50 kHz beträgt die mittlere Punktdichte im Einzelstreifen bei 50 m Flughöhe etwa 300 Punkte/m<sup>2</sup>, was einem Punktabstand von ca. 6 cm entspricht. Die räumliche Auflösung jedes Laserscanningsystems ist durch den größeren der beiden Werte (Footprintdurchmesser, Punktabstand) begrenzt und beträgt damit für das angegebene Szenario 10 cm.

Diese theoretische Auflösung wurde für den Streifen 105314 durch visuelle Kontrolle der Punktwolke in einem 3D-Editor überprüft. Dieser Streifen wurde mit den in der obigen Abschätzung angegebenen Parametern befliegen (h=50 m über Grund, Strahldivergenz=2 mrad, PRR=50 kHz). Abb. 8 zeigt zwei repräsentative Ausschnitte, die Rückschlüsse auf die effektive Auflösung erlauben. Die Abbildung enthält links den Bereich des quer über den Fluss gespannten Seils und rechts die Satteldach-förmige Holzplattenkonstruktion, die im Firstbereich einen rechten Winkel bildet. Dabei ist klar zu erkennen, dass (i) die Punktwolke entlang des Seils nicht streng linear sondern bandförmig ist (Breite des Bandes: 25cm) und (ii) der rechte Winkel der beiden Holzplatten in der Punktwolke zu einem Kreis mit einem Durchmesser von ca. 20 cm abgerundet ist. Aufgrund der großen Punktdichte kann der effektive Footprintdurchmesser anhand des linear verlaufenden Seils direkt abgelesen werden. Da die Empfindlichkeit des Sensors weit größer als 8.7 dB ist, was der Definition des nominalen Strahldurchmessers ( $10 \log(e^2)$ ) entspricht, empfängt dieser auch außerhalb des nominellen Footprints Signale, die auch zu detektierten Echos führen. Deutlich ist auch der Amplitudenabfall von der Mitte des Bandes (rot: 25-30 dB) zum Rand hin (blau: <10 dB) zu erkennen, entsprechend dem Anteil der Seilstrecke innerhalb des Gesichtsfeldes des Empfängers. Der größere effektive Footprintdurchmesser ist auch der Grund für die Abrundung des rechten Winkels in der rechten Abbildung.

## 5 Zusammenfassung

In diesem Beitrag wurden einerseits das Konzept und andererseits eine Untersuchung der Genauigkeit und Tiefenmessperformance des topo-bathymetrischen Laserscanners *RIEGL VQ-840-G* vorgestellt. Mit einem Gesamtgewicht von 12 kg eignet sich der Sensor für die Integration auf größere Multikopter-UAV Plattformen, wodurch ein flexibler Einsatz für die Erfassung kleiner bis mittelgroßer Gewässer mit moderater Wassertiefe als Alternative zur klassischen flugzeugbasierten Laserbathymetrie gegeben ist. Der Klasse 3B Laser operiert mit grünem Licht ( $\lambda=532$  nm), weist eine Pulsrepetitionrate von 50-200 kHz, eine Strahldivergenz von 1-6 mrad und ein Empfängergesichtsfeld von 6-18 mrad auf. Bei einer Flughöhe von 50 m beträgt die Punktdichte damit etwa 100-400 Punkte/m<sup>2</sup> und der Durchmesser des Abtastflecks 5-30 cm.

Zur Evaluierung des Laserscanners wurde im August 2019 eine Messkampagne an der Pielach im Bereich Neubacher Au (Niederösterreich) durchgeführt. Neben der UAV-basierten Laserdatenerfassung wurden auch Pass- und Kontrollpunkte im trockenen und benetzten Bereich installiert und mittels RTK-GNSS bzw. Totalstation terrestrisch vermessen. Gegenüber diesen Referenzpunkten ergab sich eine absolute Höhenabweichung an Land von 1 cm und Unterwasser von 0-8 cm. Bei den bathymetrischen Kontrollpunkten konnte ein tiefenabhängiger Fehler festgestellt werden. Die

Erforschung der Ursache für diesen systematischen Effekt ist Gegenstand laufender Untersuchungen.

Anhand von vier in unterschiedlichen Tiefen angeordneten  $1 \times 1 \text{ m}^2$  großen Metallplatten konnte die vom Hersteller angegebene 2-fache Secchi-Tiefe als maximale Eindringtiefe bestätigt werden. Eine vierte, in 2.3-facher Secchi-Tiefe installierte Platte konnte dagegen in der Punktwolke nicht identifiziert werden. Es besteht aber die Erwartung, dass die Eindringtiefe durch Mittelung von räumlich/zeitlich benachbarten Wellenformen weitere erhöht werden kann. Die erzielte 2-fache Secchi-Tiefe stellt aber bereits jetzt eine Steigerung gegenüber topo-bathymetrischer Datenerfassung vom bemannten Flugzeug aus dar, wo die Eindringtiefe eher bei der 1.5-fachen Secchi-Tiefe liegt.

Weitere Vorteile der UAV-basierten Erfassung sind die hohe Punktdichte und der in Folge der geringeren Messdistanz wesentlich kleinere Footprintdurchmesser. Entsprechende Untersuchungen der planimetrischen Auflösung haben ergeben, dass trotz geringem Messpunktstand und geringer Strahldivergenz, die effektive Auflösung etwa um einen Faktor 2 geringer ist, als von der im Datenblatt angegebenen Strahldivergenz her zu erwarten war. Bei zwei im rechten Winkel zu einander positionierten Holzflächen wies die 3D Punktwolke eine etwa halbkreisförmige Ausrundung mit einem Durchmesser von ca. 20 cm auf.

Generell kann festgehalten werden, dass sich der Sensor durch die hohe Genauigkeit sowie die gute Tiefenmessperformance für vielfältige Aufgaben im Bereich der präzisen Erfassung von Gewässern mit moderater Tiefe und Ausdehnung eignet, z.B. als Grundlage für Überflutungssimulationen, Monitoring von fluvialen Veränderungen, Habitatmodellierung und dergleichen mehr.

## 6 Danksagung

Die Arbeiten von Gottfried Mandlbürger zu diesem Beitrag wurden durch Mitteln der Deutschen Forschungsgesellschaft (DFG) im Rahmen des Forschungsprojektes „Bathymetrievermessung durch Fusion von Flugzeuglaserscanning und multispektralen Luftbildern“ unterstützt.

## 7 Literaturverzeichnis

- APPLANIX, 2020: Homepage Applanix POSPac. <https://www.applanix.com/products/pospac-mms.htm>.
- ASTRALiTE, 2020: Homepage der ASTRALiTe Inc. <https://www.astralite.net>.
- FUGRO, 2020: Homepage des Fugro Rapid airborne multibeam mapping system (RAMMS). <https://www.fugro.com/about-fugro/our-expertise/innovations/rapid-airborne-multibeam-mapping-system>.
- GLIRA, P., PFEIFER, N., BRIESE, C. & RESSL, C., 2015: Rigorous strip adjustment of airborne laserscanning data based on the ICP algorithm. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **II-3/W5**, 73-80.
- GLIRA, P., PFEIFER, N. & MANDLBURGER, G. 2016. Rigorous Strip Adjustment of UAV-based Laserscanning Data Including Time-Dependent Correction of Trajectory Errors. *Photogrammetric Engineering & Remote Sensing* **82(12)**, 945-954.

- GUENTHER, G.C., CUNNINGHAM, A.G., LAROQUE, P.E. & REID, D.J., 2000: Meeting the accuracy challenge in airborne lidar bathymetry. In: Proceedings of the 20th EARSeL Symposium: Workshop on Lidar Remote Sensing of Land and Sea, Dresden.
- LEICA, 2020: Homepage Leica Auswertesoftware. <https://leica-geosystems.com/de-at/products/total-stations/software>.
- MANDLBURGER, G., HAUER, C., WIESER, M., & PFEIFER, N., 2015: Topo-Bathymetric LiDAR for Monitoring River Morphodynamics and Instream Habitats—A Case Study at the Pielach River. *Remote Sens.* **7**(5), 6160-6195, <https://doi.org/10.3390/rs70506160>.
- MANDLBURGER, G., PFENNIGBAUER, M., WIESER, M., RIEGL, U. & PFEIFER, N., 2016: Evaluation of a novel UAV-borne topo-bathymetric laser profiler. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **41**(B1), 933 - 939.
- MOBLEY, C. D., 1994: *Light and Water: Radiative Transfer in Natural Waters*. San Diego, CA: San Diego Academic Press.
- PFENNIGBAUER, M., WOLF, C., WEINKOPF, J. & ULLRICH, A., 2014: Online waveform processing for demanding target situations. In *Proc. SPIE*, **90800J**.
- PIX4D, 2020: Homepage Pix4DMapper. <https://www.pix4d.com/de/produkt/pix4dmapper-photogrammetrie-software>.
- RESOP, J.P., LEHMANN, L. & HESSION, W.C., 2019: Drone Laser Scanning for Modeling Riverscape Topography and Vegetation: Comparison with Traditional Aerial Lidar. *Drones* **3**(35), 1-15.
- RESSL, C., KAGER, H. & MANDLBURGER, G., 2008: Quality Checking Of ALS Projects Using Statistics Of Strip Differences. *The International Society for Photogrammetry and Remote Sensing* **37**(B3), 1682-1750.
- RIEGL, 2020: Datenblatt des topo-bathymetrischen Laserscanners VQ-840-G. [http://www.riegl.com/uploads/tx\\_pxpriegldownloads/RIEGL\\_VQ-840-G\\_Preliminary-Datasheet\\_2019-09-02.pdf](http://www.riegl.com/uploads/tx_pxpriegldownloads/RIEGL_VQ-840-G_Preliminary-Datasheet_2019-09-02.pdf).

# Möglichkeiten und Grenzen der Bathymetrieableitung mittels Single Photon LiDAR

GOTTFRIED MANDLBURGER<sup>1,2</sup>

*Zusammenfassung: Neben klassischen Laserbathymetriesensoren bietet auch die Single Photon LiDAR Technologie prinzipiell die Möglichkeit der Erfassung von Gewässertiefen. In diesem Beitrag wird die Eignung von Single Photon LiDAR zur Bathymetrieableitung sowohl von der Theorie her als auch anhand praktischer Beispiele einer Datenerfassung mit dem Leica SPL100 in Wien und Niederösterreich erörtert. Die prinzipielle Eignung begründet sich durch die Verwendung grüner Laserstrahlung, dem konischen Scanmechanismus mit konstanter Strahlrichtung gegenüber der Horizontalen und der prinzipiellen Mehrzielfähigkeit der eingesetzten sehr empfindlichen Detektoren. Einschränkungen ergeben sich wegen der nicht vorhandenen Wellenformaufzeichnung und aufgrund der potenziellen Signalinteraktion eng benachbarter Laserstrahlen und deren zugeordneter Empfänger. Je nach Gewässertrübe konnte bei der durchgeführten Datenerfassung entweder limitierte Eindringtiefe oder unzureichende Signalantwort von der Wasseroberfläche festgestellt werden. Die erzielten Messtiefen liegen im Bereich der 1-fachen Secchi Tiefe.*

## 1 Einleitung

Bei konventionellen Laserscannern wird die rückgestreute Laserstrahlung von genau einem Empfänger erfasst. Die Distanzmessung entlang des Laserstrahls erfolgt dabei entweder durch geeignete Hardwarekomponenten direkt im Gerät oder durch die Aufzeichnung des gesamten zeitlichen Verlaufs des rückgestreuten Signals. Diese Art der Signalerfassung und -verarbeitung wird als Full Waveform (FWF) Laserscanning (MALLET & BRETAR 2009) bezeichnet. Alternativ dazu können Echodetektion und Entfernungsmessung auch geräteintern auf Basis von schwellwertsensitiven Hardwarekomponenten erfolgen. Unabhängig von der Art der Signalverarbeitung spricht man in beiden Fällen von linear-mode LiDAR (Light Detection and Ranging), da die verwendeten Avalanche Photodioden (APD) im linearen Bereich operieren, bei dem die empfangene optische Strahlungsstärke in etwa linear zum elektrischen Signal am Ausgang des Empfängers ist.

Demgegenüber sind in den vergangenen Jahren Systeme auf den Markt gekommen, die eine Objekterfassung beim Eintreffen eines oder weniger Photonen erlauben. Die Technik wird als Single Photon Sensitive LiDAR bezeichnet (DEGNAN 2002; AGISHEV et al. 2013; KIM et al. 2013). Im Kontext der Topographierfassung mittels Flugzeuglaserscanning haben sich zwei spezifische Varianten etabliert: (i) Geiger-Mode LiDAR (GmLiDAR) (ROMANO 2015) und (ii) Single Photon LiDAR (SPL) (DEGNAN 2016). MANDLBURGER et al. (2019) beschreiben beide Verfahren detailliert und diskutieren die Unterschiede zu konventionellem wellenformbasierten Laserscanning.

---

<sup>1</sup> TU Wien, Department für Geodäsie and Geoinformation, Wiedner Hauptstr. 8-10, A-1040 Wien, Österreich, E-Mail: gottfried.mandlbürger@geo.tuwien.ac.at

<sup>2</sup> Universität Stuttgart, Institut für Photogrammetrie, Geschwister-Scholl-Str. 24D, D-70174 Stuttgart, E-Mail: gottfried.mandlbürger@ifp.uni-stuttgart.de

Die als SPL bezeichnete Technologie verwendet Laserstrahlung im grünen Bereich des elektromagnetischen Spektrums ( $\lambda=532$  nm), welches die inhärente Eigenschaft aufweist, Wasser zu durchdringen. Ein am Markt verfügbares Gerät dieser Type ist der Leica SPL100. Obgleich der Hersteller den Sensor ausschließlich für die Erfassung von Topografie konzipiert hat, weist das System aufgrund (i) der verwendeten grünen Wellenlänge, (ii) der hohen Sensitivität des Empfängers und (iii) des kreisförmigen Scanmusters (Palmer Scanner, off-nadir Winkel:  $10-20^\circ$ ) bathymetrische Fähigkeiten auf. DEGNAN et al. (2016) berichten beispielsweise von einer Eindringtiefe von 15 m bei sehr klaren antarktischen Gewässern und 4 m bei trüberen Verhältnissen (Chesapeake Bay, Maryland, USA).

Ziel des vorliegenden Beitrages ist es daher, einerseits die technischen Grundlagen wie auch die Grenzen der Single Photon LiDAR basierten Ableitung von Gewässertiefen zu erläutern und andererseits anhand konkreter Messdaten zu illustrieren. Der Artikel ist wie folgt strukturiert: Abschnitt 2 behandelt die Theorie der Single Photon LiDAR basierten Detektion von Laserechos mit speziellem Hinblick auf den Wasserbereich. Konkrete Untersuchungsgebiete in Wien und Niederösterreich (Donau, Pielach) werden in Abschnitt 3 vorgestellt und die Ergebnisse der entsprechenden Befliegung mit dem Leica SPL100 in Abschnitt 4 präsentiert und diskutiert. Der Artikel schließt mit einer Zusammenfassung der wesentlichen Erkenntnisse in Abschnitt 5.

## 2 Theorie

Im Gegensatz zu konventionellem Laserscanning, wo für jeden Laserstrahl genau ein Detektor zur Verfügung steht, kommen sowohl bei GmLiDAR als auch bei SPL Detektorarrays zum Einsatz. Bei GmLiDAR trifft dabei die reflektierte Strahlung eines stark aufgeweiteten Laserpulses auf ein Feld von Einzelphoton-sensitiven Empfängerzellen, einem sogenannten Geiger-mode Avalanche Photodiode (GmAPD) array (CLIFTON et al. 2015). Es handelt sich also um eine Range Kamera mit beispielsweise  $32 \times 128$  Zellen (ROMANO 2015). Jede einzelne Zelle agiert dabei als binärer Detektor, der nach dem Auslösen eines ersten Echos bis zum Zurücksetzen des gesamten Empfängerarrays für den nächsten Laserpuls unempfindlich ist (ULLRICH & PFENNIGBAUER 2016; MANDLBURGER et al. 2019). Dieses Messprinzip eignet sich daher wegen der mangelnden Mehrzielfähigkeit nicht für Bathymetrieerfassung, da das von Wasserflächen zurückgestreute Lasersignal prinzipiell aus Komponenten von der Wasseroberfläche, der Wassersäule und des Gewässerbodens besteht. Außerdem kommt bei GmLiDAR infrarote Laserstrahlung zum Einsatz, die generell nicht ins Wasser eindringt.

Bei der SPL Technologie kommen ebenfalls mehrere Empfänger pro Laserschuss zum Einsatz. Hier wird allerdings ein wasserdurchdringender grüne Laserpuls durch ein diffraktives, optisches Element in ein Bündel von  $10 \times 10$  Teilstrahlen aufgefächert (STOKER et al. 2016, MANDLBURGER et al. 2019). Jeder einzelne Teilstrahl ist stark gebündelt. Die Strahldivergenz beträgt  $0.08$  mrad (LEICA 2020), wodurch der Durchmesser des Abtastflecks jedes Teilstrahls bei einer typischen Flughöhe von 4000 m über Grund ca. 30 cm beträgt. Für jeden Teilstrahl steht ein eigener Empfänger bestehend aus mehreren hundert Einzelphoton-sensitiven Zellen (Silicon Photo-Multiplier, SiPM), zur Verfügung (DEGNAN 2018). Einzelne dieser Geiger-mode Zellelemente können dabei bei der Signalantwort von der Wasseroberfläche durchbrechen, andere erst später beim Eintreffen

des reflektierten Signals vom Gewässergrund. Diese Technologie ist daher inhärent mehrzielfähig, was die Grundvoraussetzung für die korrekte Erfassung von Unterwasser-Topografie darstellt.

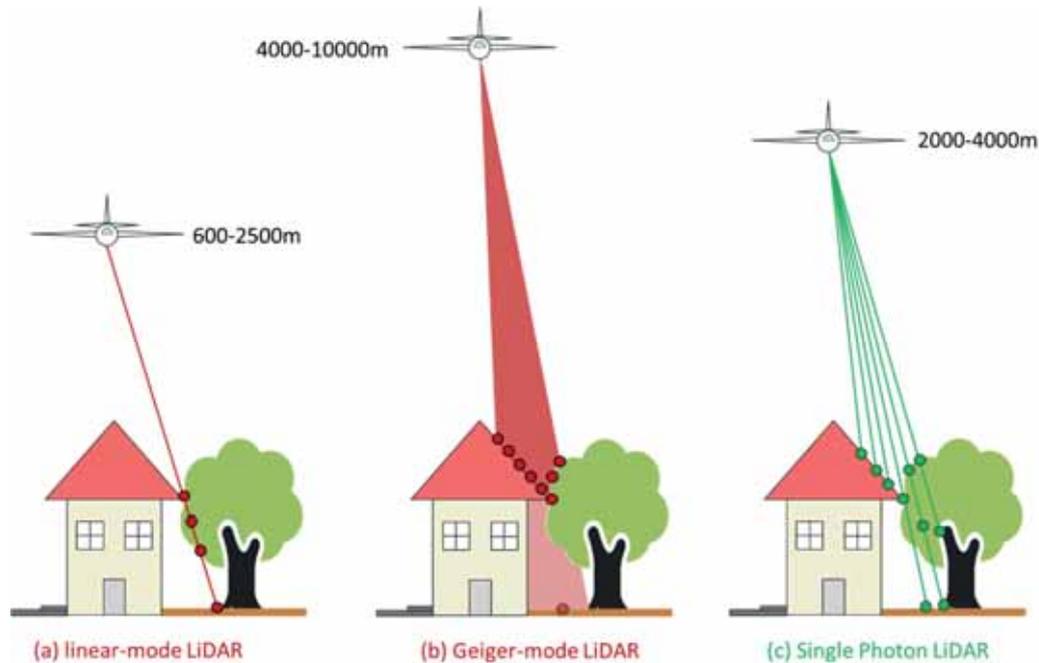


Abb. 1: Prinzipskizze eines einzelnen Laserpulses bei (a) konventionellem Laserscanning, (b) Geiger-mode LiDAR und (c) Single Photon LiDAR

Abb. 1 zeigt die Prinzipskizze eines Einzelpulses für alle drei genannten Erfassungsmethoden (konventionelles Laserscanning, GmLiDAR, SPL). In Abb. 1b ist dabei die relativ starke Strahlaufweitung beim GmLiDAR angedeutet und in Abb. 1c die Auffächerung des Einzelpulses in Teilstrahlen beim SPL visualisiert. Das Prinzip der Echodetektion bei konventionellem Laserscanning und bei SPL ist in Abb. 2 skizziert. Beim linear-mode LiDAR kann die gesamte Wellenform aufgezeichnet werden und entweder online (PFENNIGBAUER ET AL. 2014) oder im postprocessing (MALLETT & BRETAR, 2009) ausgewertet werden. Dies führt dazu, dass im skizzierten Fall auch das schwächere Ziel (2. Ast) detektiert werden kann. Beim SPL liefern die vielen Einzelzellen ein Summensignal am Ausgang des SiPM Detektors. Dies ist durch das Bündel an blauen Linien von jeder Zelle zum gemeinsamen Ausgang schematisch angedeutet. Die Detektion selbst erfolgt mittels Schwellwertverfahren. In der dargestellten Situation ist das Signal des zweiten Ziels dabei beispielsweise zu schwach, um zu einem Messpunkt zu führen.

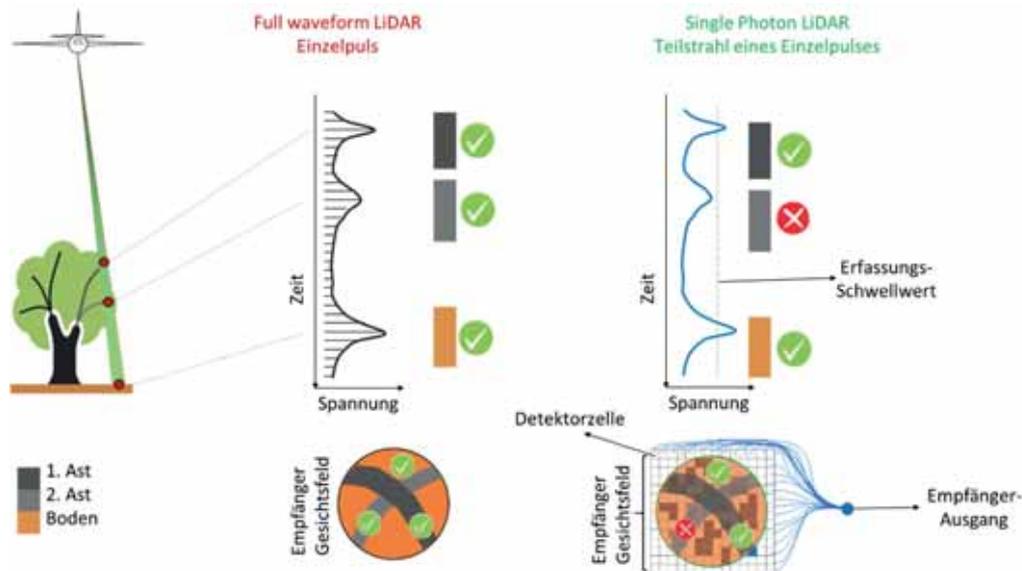


Abb. 2: Schematische Abbildung der Signalerfassung beim Full Waveform Laserscanning (links) und Single Photon LiDAR (rechts)

In der flugzeuggetragenen Laserbathymetrie kommen seit jeher wellenformbasierte Sensorsysteme zum Einsatz (GUENTHER et al. 2000). Die Interaktion der Laserstrahlung mit dem Medium Wasser ist komplex (PFEIFER et al. 2016). Aufgrund des hohen Anteils an gerichteter Reflexion ist die Anteil von der Wasseroberfläche stark richtungsabhängig. In der Wassersäule kommt es durch Volumenstreuung an Partikeln im Wasser zu einer nichtlinearen Strahlaufweitung und einer generell exponentiellen Signaldämpfung, wodurch die Stärke des vom Gewässerboden zurückgestreuten Signals mit zunehmender Wassertiefe massiv abnimmt. Optimale Ergebnisse sind daher generell mittels bathymetriespezifischer Wellenformauswertung zu erreichen (ALLOUIS et al. 2010; SCHWARZ et al. 2019). Allerdings werden heute auch Scanner für bathymetrische Zwecke genutzt, bei denen die Echodetektion auf Schwellwertverfahren basiert und die grüne Wellenlänge als Teil eines multispektralen Scanners eher für die Ableitung von Vegetationsindizes ausgelegt ist (FERNANDEZ-DIAZ et al. 2016). Darüber hinaus setzen andere dezidierte Bathymetriescanner zwar Wellenformaufzeichnung ein, werten die Signale aber online im Gerät ohne besondere Berücksichtigung der spezifischen Eigenschaften von Wasserzielen aus (PFENNIGBAUER et al. 2014). Aus diesem Grund ist die bei SPL eingesetzte schwellwertbasierte Echodetektion prinzipiell als bathymetrietauglich einzustufen.

Wie bereits weiter oben erwähnt, wird bei SPL der Laserstrahl in 10 x 10 stark gebündelte Teilstrahlen aufgefächert. Beim Leica SPL100 im Speziellen beträgt die nominelle Strahldivergenz laut Datenblatt 0.08 mrad (LEICA 2020). Analyse konkreter Messdaten ergab für den Divergenzwinkel zweier benachbarter Teilstrahlen etwa 0.16 mrad. Bei einer Flughöhe von 3750 m beträgt der Durchmesser des Laserfootprints am Boden daher 30 cm bei einem Abstand benachbarter Punkte von 60 cm. Für Anwendungen in der Topografie ist dieser Abstand gerade ausreichend, um zu verhindern, dass der reflektierte Laserpuls eines Teilstrahls vom benachbarten Detektor erfasst wird. Die Volumenstreuung in der Wassersäule führt aber bei zunehmender Wassertiefe zu

einer nicht-linearen Aufweitung des Laserfootprints. Entsprechende Prinzipskizzen zur Unterwasser-Strahlaufweitung sind in u.a. in GUENTHER et al. (2000; Abb. 4) und PFEIFER et al. (2016; Abb. 9) zu finden.

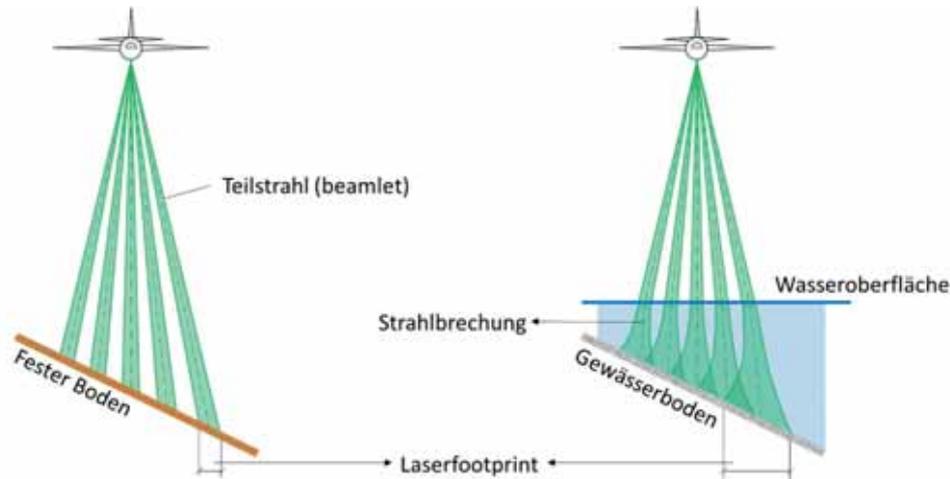


Abb. 3: Schematische Darstellung der Laserstrahlanordnung und Strahlaufweitung beim SPL100; links: Topografieaufnahme mit linearer Strahlaufweitung und nicht überlappenden Footprints; rechts: Bathymetrieaufnahme mit exponentieller Strahlaufweitung und überlappenden Footprints

Abb. 3 zeigt die Auswirkungen für die eng benachbarten Teilstrahlen des SPL Systems. Daraus ist zu erkennen, dass es im Gegensatz zu topografischen Zielen (links) bei bathymetrischen Anwendungen (rechts) durch die Strahlaufweitung zu einer Überlappung der Laserfootprints kommt. Dadurch kann aber auch das vom Gewässerboden reflektierte Signal eines Teilstrahls zu benachbarten Detektoren gelangen und dort ggf. ein Echo auslösen. Wegen der Divergenz der einzelnen Teilstrahlen (Fächer) kommt es dadurch zu einer Verfälschung der Entfernungsmessung und damit zu potenziell systematischen Fehlern in der Unterwasserpunktwolke. Der Effekt ist bei horizontalem Gewässergrund eher gering, da der gesamte Öffnungskegel des Strahlenbündels lediglich etwa 1.5 mrad beträgt. Bei großer Neigung kann die fehlerbehaftete Entfernungsmessung aber Höhenfehler im dm- bis m-Bereich verursachen. Entsprechende Beobachtungen wurden bereits an Traufenlinien von Hausdächern gemacht (MANDLBURGER et al. 2019), sind aber bei Unterwasserzielen bei den durch Volumenstreuung effektiv überlappenden Laserfootprints wesentlich wahrscheinlicher.

In Abb. 3 ist neben der Strahlaufweitung im Wasser auch die Strahlbrechung an der Wasseroberfläche angedeutet. Voraussetzung für die Refraktionskorrektur ist daher die Kenntnis der Wasseroberfläche. MANDLBURGER & JUTZI (2019) haben die Eignung von SPL zur Erfassung der Wasseroberfläche untersucht und kamen zum Schluss, dass bei typischen Flughöhen (ca. 4000 m) und Strahlneigungen ( $15^\circ$ ) nur wenige Photonen von der Wasseroberfläche beim Empfänger ankommen. Die Bestimmung der Wasseroberfläche ist daher unsicher und erfordert räumliche Aggregation und die Anwendung von statistischen Methoden für die Festlegung der lokalen Wasserspiegellage. Die mangelnde Wellenformaufzeichnung, die potenzielle Pulsüberlappung und die unsichere Wasseroberflächenbestimmung stellen somit die Einschränkungen für die SPL-basierte Bathymetrieableitung dar.

### 3 Untersuchungsgebiete

Im Auftrag der Stadt Wien (Magistratsabteilung 41, Stadtvermessung) erfolgte Ende Juli 2018 im Rahmen eines Pilotprojektes die Erfassung von etwa 30% des Wiener Stadtgebietes (MANDLBURGER et al. 2019). Der Flugblock (20 km x 8 km) besteht aus 10 Streifen mit 50%iger Überlappung. Die Befliegung erfolgte aus ca. 3800 m Flughöhe mit einem Öffnungswinkel des Scankegels von 15°, woraus sich eine Streifenbreite von 2000 m ergibt. Das Befliegungsgebiet reicht von der Lobau im Osten über den Bereich der Wiener Innenstadt bis in den westlich angrenzenden Wiener Wald. Dabei wurden auch zahlreiche Gewässerflächen (Donau, Neue Donau, Alte Donau, Heustadlwasser, etc.) aufgenommen. In derselben Mission wurden auch zwei Flugstreifen in Niederösterreich entlang der Pielach, einem voralpinen Schotterfluss in Niederösterreich (MANDLBURGER et al. 2015) erfasst. Die Untersuchungsgebiete sind in Abb. 4 dargestellt.

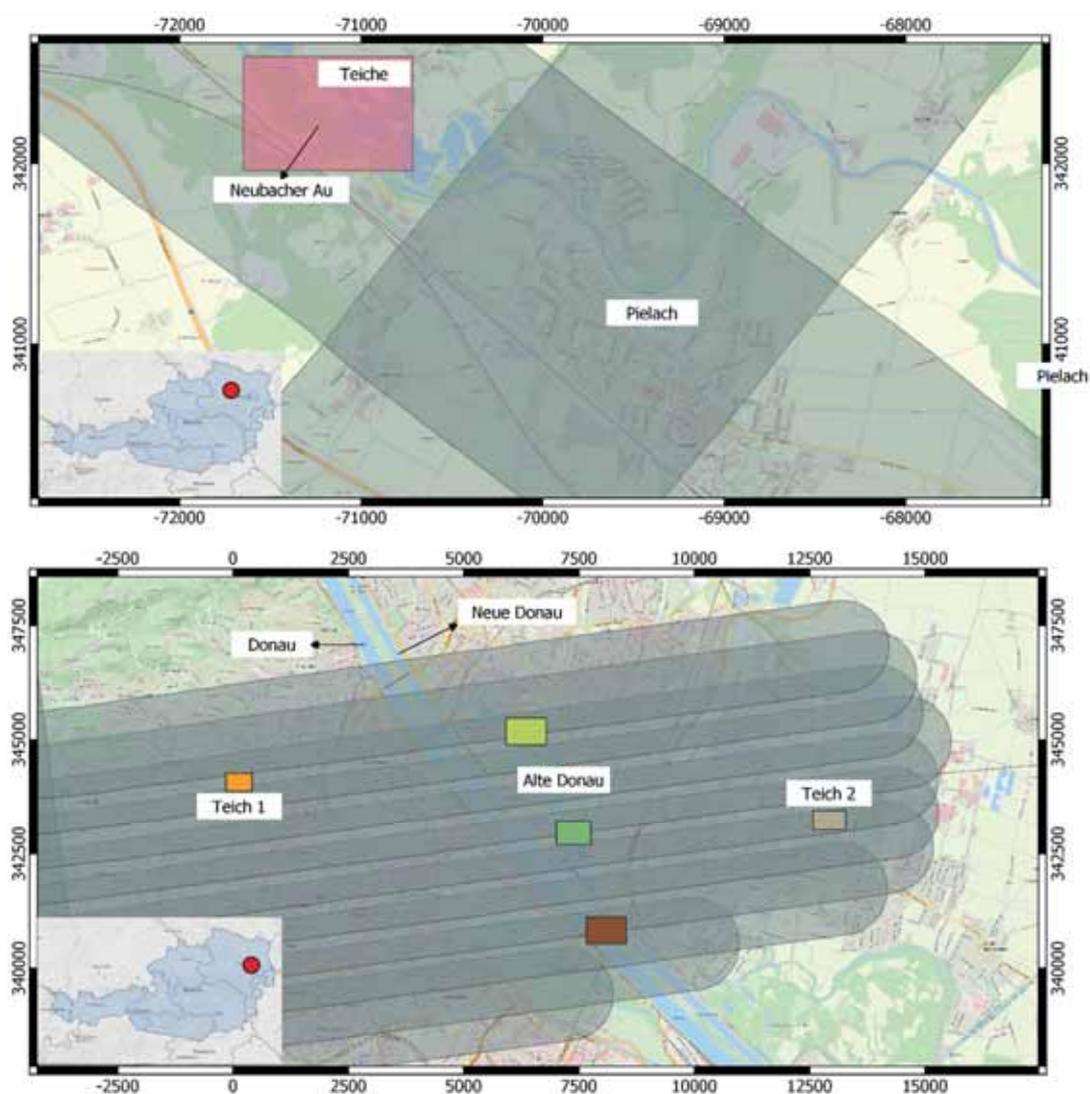


Abb. 4: Untersuchungsgebiet Wien (oben) und Niederösterreich, Pielach (unten), Übersichtskarte: basemap.at, Koordinatenbezugssystem: MGI, Gauß-Krüger M34 (EPSG: 31256)

## 4 Ergebnisse und Diskussion

In Abb. 5 (Wien) und Abb. 6 (Niederösterreich) sind ausgewählte Profilschnitte aus der SPL100 Punktwolke abgebildet. In allen Fällen beträgt die Profilbreite 3 m und die Punkte sind nach Signalamplitude eingefärbt, wobei rot für hohe und blau für niedrige Amplitude steht. Die Punktwolken sind nicht refraktionskorrigiert, und daher sind die scheinbaren Gewässersohlpunkte tendenziell zu tief.

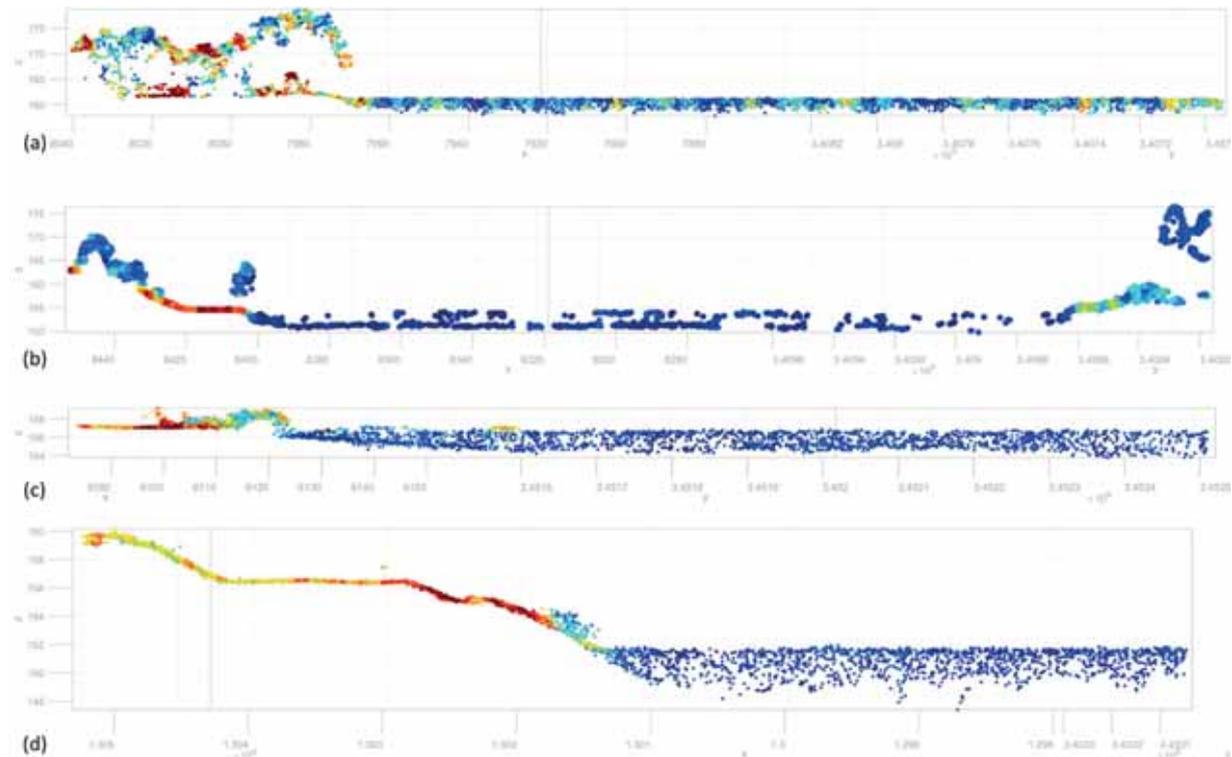


Abb. 5: Untersuchungsgebiet Wien, Profilschnitte durch SPL100 3D-Punktwolke eingefärbt nach Signalamplitude (rot: hoch, blau: niedrig) (a) Donau, (b) Neue Donau, (c) Alte Donau, (d) Teich 2

Abb. 5a zeigt einen Ausschnitt der Donau, die eine verhältnismäßig große Trübe aufweist. Die Erfassung von Gewässertiefen mittels Laserbathymetrie ist generell bei Flüssen dieser Größe auch für darauf spezialisierte Sensoren problematisch und nur im unmittelbaren Uferbereich möglich. Die SPL Punktwolke zeigt hier ein etwa 2 m breites Band an Punkten, wobei Sohlpunkte nur bis zu einer Tiefe von ca. 1 m reichen, bevor der Gewässerboden im Rauschen verschwindet. Die Wasseroberfläche hingegen bildet sich relativ klar ab. Durch Rückstreuung an Sedimentpartikeln gelangt ausreichend Signal von knapp unterhalb der Wasseroberfläche zum Sensor, sodass die Oberfläche flächendeckend abgebildet ist. In Abb. 5b von der Neuen Donau, einem parallel zum Donaustrom verlaufenden stehenden Gewässer mit relativ klarerem Wasser, sind Echos von der Wasseroberfläche nur sporadisch zu beobachten. Im Gegensatz dazu ist aber der Gewässerboden bis zu einer Tiefe von ca. 2.5 m deutlich abgebildet. Dies unterstreicht die prinzipielle Eignung

des Sensors für die Kartierung von Unterwasser Topografie, zeigt aber andererseits auch die Grenzen, da für präzise Bathymetrieerfassung sowohl Echos vom Gewässerboden als auch verlässliche Information von der Wasseroberfläche erforderlich sind.

Abb. 5c und d schließlich zeigen zwei weitere Beispiele von stehenden Gewässern (c: Alte Donau, d: künstlicher Teich). Beide Punktwolken weisen eine gute Abdeckung mit Punkten nahe der Wasseroberfläche und eine maximale Tiefe der klar erkennbaren Gewässerbodenpunkte von ca. 1.8 m auf. An der Alten Donau wurde etwa zum Zeitpunkt der Befliegung eine Secchi Tiefe von 1.6 m gemessen. Die maximale Eindringtiefe des Sensors entspricht in diesem Bereich daher der einfachen Secchi Tiefe.

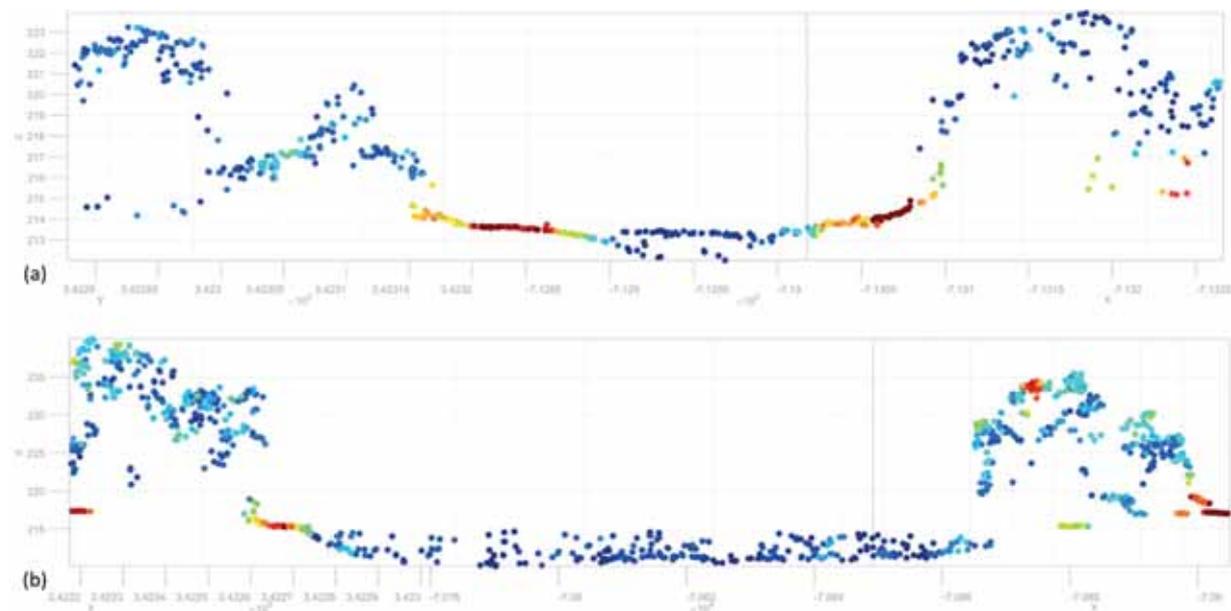


Abb. 6: Untersuchungsgebiet Niederösterreich/Pielach, Profilschnitte durch SPL100 3D-Punktwolke eingefärbt nach Signalamplitude (rot: hoch, blau: niedrig) (a) Pielach, (b) Schotterteich

In Abb. 6 sind zwei Beispiele aus dem Untersuchungsgebiet Niederösterreich/Pielach abgebildet. Abb. 6a zeigt einen Profilschnitt durch den Pielach Fluss. Die Wasseroberfläche ist in der Punktwolke klar erkennbar, die Gewässerbodenpunkte hingegen weisen in der Mitte des etwa 20 m breiten Flusses eine Lücke auf. Die Eindringtiefe beträgt etwa 1.2 m. Interessant ist in diesem Zusammenhang der flach überströmte Uferbereich. Der Flussschotter im trockenen Bereich reflektiert sehr gut (rote Punkte mit hoher Signalamplitude). In Richtung Flussmitte sind dann Punkte am Gewässerboden mit etwas geringerer Amplitude (gelb, grün) zu erkennen, zu denen es keine korrespondierenden Punkte der Wasseroberfläche gibt. Das ist ein Indiz dafür, dass die Signalantwort von der Wasseroberfläche zu schwach für flächendeckende Echodetektion ist, während die Punkte am Gewässerboden bereits durch die Dämpfung in der Wassersäule eine zunehmend geringere Signalamplitude aufweisen. Erst bei größerer Tiefe gegen die Flussmitte hin sind sowohl Oberflächenpunkte als auch Bodenpunkte zu erkennen. Prinzipiell ist dieses Verhalten auch bereits bei dezidierten topo-bathymetrischen Laserscannern mit online prozessierten Wellenformen beobachtet worden (MANDLBURGER et al. 2015).

Abb. 6b zeigt einen Profilschnitt durch einen Schotterteich mit geringer Trübung. Bodenpunkte können in diesem Fall bis zu einer Tiefe von etwa 3.6 m erkannt werden, allerdings ist auch hier lediglich der Uferbereich dicht mit Gewässerbodenpunkten besetzt und die Wasseroberfläche weist nur sporadische Echos auf, sodass deren Modellierung nur unsicher möglich ist. Für stehende Gewässer ist durch den Einsatz von statistischen Methoden und räumlicher Aggregation eine Schätzung der Wasserspiegellage in dm-Genauigkeit möglich (MANDLBURGER & JUTZI 2019). Für Fließgewässer mit geneigten Wasseroberflächen sowie für Küstengewässer bei entsprechendem Wellengang hingegen führt räumliche Aggregation zu lokalen Punktlagefehlern der Gewässerbodenpunkte infolge des fehlenden Detailreichtums der Oberflächenbeschreibung.

## 5 Zusammenfassung und Ausblick

In diesem Beitrag wurden die Möglichkeiten und Grenzen der Bathymetrieerfassung mittels Single Photon LiDAR aus Sicht der Theorie und anhand von konkreten Messergebnissen diskutiert. Prinzipiell kann festgehalten werden, dass Single Photon LiDAR bathymetrische Fähigkeiten aufweist, die sich aufgrund folgender Eigenschaften ergeben: (i) Verwendung von wasserdurchdringender grüner Laserstrahlung ( $\lambda=532$  nm), (ii) Verwendung eines konischen Scanmechanismus mit konstanter Neigung gegenüber der Nadirrichtung von  $10-20^\circ$ , dem allgemein anerkannten Optimum für bathymetrische Anwendungen, und (iii) Einsatz eines mehrzielfähigen Detektors für jeden einzelnen Teilstrahl bestehend aus einem Array von Einzelphoton-sensitiven Zellen. Als einschränkende Parameter sind zu nennen: (i) technologisch bedingte, mangelnde Aufzeichnung der vollen Wellenform, (ii) damit verbunden keine Möglichkeit der gewässerspezifischen Wellenformanalyse und (iii) potenzielles Empfangen des durch Volumenstreuung im Wasser breiter rückgestreuten Signals von benachbarten Teilstrahlen.

Die Analyse der 3D-Punktwolken einer konkreten Messkampagne in Wien (Donau, Neue Donau, Alte Donau, künstliche Teiche) und Niederösterreich (Pielach Fluss, Schotterteiche) hat ergeben, dass die zu erwartende Eindringtiefe etwa im Bereich der 1-fachen Secchi Tiefe liegt und die Erfassung der Wasseroberfläche stark von der Trübung des Gewässers beeinflusst zu sein scheint. Mit dem vorliegenden Datenmaterial konnte keine Gegenüberstellung der Genauigkeiten und tatsächlich erzielbaren Messtiefen angestellt werden. Dies wird Gegenstand künftiger Untersuchungen sein. Es kann aber abschließend festgestellt werden, dass der SPL100 Sensor neben seiner intendierten Verwendung als Laserscanner für großflächige Topografieerfassung auch als Nebenprodukt Tiefen von seichten und klaren Gewässern liefert. Genauigkeit und Eindringtiefe sind allerdings gegenüber dezidierten Bathymetriescannern eingeschränkt.

## 6 Danksagung

Die vorliegende Arbeit wurde durch Mittel der Deutschen Forschungsgesellschaft (DFG) im Rahmen des Forschungsprojektes „Bathymetrieermessung durch Fusion von Flugzeuglaserscanning und multispektralen Luftbildern“ unterstützt. Der Autor dankt der Stadt Wien, Magistratsabteilung 41 (Stadtvermessung Wien) für die Bereitstellung der SPL Punktwolken.

## 7 Literaturverzeichnis

- AGISHEV, R., COMERÓN, A., BACH, J., RODRIGUEZ, A., SICARD, M., RIU, J. & ROYO, S., 2013: Lidar with SiPM: Some capabilities and limitations in real environment. *Optics and Laser Technology*, **49**, 86-90.
- ALLOUIS, T., BAILLY, J.-S., PASTOL, Y. & ROUX, C. L., 2010: Comparison of LiDAR waveform processing methods for very shallow water bathymetry using raman, near-infrared and green signals. *Earth Surface Processes and Landforms* **35**(6), 640-650.
- CLIFTON, W., STEELE, B., NELSON, G., TRUSCOTT, A., ITZLER, M. & ENTWISTLE, M., 2015: Medium altitude airborne geiger-mode mapping LIDAR system. *Proc. SPIE*, **9465**.
- DEGNAN, J., 2002: Photon-counting multikilohertz microlaser altimeters for airborne and spaceborne topographic measurements. *Journal of Geodynamics* **34**(3-4), 503-549.
- DEGNAN, J., 2016: Scanning, Multibeam, Single Photon Lidars for Rapid, Large Scale, High Resolution, Topographic and Bathymetric Mapping. *Remote Sensing* **8**(11), 923-958.
- DEGNAN, J., 2018: Airborne And Spaceborne Single Photon 3D Imaging Lidars. In *Proceedings of 21<sup>st</sup> International Workshop on Laser Ranging*, Canberra, Australia.
- FERNANDEZ-DIAZ, J., CARTER, W., GLENNIE, C., SHRESTHA, R., PAN, Z., EKHTARI, N., SINGHANIA, A., HAUSER, D. & SARTORI, M., 2016: Capability Assessment and Performance Metrics for the Titan Multispectral Mapping Lidar. *Remote Sensing* **8**(936), 1-34.
- GUENTHER, G. C., CUNNINGHAM, A. G., LAROQUE, P. E. & REID, D. J., 2000: Meeting the accuracy challenge in airborne lidar bathymetry. *20th EARSeL Symposium: Workshop on Lidar Remote Sensing of Land and Sea*, Dresden.
- KIM, S., LEE, I. & KWON, Y. J., 2013: Simulation of a geiger-mode imaging LADAR system for performance assessment. *Sensors* **13**(7), 8461–8489.
- LEICA, 2020: Produktbeschreibung des Leica SPL100 LiDAR Sensors, <https://leica-geosystems.com/de-at/products/airborne-systems/topographic-lidar-sensors/leica-spl100>
- MALLET C. & BRETAR, F., 2009: Full-waveform topographic lidar: State-of-the-art. *ISPRS Journal of Photogrammetry and Remote Sensing* **64**(1), 1-16.
- MANDLBURGER, G., LEHNER, H. & PFEIFER, N., 2019: A Comparison of Single Photon and Full Waveform LiDAR. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **IV-2/W5**, 397-404.
- MANDLBURGER, G. & JUTZI, B., 2019: On the Feasibility of Water Surface Mapping with Single Photon LiDAR. *ISPRS International Journal of Geo-Information* **8**(188), 1-23.
- PFENNIGBAUER, M., WOLF, C., WEINKOPF, J. & ULLRICH, A., 2014: Online waveform processing for demanding target situations. In *Proc. SPIE*, **90800J**.
- PFEIFER, N., MANDLBURGER, G. & GLIRA, P., 2016: Laserscanning. *Handbuch der Geodäsie: Photogrammetrie und Fernerkundung*, C. Heipke (Hrsg.), Springer, Berlin Heidelberg, 431-481.
- ROMANO, M., 2015: Geiger Mode LiDAR. *Proceedings of the ASPRS IGTF Conference*, ISBN: 9781510804579.
- SCHWARZ, R., MANDLBURGER, G., PFENNIGBAUER M. & PFEIFER, N., 2019: Design and evaluation of a full-wave surface and bottom-detection algorithm for LiDAR bathymetry of very shallow waters. *ISPRS Journal of Photogrammetry and Remote Sensing* **150**, 1-10.
- STOKER, J. M., ABDULLAH, Q. A., NAYEGANDHI, A. & WINEHOUSE, J., 2016: Evaluation of Single

Photon and Geiger Mode Lidar for the 3D Elevation Program. *Remote Sensing* **8(9)**, 716-767.

ULLRICH, A. & PFENNIGBAUER, M., 2016: Linear LIDAR versus Geiger-mode LIDAR: impact on data properties and data quality. In *Proc. SPIE* **9832**, 983204-983217.

# Erfassung der Hauptreflektordeformation eines Radioteleskops durch UAV-gestützte Nahbereichsphotogrammetrie

ANSGAR GREIWE<sup>1</sup>, RAINER BRECHTKEN<sup>1</sup>, MICHAEL LÖSLER<sup>2</sup>,  
CORNELIA ESCHELBACH<sup>2</sup> & RÜDIGER HAAS<sup>3</sup>

*Zusammenfassung: Deformationen am Hauptreflektor eines Radioteleskops verursachen systematische Abweichungen in den Messungen und führen zu einer Verfälschung der abgeleiteten globalen Position und somit zu einem verzerrten Netzmaßstab im globalen geodätischen Bezugsrahmen. Um diese Systematiken zu minimieren, ist ein stabiler bzw. modellierbarer Strahlengang über den gesamten Arbeitsbereich des Teleskops notwendig. Zur messtechnischen Erfassung des Hauptreflektors konventioneller Radioteleskope werden gegenwärtig häufig Laserscanner verwendet, die eine Einzelpunktgenauigkeit von wenigen Millimetern erreichen.*

*Für diesen Beitrag wurde erstmals das elevationsabhängige Verformungsverhalten eines modernen, kompakten VGOS-spezifizierten VLBI-Radioteleskops photogrammetrisch untersucht. Um den Hauptreflektor des Teleskops in mehreren Neigungen zu erfassen, erfolgte die photogrammetrische Datenerhebung durch ein Unmanned Aircraft Vehicle (UAV) als Sensorplattform für eine leichtgewichtige Kompaktkamera. Im Rahmen der Untersuchungen wurde eine Einzelpunktgenauigkeit von ca. 50 µm in 21 unterschiedlichen Bildverbänden erreicht.*

## 1 Einleitung

Das Verfahren zur Interferometrie auf langen Basislinien (VLBI) zählt zu den geodätischen Raumtechniken, die zur Realisierung eines erdfesten geodätischen Bezugsrahmens (bspw. International Terrestrial Reference Frame, ITRF) herangezogen werden. Radioteleskope bilden dabei die erdfesten Sensoren, die sich durch Drehungen um eine Azimut- und Elevationsachse auf Radioquellen, z. B. Quasare, ausrichten lassen. Durch Korrelation der von mehreren Radioteleskopen gleichzeitig aufgezeichneten Radiosignale der Quasare lassen sich u. a. die interkontinentalen Raumstrecken zwischen den Teleskopen mit Millimetergenauigkeit ableiten.

Im Rahmen der Agenda VLBI2010 wurden durch NIELL et al. (2006) Spezifikationen für eine neue Generation von Radioteleskopen mit kompakter Bauweise und erhöhter Ausrichtungsgeschwindigkeit formuliert. Die Erweiterung des internationalen VLBI-Netzes durch diesen neuen Teleskoptyp soll zu der durch GGOS (Global Geodetic Observing System) angestrebten Steigerung der Positionsgenauigkeit im globalen geodätischen Bezugsrahmen führen (vgl. ROTHACHER et al. 2009), und wird häufig als VLBI Global Observing System (VGOS) bezeichnet. PETRACHENKO et al. (2009) spezifizieren die Mindestanforderungen für VGOS-Teleskope, um die Systemsensitivität zu erhöhen und die angestrebte Positionsgenauigkeit von 1 mm im globalen terrestrischen Referenzrahmen zu erreichen. Hierzu zählt auch ein stabiler bzw. modellierbarer Strahlengang über

---

<sup>1</sup> Hochschule Bochum, Labor für Photogrammetrie & Laserscanning, Lennerhofstr. 140, D-44801 Bochum, E-Mail: [Ansgar.Greiwe, Rainer.Brechtken]@hs-bochum.de

<sup>2</sup> Frankfurt University of Applied Sciences, Labor für Industrielle Messtechnik, E-Mail: [Michael.Loesler, Cornelia.Eschelbach]@fb1.fra-uas.de

<sup>3</sup> Chalmers University of Technology, Gothenburg, SE, E-Mail: rudiger.haas@chalmers.se

den gesamten Elevationsbereich des Teleskops, sodass für alle Teleskoporientierungen die Signalwegänderungen  $300 \mu\text{m}$  (RMS) nicht überschreiten. Im Gegensatz zu konventionellen Radioteleskopen, die Brennweitenvariation von z. T. mehreren Zentimetern aufweisen (z. B. SARTI et al. 2009), liegen die zu erwartenden Deformationen aufgrund der kompakten Bauweise für VGOS-Teleskope im unteren Millimeterbereich. Messtechnische Untersuchungen von VGOS-Teleskopen fehlen jedoch bisher.

Deformationen am Hauptreflektor werden neben Temperaturänderungen oder Wind vor allem durch Lastfalländerungen hervorgerufen und wirken sich auf die Struktur und damit auf die Empfangseigenschaften der Teleskope aus. Durch die Verlagerung des Eigengewichts der Antenne während des VLBI-Messprogramms wird der Antennenbrennpunkt in Abhängigkeit von der jeweiligen Elevationsstellung des Hauptreflektors verändert. Mit einer geeigneten Messmethode ist somit das elevationsabhängige Verformungsverhalten eines Teleskops zu erfassen und die Brennpunktvariation zu bestimmen.



Abb. 1: Nordöstliches Onsala Twin Teleskop (ONSA13NE) mit Hexacopter (Hexapilots HP-TS960) während der Befliegung in der  $0^\circ$  Elevationsposition. Im Hintergrund erkennbar: Südliches Teleskop in  $90^\circ$ -Stellung (Opferschalenstellung)

Für diese Studie wurde erstmals das Deformationsverhalten eines VGOS-spezifisierten Teleskops untersucht. Als Untersuchungsobjekt diente eines der 2017 in Betrieb genommenen Onsala Twin Teleskope (OTT; s. Abb. 1) am Onsala Space Observatory in Schweden (HAAS 2013). Das Observatorium befindet sich 40 km südlich von Göteborg. Die beiden baugleichen Radioteleskope stehen in einem Abstand von ca. 70 m zueinander, direkt an der Küste des Kattegats. Der Hauptreflektor hat die Form eines sogenannten Ring-Focus-Paraboloids und besitzt einen Durchmesser von 13,2 m. Der Scheitelpunkt des Hauptreflektors befindet sich in etwa auf 10 m Höhe über dem

Grund. Angestrebt wurde die Untersuchung der Antennengeometrie bei Lastfalländerung, also bei Änderung der Teleskopausrichtung zwischen  $0^\circ$  und  $90^\circ$  Elevation.

## 2 Erfassung der Reflektoroberfläche

Um die Verformung der Hauptreflektoroberfläche in unterschiedlichen Elevationsstellungen zu erfassen, könnte ein terrestrischer Laserscanner (TLS) nahe des Subreflektors an dessen Aufhängung kardanisich befestigt werden (HOLST et al. 2017; BERGSTRAND et al. 2019). Durch dieses Aufnahmeverfahren wird die Oberfläche des Hauptreflektors mit einer Genauigkeit von ca. 1,5 mm erfasst, hervorgerufen durch zufällige und systematische Anteile (HOLST et al. 2017). Nachteil dieses Verfahrens ist die nicht vollständige Erfassung der Antennenoberfläche, da der Tubus mit der Empfangseinheit große Abschattungen auf der zu scannenden Fläche erzeugt (Abb. 1).

Aufgrund der hohen Genauigkeitsanforderungen für diese VGOS-Radioteleskope wurde für diese Untersuchungen die photogrammetrische Erfassung als alternatives Aufnahmeverfahren gewählt. Vorteil der photogrammetrischen Aufnahme gegenüber einer Aufnahme per TLS ist die vollständige Erfassung der Oberfläche durch die wechselnden Kamerastandpunkte, die durch die Aufnahmekonfiguration des photogrammetrischen Blockverbandes vorgegeben werden. Somit werden auch Oberflächenpunkte unterhalb etwaiger Streben am Subreflektor erfasst.

Die Nahbereichsphotogrammetrie erlaubt - unter optimalen Voraussetzungen (metrische Kamerasysteme, optimale Aufnahmekonfiguration) - Einzelpunktgenauigkeiten von  $5 \mu\text{m} + 5 \text{ ppm}$  zu erreichen (LUHMANN 2018). Dieses Messverfahren wird aus diesem Grund auch eingesetzt, um die einzelnen Teleskoppaneele des Hauptreflektors formschlüssig beim Aufbau des Teleskops vor Ort zu justieren (Abb. 2).



Abb. 2: Einsatz eines Krans zur Montage und Ausrichtung der Paneele

Für das untersuchte Teleskop wurden die Paneele durch den Teleskophersteller in einer Elevationsstellung von  $34^\circ$  nach der Montage abschließend eingemessen und so der Nachweis für die Einhaltung der Sollgeometrie erbracht. Bei dieser Messung werden Abweichungen von der Sollgeometrie in anderen Elevationsstellungen des Hauptreflektors nicht ermittelt.

## 2.1 UAV gestützte Aufnahme

Um die erforderlichen Kamerastandpunkte für eine photogrammetrische Aufnahme zu realisieren, ist bei geringen Elevationen des Hauptreflektors die Verwendung eines Krans denkbar, wie in Abb. 2 dargestellt. Die Aufnahmekonfiguration entspräche in etwa der einer Fassadenaufnahme. Wird der Hauptreflektor in einer Elevation von mehr als  $40^\circ$  geneigt, ist die Positionierung per Kran für einen Rundumverband als Aufnahmekonfiguration herausfordernd.

Um Aufnahmen für den gesamten Elevationsbereich des Teleskops zu erhalten, wurde erstmals ein Hexacopter eingesetzt, der mit einem Gesamtgewicht von ca. 10 kg und einem Durchmesser von ca. 1,8 m in der Lage ist, Nutzlasten von bis zu 2500 g zu tragen (Abb. 3).



Abb. 3: Hexapilots HP-TS960 (links) und für die Aufnahmen eingesetzter Kameragimbal (rechts) zur Orientierung der Kamera auf das Messobjekt

Das UAV wird durch eine Funksteuerung (2,4 GHz) durch den Piloten manuell gesteuert. Das Kamerabild kann durch einen Monitor am Boden betrachtet werden und die Auslösung der Aufnahmen manuell per Funksteuerung erfolgen. Zudem verfügt der Copter über einen halbautonomen Flugmodus, bei dem die Flugtrajektorie in Form von Wegpunkten, die Fluggeschwindigkeit und die geplanten Auslösezeitpunkte für die Aufnahmen durch ein Wegpunktskript in Form einer Textdatei vorgegeben werden. Für die vorliegende Untersuchung wurde die Flugtrajektorie vorab geplant und automatisiert durch den Copter unter den definierten Geschwindigkeiten geflogen. Aufgrund der schwierigen Windverhältnisse am Messobjekt und der kurzen Aufnahmedistanzen erfolgte die Ausrichtung und Auslösung der Kamera manuell durch den Piloten.

Das System ist modular aufgebaut und lässt sich individuell konfigurieren. Halterung und Kamera können problemspezifisch und insbesondere herstellerunabhängig eingesetzt werden. Ausgestattet mit zwei 10.000 mAh Akkus beträgt die Flugzeiten für eine Mittelformatkamera (2500 g) ca. 15 Minuten. Eine Kompaktkamera (Sensor im APS-C Format) ermöglicht durch ihr geringes Gewicht von ca. 400 g mit entsprechender Halterung eine Flugzeit von 25 Minuten.

## 2.2 Kamerasystem

Für ein Kamerasystem zur Nahbereichsphotogrammetrie existieren unterschiedliche Anforderungen im Hinblick auf die optische Abbildungsgüte, geometrische Stabilität, Temperaturunempfindlichkeit, Auflösungsvermögen und Schockresistenz. Insbesondere eine feste Sensor-Objektiv-Verbindung ist unabdingbar, um gute Ergebnisse zu erzielen. Diese Forderungen werden sicherlich

von vielen Systemen am Markt erfüllt, lediglich das Gewicht dieser Kameras ermöglicht kaum Flugzeiten von 25 Minuten mit dem eingesetzten UAV.

Die Forderung nach einer festen Sensor-Objektiv-Verbindung erfüllt noch am besten die Merrill-Baureihe des Herstellers Sigma. Diese Kamera wird mit einer fest montierten, nicht wechselbaren Festbrennweiten-Optik ausgeliefert. Das Objektiv verfügt bis auf die Fokussierlinse über keine variablen Bauteile (Objektivtubus fest, s. Abb. 4). Ebenso fehlt der Kamera ein Bildstabilisator. Objektiv und Sensor sind fest mit dem Gehäuse verbunden und führen zu einem stabilen Bildhauptpunkt. In einer Serie von Testfeldkalibrierungen an fünf unterschiedlichen Tagen wurde für die eingesetzte Kamera eine Standardabweichung der Kamerakonstante von  $4,8 \mu\text{m}$  und für den Bildhauptpunkt von  $1,1 \mu\text{m}$  in x- sowie  $2,2 \mu\text{m}$  in y-Richtung ermittelt. Aufgrund dieser Werte wurde im Rahmen der Erfassung vor Ort davon ausgegangen, dass es sich zumindest um eine geometrisch stabile Kamera zwischen einem Ein- und Ausschaltvorgang handelt (Fokussierlinse wird beim Einschalten bewegt).



Abb. 4: Der Objektivtubus (links) der Sigma DP3 Merrill ist fest mit dem Gehäuse verbaut

Im Hinblick auf das geforderte Auflösungsvermögen der Kamera sind sowohl Objektiv als auch Sensor maßgebliche Komponenten. Das Objektiv ist bei der Merrill Baureihe vom Hersteller speziell auf die Geometrie (Auflagemaß) der Kamera berechnet. Bereits bei Offenblende verfügt die Kamerareihe über ein sehr gutes Abbildungsverhalten (SPRECKELS et al. 2016). Zudem verfügt das Objektiv über einen Zentralverschluss und nicht über einen Rolling-Shutter, der insbesondere bei den hochfrequenten Schwingungen am Copter, hervorgerufen durch die Propeller, die Bildmessgenauigkeit beeinflussen würde.

Für den Mikrokontrast ist zudem der Sensor maßgeblich. Die verwendete Merrill-Baureihe verfügt über einen Foveon-Sensor in APS-C Größe ( $16 \text{ mm} \times 24 \text{ mm}$ ) anstelle eines Bayer-Patterns, der in allen sonstigen CCD- oder CMOS-Kamerasensoren verwendet wird (GREIWE & GEHRKE 2013a). Die Farbfilteranordnungen eines Bayer-Patterns erfordert Farbbinterpolationen mit Nachbarpixeln, um die volle RGB-Farbinformation zu erhalten. Im Gegensatz hierzu erfasst ein Foveon-Sensor für jedes Pixel die volle RGB-Farbinformation ohne Interpolation (vgl. GREIWE & GEHRKE, 2013b). Durch diese Sensortechnik weisen die Bilder einen besseren Mikrokontrast auf und ermöglichen folglich auch eine genauere Bildmessung der Zielmarken.

### 2.3 Signalisierung

Für die photogrammetrische Aufnahme wurde das Objekt mit 72 Zielmarken signalisiert, welche gleichmäßig auf der Hauptreflektoroberfläche verteilt waren. Zum Einsatz kamen 12 bit codierte Schwarz-Weiß-Marken, auf Haftetiketten mit einem Durchmesser von ca. 15 cm (Abb. 5). Die Anordnung der Zielmarken erfolgte in drei konzentrischen Kreisen mit  $2 \times 24$  und  $1 \times 12$  Elementen. Somit wurde pro Paneel des Hauptreflektors mindestens eine Zielmarke angebracht. Weitere 12 Marken waren in der Mitte des Hauptreflektors im Bereich des Tubus und vier Zielmarken am Subreflektor angebracht. Zudem wurde ein Kreuz mit sechs codierten Marken am Subreflektor positioniert.



Abb. 5: Zielmarken am Hauptreflektor, Kreuz mit Passpunkten zur Lagerung sowie Maßstäbe vorne am Subreflektor, an den Streben und am Rand des Hauptreflektors

Die Koordinaten der sechs Punkte wurden zuvor im Labor photogrammetrisch bestimmt und als Datumspunkte zur Definition des lokalen Koordinatensystems der 3D Punkte der jeweiligen Bündelblockausgleichung verwendet. Die Maßstabsinformation für die Auswertung wurde durch insgesamt sechs kalibrierte Kohlefasermaßstäbe mit einer Länge von ca. 1,3 m realisiert. Die Maßstäbe umschließen das Messvolumen, indem zwei an den Streben zum Subreflektor, drei am Rand des Hauptreflektors und einer am Subreflektor angebracht wurden.

### 2.4 Aufnahmekonfiguration

Für die Aufnahmen wurde eine Pixelgröße von 1 mm am Objekt angezielt. Dies führt bei dem zu Verfügung stehenden Foveon Chip mit einem Pixelpitch von  $5 \mu\text{m}$  zu einem Bildmaßstab von 1:200. Um aus flugtechnischer Sicht eine größtmögliche Aufnahmedistanz zu erreichen, wurde die DP3 mit einer Brennweite von 50 mm anstelle der alternativ zur Verfügung stehenden DP2

(Brennweite 30 mm) gewählt. Mit Blende 8.0 wurde das Objektiv dreifach abgeblendet und auf 10 m Aufnahmeentfernung fokussiert. Dies führte zu einer Schärfentiefe von ca. 21 m. Der vordere Schärfe­punkt lag bei ca. 6 m, der hintere Schärfe­punkt bei ca. 27 m Aufnahmeentfernung. Die Belichtungszeit wurde aufgrund der kleinen Blende mit 1/640 s. eingestellt. Die verwendete Fluggeschwindigkeit von 0,2 m/s führte bei dieser Verschlusszeit zu einer Bewegungsunschärfe von ca. 1/3 Pixel während der Belichtung.

Die eingesetzte Kamera ist zwar leichtgewichtig, aber nicht über einen längeren Zeitraum geometrisch stabil. Da die Stromversorgung der Kamera durch den Hexacopter erfolgte, war es erforderlich, die photogrammetrische Erfassung einer Elevationsposition des Hauptreflektors in einem Flug ohne Zwischenlandung bzw. Akkuwechsel durchzuführen.

Resultierend aus der maximalen Flugzeit von 25 Minuten und der maximalen Fluggeschwindigkeit von 0,2 m/s wurde eine Flugkonfiguration bestehend aus zwei horizontalen Flugbahnen vor dem Subreflektor und zwei Kreisen geplant (Abb. 6). Die kreisrunden Flugbahnen mit einem Radius von 6,5 m bzw. 11,5 m waren um den Elevationswinkel des Teleskops geneigt und lagen zentrisch um die optische Achse des Teleskops. Die kürzeste Aufnahme­distanz (Abstand zwischen Fluglinien und Subreflektor) betrug ungefähr 7 m, die weiteste Distanz 24 m (Abstand zwischen äußeren Kreis und Hauptreflektor).

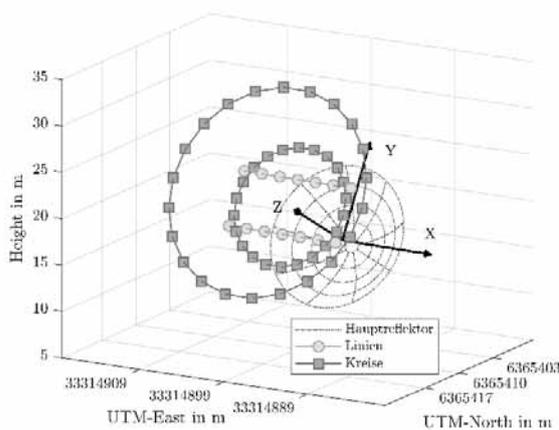


Abb. 6: Geplante (links) und tatsächlich geflogene Aufnahme­konfiguration für eine Elevationsposition

Die Kamera wurde jeweils manuell auf die Teleskopoberfläche ausgerichtet und ausgelöst. Mit einer Bildfolgezeit von ca. 5 Sekunden führte dies zu einem Aufnahme­abstand von ca. 1 m. Kreisbahnen und horizontale Linien führen zu einem Gesamtflugweg von ca. 150 m. Somit sind für einen Flug mindestens 150 Aufnahmen möglich.

Das Teleskop wurde zwischen 0° und 90° in 10° Schritten in Elevation verfahren. Alle zehn Positionen wurden zweimal erfasst, wobei zur Steigerung der Unabhängigkeit der Ergebnisse die Wiederholungsmessung nie direkt im Anschluss erfolgte. Ergänzt wurde die Datenreihe durch einen Datensatz in 34° Elevation, die der Justierposition des Herstellers entspricht. Die ganze Messkampagne lieferte insgesamt 21 getrennt voneinander auswertbare Datensätze.

## 2.5 Auswertung und Ergebnisse

Die einzelnen Bildverbände wurden mit AICON 3D Studio ausgewertet. Pro Epoche standen für die Auswertung nach Löschung unscharfer Bilder ca. 130 Aufnahmen zur Verfügung (mindestens 117, in zwei Fällen 227).

Wie in Abb. 7 ersichtlich, sind drei Flüge mit mehr als 200 Aufnahmen durchgeführt worden. Hierbei handelt es sich um drei am ersten Flugtag durchgeführten Flüge bei 10°, 20° und 90° Elevation, bei denen die zu Verfügung stehende Flugzeit von etwa 25 Minuten maximal ausgenutzt wurde. Am zweiten Flugtag wurden die restlichen Elevationsstellungen des Teleskops befliegen und die in Abbildung 6 dargestellte Flugkonfiguration strikt eingehalten. Dies ermöglichte die Reduktion der Flugzeit um ca. fünf Minuten auf ca. 20 Minuten Gesamtflugzeit und die Aufnahme der geplanten 150 Aufnahmen.

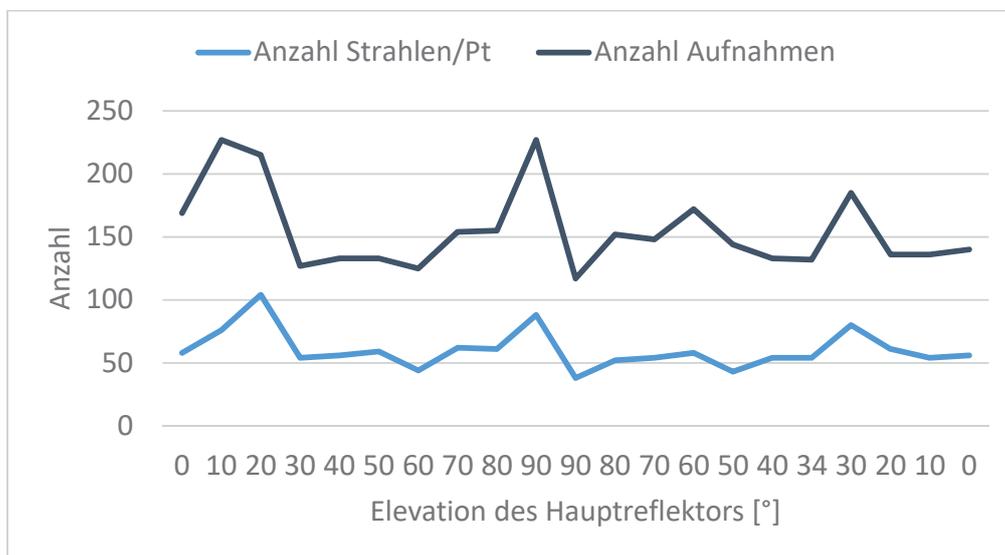


Abb. 7: Anzahl der Aufnahmen und Strahlen pro Punkt

Die Bündelblockausgleichung wurde für diese Untersuchung nur mit den signalisierten Zielmarken durchgeführt. Zusätzliche von AICON erkannte Punkte wurden nicht in die Auswertung aufgenommen, da es sich teilweise um fehlinterpretierte Panneelzwischenräume oder versenkte Schraubenköpfe handelte. In der Regel wurde jede Zielmarke in ca. 50 Aufnahmen, mindestens jedoch in 38 Aufnahmen gemessen.

Für jeden Bildverband wurde die Kamerakalibrierung in situ durchgeführt und pro Bildverband ein zugehöriges Kameramodell im Zuge der Bündelblockausgleichung bestimmt.

Die Standardabweichung in Aufnahme­richtung (z) war erwartungsgemäß am höchsten, wie aus der Größe und Orientierung der resultierenden Konfidenz­ellipsoide der signalisierten Ziel­marken nach der Bündelblockausgleichung in Abb. 8 ersichtlich.

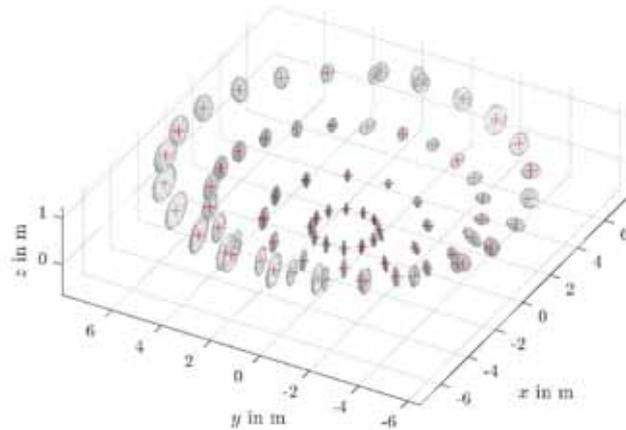


Abb. 8: Resultierende Konfidenz­ellipsoide nach der Bündelblockausgleichung für die signalisierten Ziel­marken auf der Hauptreflektoroberfläche

Das quadratische Mittel der Standardabweichung eines Einzelpunktes in z-Richtung betrug in fast allen Epochen weniger als 50  $\mu\text{m}$ , die maximale Standardabweichung weniger als 100  $\mu\text{m}$ . Nicht alle der 21 Flugmissionen konnten optimal durchgeführt werden, da am Teleskop schwierige Windverhältnisse (Windschatten/Wirbelschleppen) herrschten. Zudem musste die Ausrichtung der Kamera in jeder Elevationsstellung neu angepasst werden, da die 90°-Stellung einem Flug mit Nadiraufnahmen, die 0° Stellung dem Anwendungsfall der Fassadenaufnahme entspricht. Insbesondere die Elevationen ab 40° bis 80° erfordern einige Übung des Piloten, somit sind die Genauigkeiten in der zweiten Befliegung dieser Elevationen bei annähernd gleicher Anzahl der Aufnahmen höher, wie in Abb. 9 im rechten Teil des Diagramms mit abnehmenden Elevationswerten dargestellt.

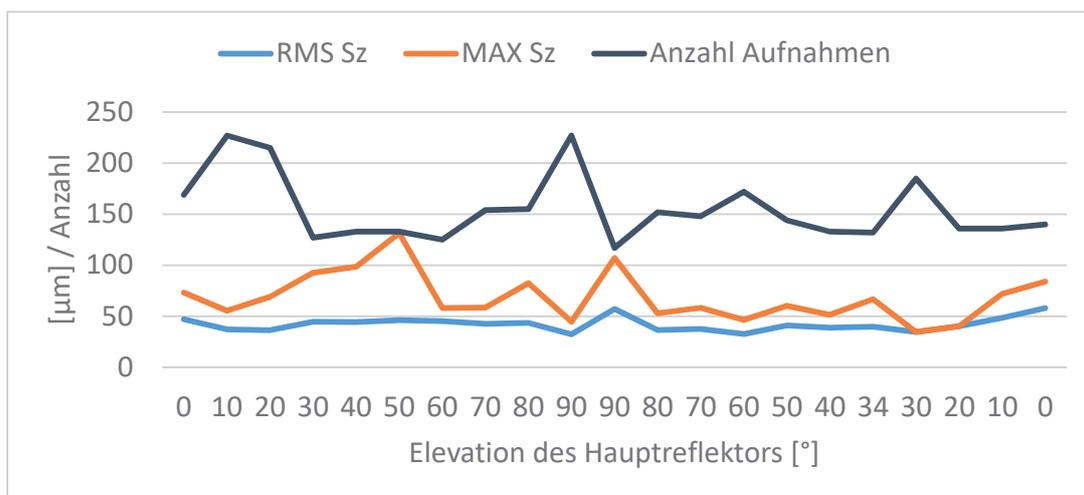


Abb. 9: Quadratisches Mittel und Maximale Standardabweichung der Oberflächenpunkte in z-Richtung

Die Abnahme der Genauigkeiten in den niedrigen Elevationsstellungen des Teleskops lassen sich durch die Aufnahmekonfiguration erklären. In den niedrigen Elevationsstellungen konnte keine optimale Blockkonfiguration geflogen werden (Abb. 10). Die geplante Aufnahmekonfiguration (Kreise, Abb. 6) hätte aufgrund der niedrigen Flughöhe vor der nahezu senkrecht stehenden Hauptreflektoroberfläche entweder zu negativen oder sehr geringen Flughöhen geführt. Vollständige Kreise konnten ab einer Elevationsstellung von  $30^\circ$  geflogen werden, in niedrigeren Elevationsstellungen führte eine Mindestflughöhe zu Halbkreisen, was zu schlechterer Punktgenauigkeit bei vergleichbarer Aufnahmeanzahl führte.

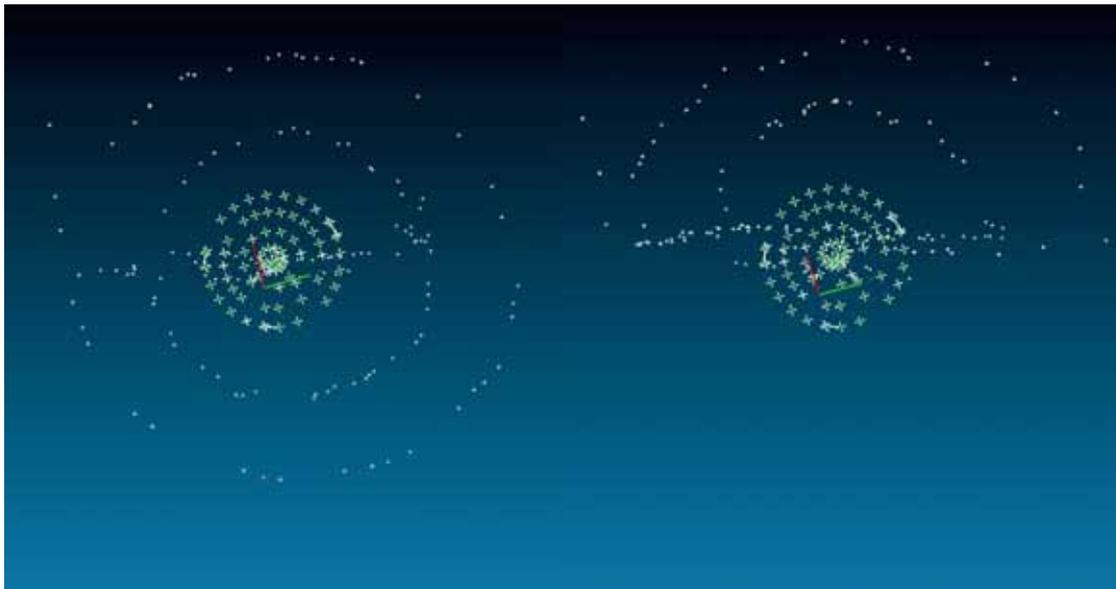


Abb. 10: Flugkonfiguration in Aufnahme­richtung für  $30^\circ$  und  $0^\circ$  Elevationsstellung

### 3 Fazit & Ausblick

Die Erfassung der Oberflächenpunkte für eine Deformationsanalyse ist erstmals mit einem UAV erfolgreich durchgeführt worden. Hierbei kam eine geeignete Kompaktkamera als Sensor zum Einsatz, welche durch das geringe Gewicht die hohe Anzahl an Aufnahmen ermöglichte (Flugzeit) und trotzdem eine hohe Bildqualität und eine zumindest kurzzeitige geometrische Stabilität während der Aufnahme aufweist. Die geforderten Punktgenauigkeiten sind in allen 21 Flugkampagnen konstant erreicht worden.

Limitierende Faktoren sind hierbei die Aufnahmekonfiguration in niedrigen Elevationsstellungen und die Anzahl der Aufnahmen pro Bildverband. Ersteres kann in Zukunft durch eine optimierte Aufnahmekonfiguration erreicht werden. Die Anzahl der Aufnahmen pro Verband sind entweder durch eine längere Flugzeit oder durch eine Kamera mit schnellerer Bildfolgezeit möglich, die allerdings den geschilderten Anforderungen (Gewicht, Stabilität, Optik) genügen muss.

Für jeden Datensatz konnte ein Ring-Fokus-Paraboloid geschätzt werden. Abbildung 11 zeigt die daraus abgeleiteten Brennweiten in Abhängigkeit der Elevationsstellung des Teleskops. Die größte Brennweite liegt erwartungsgemäß bei einer Elevation von  $90^\circ$ , da hier der Antennenrand von der Schwerkraft nach unten gezogen und die Antenne aufgeweitet wird. Mit dem Elevationswinkel

fällt auch die Brennweite um bis zu 2,3 mm. Diese Spannweite ist um mehr als das 10-fache kleiner als in der Literatur dokumentierte Variationen konventioneller Radioteleskope (z. B. SARTI et al. 2009).

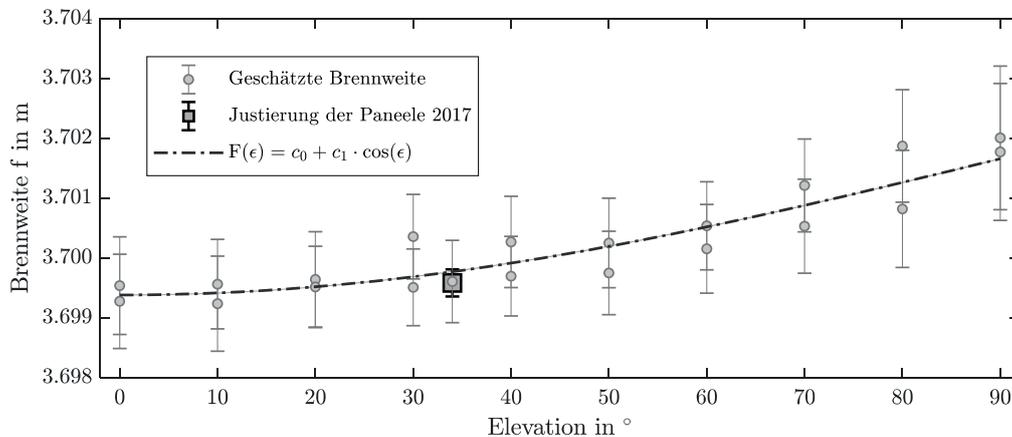


Abb.11: Geschätzte Brennweiten in Abhängigkeit der Teleskopelevation; der funktionale Zusammenhang zwischen Elevation und Brennweite ist als gestrichelte Linie dargestellt, das Unsicherheitsband beträgt  $2\sigma$ ; ein schwarzer Stern symbolisiert das Ergebnis aus LÖSLER et al. (2019).

## 4 Dank

Die Autoren danken Lars Wennerbäck und Christer Hermansson vom Onsala Space Observatory für ihre Unterstützung bei der Präparierung des Radioteleskops.

## 5 Literaturverzeichnis

- BERGSTRAND, S., HERBERTSSON, M., RIECK, C., SPETZ, J., SVANTESSON, C.G. & HAAS, R., 2019: A gravitational telescope ldeformation model for geodetic VLBI. *Journal of Geodesy* **93**(5), 669-680. doi:10.1007/s00190-018-1188-1.
- GREIWE, A. & GEHRKE, R., 2013a: Foveon Chip oder Bayer Pattern – geeignete Sensoren zur Aerophotogrammetrie mit UAS. In: Luhmann, T., Schumacher, C. (Hrsg.): *Photogrammetrie - Laserscanning – Optische 3D-Messtechnik: Beiträge der 12. Oldenburger 3D-Tage 2013*, Wichmann, Offenbach, 334-343. ISBN: 978-3879075287.
- GREIWE, A. & GEHRKE, R., 2013b: Kameras zur 3D-Modellierung mit UAS. In: Strobl, J., Blaschke, T., Griesebner, G., Zagel, B. (Hrsg.): *Angewandte Geoinformatik 2013: Beiträge zum 25. AGIT-Symposium Salzburg*, Wichmann, Offenbach, 41-46.
- HAAS, R., 2013: The Onsala twin telescope project. In: Zubko, N., Poutanen, M. (Hrsg.): *Proceedings of the 21st European VLBI for Geodesy and Astrometry (EVGA) Working Meeting*, 61-66. ISBN: 978-9517112963.
- HOLST, C., SCHUNCK, D., NOTHNAGEL, A., HAAS, R., WENNERBÄCK, L., OLOFSSON, H., HAMMARGREN, R. & KUHLMANN, H., 2017: Terrestrial Laser Scanner Two-FaceMeasurements for Analyzing the Elevation-Dependent Deformation of the Onsala Space Observatory 20-

- m Radio Telescope's Main Reflector in a Bundle Adjustment. *Sensors* **17**(8), 1833, 1-21. doi:10.3390/s17081833.
- LÖSLER, M., HAAS, R., ESCHELBACH, C. & GREIWE, A., 2019: Gravitational Deformation of Ring-Focus Antennas for VGOS - First Investigations at the Onsala Twin Telescopes Project. *Journal of Geodesy* **93**(10), S. 2069-2087. doi: 10.1007/s00190-019-01302-5.
- LUHMANN, T., 2018: Nahbereichsphotogrammetrie - Grundlagen, Methoden und Anwendungen, Wichmann, Berlin, 4. Auflage. ISBN: 978-3879074792.
- NIELL, A., WHITNEY, A., PETRACHENKO, B., SCHLÜTER, W., VANDENBERG, N. & HASE, H., 2006: Current and Future Requirements for Geodetic VLBI Systems. In: Behrend, D., Bayer, K.D. (Hrsg.): *International VLBI Service for Geodesy and Astrometry 2005 Annual Report*, 13-40. NASA/TP-2006-214136.
- PETRACHENKO, B., NIELL, A., BEHREND, D., COREY, B., BÖHM, J., CHARLOT, P., COLLILOUD, A., GIPSON, J., HAAS, R., HOBIGER, T., KOYAMA, Y., MACMILLAN, D., MALKIN, Z., NILSSON, T., PANY, A., TUCCARI, G., WHITNEY, A., & WRESNIK, J., 2009: Design aspects of the VLBI2010 system. NASA/TM-2009-214180.
- ROTHACHER, M., BEUTLER, G., BEHREND, D., DONNELLAN, A., HINDERER, J., MA, C., NOLL, C., OBERST, J., PEARLMAN, M., PLAG, H-P., RICHTER, B., SCHÖNE, T., TAVERNIER, G. & WOODWORTH, P.L., 2009: The future Global Geodetic Observing System. In: Plag, H.-P., Pearlman, M. (Hrsg.): *The Global Geodetic Observing System. Meeting the Requirements of a Global Society on a Changing Planet in 2020*, Springer, Berlin, 237-272. doi:10.1007/978-3-642-02687-4\_9.
- SARTI, P., LUCA VITTUARI, L. & ABBONDANZA, C., 2009: Laser Scanner and Terrestrial Surveying Applied to Gravitational Deformation Monitoring of Large VLBI Telescopes' Primary Reflector. *Journal of Surveying Engineering* **135**(4), 136-148, doi: 10.1061/(ASCE)SU.1943-5428.00000008.
- SPRECKELS, V., SCHLIENKAMP, A., GREIWE, A. & EBERHARDT, L., 2016: Eignung von ALS, Aero- und UAS-Photogrammetrie zur Früherkennung und Erfassung von Tagesbrüchen. Dreiländertagung der DGPF, der OVG und der SGPF in Bern, Schweiz – Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V., Band **25**, 97-112.

# Automatisierte semantische Modellierung von Innenräumen aus Bildern und abgeleiteten Punktwolken basierend auf Deep Learning Methoden

LARS S. OBROCK<sup>1</sup> & EBERHARD GÜLCH<sup>1</sup>

*Zusammenfassung: In diesem Paper präsentieren wir eine Weiterentwicklung unseres bisherigen Ansatzes, photogrammetrisch erzeugte Punktwolken mit aus Bildern extrahierten semantischen Informationen anzureichern, um eine Automatisierung der BIM-Modellierung zu ermöglichen.*

*Zur semantischen Segmentierung von Bildern verwenden wir nun das Neuronale Netz DeepLabv3+, um Bauteile und Objekte von Innenräumen zu extrahieren. Während der photogrammetrischen Rekonstruktion projizieren wir die segmentierten Kategorien in die Punktwolke. Dabei auftretende Interpolationen korrigieren wir automatisiert und erreichen bei der Klassifizierung der Punktwolke eine mIoU von 51,9 %. Diese Informationen nutzen wir, um die Punktwolke auszurichten, den Maßstab zu korrigieren und weitere Informationen zu extrahieren.*

*Durch unsere Untersuchungen sehen wir die Grundlagen für eine automatisierte BIM-Modellierung basierend auf einer Kombination von Photogrammetrie und Deep Learning bestätigt.*

## 1 Einführung

Die Digitalisierung des Bausektors schreitet immer weiter voran und vollzieht mit dem Building Information Modeling (BIM) den Schritt von zweidimensionalen Plänen auf Papier zu umfassenden, dreidimensionalen digitalen Gebäudemodellen. Diese BIM-Modelle sind das zentrale Element und bilden den gesamten Zyklus eines Gebäudes ab, von der Planung über den Betrieb bis zum Abriss. In ihnen sind neben den dreidimensionalen Bauteil- und Objektgeometrien auch die wesentlichen semantischen Informationen hinterlegt. Jedoch vollzieht sich die Einführung von BIM derzeit quasi nur bei der Planung von Neubauten. Eine großflächige Erfassung von bereits bestehenden Gebäuden als BIM-Modell ist aufgrund der Komplexität von Aufnahme und Auswertung bisher ein untergeordnetes Thema.

Der große Nachteil liegt hierbei besonders darin, dass die geometrischen und die für ein BIM-Modell essentiell wichtigen semantischen Informationen derzeit getrennt ermittelt und manuell zu einem Modell zusammengeführt werden müssen. An dieser Stelle setzen wir an, um eine (Teil-)Automatisierung der Modellierung dieser in ihrem „As-Build“ bzw. „As-Is“ Zustand zu entwickeln und dadurch ihre Erstellung zu vereinfachen. Besonders die Verbindung von Photogrammetrie und Bildanalyseverfahren bieten aufgrund der großen Informationstiefe der zugrundeliegenden Bilder ein sehr hohes Potential, diese Extraktion von geometrischen und semantischen Informationen automatisiert zu ermöglichen.

---

<sup>1</sup> Hochschule für Technik Stuttgart, Labor für Interpretation und Messung bildgebender Sensoren (LI-MES), Schellingstraße 24, D-70174 Stuttgart  
E-Mail: [Lars.Obrock, Eberhard.Guelch]@hft-stuttgart.de

In OBROCK & GÜLCH (2018) wurden hierzu erste Ansätze veröffentlicht. Diese basierten auf der Anwendung von Deep Learning zur Segmentierung von acht Innenraumbauteilen und Objekten unter Verwendung eines Fully Convolutional Networks (SHELHAMER et al. 2016). Die auf diese Weise segmentierten Bilder wurden als Ersatz für den blauen Kanal in das Originalbild eingefügt. Basierend auf diesen Falschfarbenbildern wurde durch photogrammetrische Verfahren eine Punktwolke erstellt. Jedoch trat in der Punktwolke durch den Verlust an Tiefeninformationen im blauen Kanal eine erhöhte Zahl von Fehlern auf. Trotzdem wurde dabei die Umsetzbarkeit dieses Ansatzes bestätigt, da die zuvor segmentierten Objektkategorien in den Farbwerten der Punktwolke enthalten sind und somit die für BIM so wichtigen semantischen Informationen. In dieser Untersuchung fokussieren wir uns auf die Verbesserung und Erweiterung des zuvor entwickelten Ansatzes.

Aufbauend auf verbesserten Deep Learning Architekturen zur semantischen Segmentierung, einer Anpassung des Transfers der extrahierten Informationen in eine automatisiert erstellte Punktwolke und eine anschließende Analyse und Bearbeitung dieser sind uns weitere Schritte hin zu einer Automatisierung der BIM-Modell Erstellung gelungen.

## 2 Verwandte Arbeiten

Der in der Baubranche stattfindende Wandel hin zu einer Digitalisierung der Arbeitsschritte vollzieht sich rasant durch die Nutzung von Building Information Modeling und dem hierbei zentralen BIM-Modell.

In Deutschland wurden hierfür mehrere Richtlinien, Leitfäden und Untersuchungen veröffentlicht, welche mehrheitlich die Einführung und Umsetzung von BIM in unterschiedlichen Fachbereichen behandeln, z.B. (EGGER et al. 2013; ESCHENBRUCH et al. 2014; KADEN et al. 2017; BORRMANN et al. 2015, 2015; BRAMANN et al. 2015b). Mit dem „Stufenplan Digitales Planen und Bauen“ führte das Bundesministerium für Verkehr und Digitale Infrastruktur die BIM-Methode stufenweise in die Planungsprozesse der öffentlichen Infrastruktur ein (BRAMANN et al. 2015a). Die Auswirkungen von BIM auf die Durchführung von Infrastrukturprojekten wird in der Einführungsphase von diesem ausführlich untersucht.

Das Modellieren von Bestandsgebäuden basiert derzeit auf ihrer Aufnahme durch geodätische Messgeräte. Mögliche Vorgehensweisen werden unter anderem in (BORRMANN et al. 2015; CLEMEN & EHRICH 2014; KADEN et al. 2017) beschrieben. Der gesteigerte Bedarf an geometrischen und semantischen Informationen für BIM-Modelle bleibt jedoch unbeachtet. Im Umkehrschluss führt dies dazu, dass die Aufnahme und Auswertung der Messdaten sehr komplex ist und entsprechend längere Zeit in Anspruch nehmen kann.

Im gesamten Feld des Deep Learning und auch speziell bei den für die Computer Vision und Bildanalyse sehr gut geeigneten Convolutional Neural Networks (CNN) wurden seit (KRIZHEVSKY et al. 2012) gewaltige Fortschritte gemacht. So steigerten sich die Genauigkeiten bei der Bildklassifizierung durch die Verwendung von weiterentwickelten Netzwerkarchitekturen stark (z.B. SZEGEDY et al. 2014; SIMONYAN & ZISSERMAN 2015; HE et al., 2015; XIE et al. 2017; HAUNG et al. 2018). Entwickelt wurde neben der reinen Klassifikation auch eine Lokalisierung von Objekten in den Bildern. Zu diesem Zweck werden in der Objekt Detektion genannten Bildanalyse Bounding Boxen zur Unterteilung in kleinere Bereiche gesucht, deren Inhalt klassifiziert und somit

eine Lokalisation geschaffen (GERSHICK et al. 2014; GERSHICK 2015; REN et al. 2016). Die semantische Segmentierung setzt im Gegensatz dazu auf eine pixelgenaue Klassifizierung des gesamten Bildes. Diese eignet sich für die Projektion der semantischen Informationen in die Punktwolke. Eines der ersten für die semantische Segmentierung entwickelten Neuronalen Netze stammt von (SHELHAMER et al. 2016) und heißt Fully Convolutional Network (FCN). Dieses bildet die Grundlage für viele weitere Architekturen, mit denen die erreichten Genauigkeiten mehr und mehr gesteigert wurden (JÉGOU et al. 2017; ZHAO et al. 2017; CHEN et al. 2016). In dieser Untersuchung verwenden wir in DeepLabv3+ (CHEN et al. 2018) eine Architektur, welche in Benchmarks sehr gute Genauigkeiten erreicht.

In (BOULCH et al. 2017) werden von vermaschten Punktwolken RGB- und Tiefenbilder erstellt, diese mit einem CNN segmentiert und anschließend über eine Projektion auf das Mesh wieder in die Punktwolke übertragen.

### 3 Methoden

#### 3.1 Bildsegmentierung

Eine qualitativ hochwertige und semantisch angereicherte Erfassung in drei Dimensionen ist von essentieller Bedeutung für die automatisierte, vollständige Rekonstruktion aller wichtigen Bauteile und Objekte eines Bestandsgebäudes als BIM-Modell.

Aufbauend auf den in (OBROCK & GÜLCH 2018) beschriebenen Methoden bilden Innenraumbilder die Basis der Untersuchung. Bei den damals verwendeten Kategorien handelte es sich nur um einen Grundstock an wichtigen Objekten. Um ein möglichst vollständiges BIM-Modell zu erzeugen, erweiterten wir diese auf insgesamt 24 für Innenräume wichtige Bauteile und Objekte.

Die Auswahl umfasste die am häufigsten in Innenräumen vorkommenden Objekte und richtete sich nach den Erfahrungen der zuvor erfolgten Untersuchung. Die Kategorien lassen sich grob in „Raum formende Bauteile“, „Verbindende Bauteile“, „Feste Objekte von Interesse“ und „Bewegliche Objekte von Interesse“ aufteilen. Eine Übersicht über die getroffene Auswahl findet sich in Tabelle 1.

Tab. 1: Übersicht über die für Innenräume ausgewählten wichtigen Bauteile und Objekte.

Raum formende Bauteile	Verbindende Bauteile	Feste Objekte von Interesse		Bewegliche Objekte von Interesse
Boden	Tür	Lichtschalter	Steckdose	Poster
Wand	Fenster	Lampe	Stütze/Säule	Feuerlöscher
Decke		Heizung	Rohre	Teppiche
		Treppenstufen	Geländer	Schrank
		Waschbecken	Toilette	Regal
		Kabelkanal	Feuermelder	Tisch
		Rauchmeldersirene		Stuhl

Allerdings konnte mit der „Wand“ eines der Hauptmerkmale eines Innenraums nicht berücksichtigt werden, da eine Unterscheidung zwischen ihr und der „Decke“ kaum möglich war und sie

einen großen Anteil in den Bildern einnimmt. Dies führte zu einer zu starken Gewichtung dieser Klasse und resultierte deshalb in einem negativen Einfluss auf die Qualität der Segmentierung der anderen Kategorien.

Basierend auf diesen Kategorien wurde für das spätere Training des Neuronalen Netzes ein Trainingsdatensatz erstellt. Dieser besteht aus ca. 300 Bildern, von welchen ein großer Teil selber aufgenommen und der Rest aus dem Internet bezogen wurde. Zu diesen wurde manuell das „ground truth“-Bild segmentiert und so jedem Pixel eine eindeutige Kategorie zugewiesen. Da dies für das Training eines Neuronalen Netzes wenige Bilder sind, wurde der Datensatz mit Hilfe von Daten Augmentierung (data augmentation) auf fast 18.000 Bilder erweitert. Hierfür wurden die Bilder durch Zerschneiden, Drehen, Spiegeln, Helligkeitsänderung und zufälligem Rauschen verändert. Der so erzeugte Datensatz ist für ein Neuronales Netz noch immer recht klein, ermöglicht aber durch fine-tuning eines vortrainierten Modells (pre-trained model) das Training.

Die rapiden Fortschritte im Bereich des Deep Learning nehmen wir zum Anlass, eine andere Struktur als Neuronales Netz heranzuziehen. Anstelle eines Fully Convolutional Networks wurde das im Jahr 2018 erschienene und stark weiterentwickelte DeepLabv3+ (CHEN et al. 2018) als Grundlage für das Training verwendet. Dieses baut ebenfalls auf einer Encoder-Decoder Struktur auf. Zur Feature Extraktion in der Encoder Phase setzt es auf die Xception Architektur und erweitert diese mit Atrous Spatial Pyramid Pooling zur besseren Einbeziehung der erweiterten Nachbarschaft des untersuchten Neurons. Durch Verwendung des Decoders wird die zuvor reduzierte Ausdehnung der Layer, basierend auf aus dem Encoder abgeleiteten Werten wieder hochskaliert. So ermöglicht dieses Neuronale Netz eine gute Qualität der Segmentierungen und präzise Kanten zwischen unterschiedlichen Bereichen.

Zur semantischen Segmentierung der Bilder wurde das DeepLabv3+ basierend auf der Xception 65 Architektur herangezogen. Zu dieser Architektur wurde das zugehörige auf ImageNet und COCO Daten vortrainierte Modell genommen. Aufbauend auf diesem erfolgte das Training durch fine-tuning.

### **3.2 Klassifizierte Punktwolke**

Zur automatisierten Erstellung eines BIM-Modells werden dreidimensionale, semantisch angereicherte Informationen benötigt. Durch die Verwendung von Photogrammetrie und digitaler Bildzuordnung können die Bilder als Grundlage zur Erstellung einer dreidimensionalen Punktwolke verwendet werden.

Als zentrales Element eines Gebäudes steht die Rekonstruktion von Innenräumen im Fokus der Untersuchungen. Als unabhängiges Vergleichsobjekt wurde zur generellen Darstellung eines solchen ein anderes, typisches Büro in einem öffentlichen Gebäude ausgewählt. Dabei wurde darauf geachtet, dass keines der Bilder im Trainingsdatensatz aus dem Inneren dieses Büros stammt, um die Möglichkeit eines Overfittings zu diesem und damit dessen Auswirkungen auszuschließen. Von diesem Büro wurde unter photogrammetrischen Gesichtspunkten eine Vielzahl an Bildern aufgenommen. Diese Bilder werden durch das trainierte Modell segmentiert, um pixelgenau die in ihnen enthaltenen Objekte verorten zu können. In einem eigens entwickelten Prozessschritt werden kleinste Flächen, die keine realen Objektteile darstellen können, aus den erzeugten Bildern herausgefiltert.

In OBROCK & GÜLCH (2018) wurden vor der Erzeugung der Punktwolke die segmentierten Bilder anstelle des Blauen Kanals in die Originalbilder eingefügt. Die dabei entstehenden Falschfarbenbilder waren noch immer geeignet, eine Punktwolke zu erzeugen, in welcher die segmentierten semantischen Informationen enthalten waren. Jedoch sind durch den Informationsverlust der Tiefeninformationen dieses Kanals einige Fehler in der Punktwolke aufgetreten. Außerdem waren sie durch die Komprimierung der Klasseninformationen im Blauen Kanal für den Menschen nicht mehr einfach visuell zu unterscheiden, da sie alle in leicht unterschiedlichen Lila-Tönen dargestellt wurden. Um diesen Einschränkungen zu begegnen, haben wir für diese Untersuchung unsere Vorgehensweise angepasst. So werden den einzelnen Kategorien keine IDs mehr zugewiesen, um diese in den Blauen Kanal einzufügen. Die durch DeepLabv3+ segmentierten Bilder werden als RGB-Bilder gespeichert und die Kombination aller drei Farbkanäle steht eindeutig für eine Kategorie.

Als Grundlage für die nächsten Schritte erstellen wir unter Verwendung der Originalbilder eines Innenraums in Agisoft Metashape (AGISOFT 2020) automatisiert und ohne Platzierung von Passpunkten eine Punktwolke. Um die Kategorien in diese zu überführen, ist es möglich, die Originalbilder durch die segmentierten Bilder zu ersetzen und die Punktwolke auf Grundlage der zuvor ermittelten Kameraposition und Rotation einzufärben zu lassen. Eine Übertragung findet nicht mehr direkt bei der Punktwolkenerstellung aus einem angereicherten Farbkanal statt, sondern aus der Projektion der segmentierten Bilder auf die Punktwolke. Somit findet der Informationstransfer der semantischen und geometrischen Informationen nicht mehr in einem, sondern in zwei getrennt aber direkt aufeinander aufbauenden Schritten statt. Die hierdurch ermöglichte Reduktion des Informationsverlustes bringt viele Vorteile.

Bei der Generierung der Punktwolke durch Photogrammetrie stehen alle Farbkanäle verlustfrei zur Verfügung. Dies ermöglicht eine bessere Qualität beim Extrahieren und Verknüpfen einzelner prägnanter Bildbereiche und erlaubt auf diese Weise eine bessere Rekonstruktion der gesamten Punktwolke. Außerdem lassen sich die Farbwertkombinationen der einzelnen Kategorien in größerem Abstand zueinander platzieren. Diese sind sehr gering, wenn 24 Klassen auf 256 mögliche Farbwerte eines Kanals aufgeteilt werden sollen. Im Gegensatz dazu ermöglicht die Kombination von Werten aus drei Kanälen deutlich größere euklidische Abstände zwischen den Farbwerten der Kategorien, da sie als Punkte im dreidimensionalen Raum betrachtet werden können. Diese Verteilung auf alle Farbkanäle birgt auch für den Betrachter den größten Vorteil, da alle Farben deutlicher voneinander zu unterscheiden und die segmentierten Klassen einfach festzustellen sind.

Beim Transfer der semantischen Informationen werden die Farbwerte der Pixel durch die Bilder bestimmt, aus denen sie generiert wurden. Ist die in diesen vorgenommene Segmentierung an den sich überschneidenden Stellen nicht übereinstimmend oder sind die Kameraausrichtungen im dreidimensionalen Raum nicht perfekt zueinander, treten durch Interpolation abweichende Farbwerte in den erzeugten Punkten auf.

Deshalb passten wir die RGB-Farbwerte für die Kategorien bereits in den segmentierten Bildern so an, dass deren minimaler euklidischer Abstand untereinander und zu möglichen interpolierten Farben maximal groß ist. Auf diese Weise wird eine einigermaßen gleichmäßige Verteilung der Farbwerte im dreidimensionalen Raum erreicht und die eindeutige Zuordnung der Kategorien auch

bei Interpolationen verbessert. In der Punktwolke auftretende Ausreißerpunkte werden herausgefiltert.

In einem weiteren Arbeitsschritt müssen die aufgetretenen Interpolationen eliminiert werden. Durch die Betrachtung der Farbwerte der einzelnen Punkte im dreidimensionalen Raum und der aus diesen abgeleiteten euklidischen Distanz zu den Farbwerten der Kategorien wird eine eindeutige Zuweisung ermöglicht. Auch werden die Distanzen der Farbwerte eines Punktes zu den Farbwerten errechnet, welche durch die Interpolation einer Kategorie mit Background-Pixeln entstehen würden. Liegen diese aus den Farbwerten ermittelten Distanzen innerhalb eines festgelegten Maximalwerts, erfolgt eine direkte Zuordnung zu der entsprechenden Kategorie.

Im Fall, dass sich keine direkte Zuordnung ergibt, wird für die Umgebung der Punkte untersucht. Erreicht eine Kategorie bei diesen eine deutliche Mehrheit, wird sie für den untersuchten Punkt übernommen. Falls dies nicht der Fall ist, wird eine Kombination der beiden Herangehensweisen gewählt. Dafür wird untersucht, ob die Distanz der Farbwerte des untersuchten Punktes zu der am häufigsten in der Umgebung vorkommenden Kategorie in eines erweiterten Maximalwerts zu finden sind. Sind durch diese Distanz und Nachbarschaftsuntersuchungen keine Kategorien ermittelbar, werden die untersuchten Punkte als unbestimmte Background-Punkte geführt.

In der auf diese Weise erstellten Punktwolke ist jedem Punkt eine eindeutige Kategorie zugeordnet. Somit wurde die Verknüpfung von dreidimensionalen und semantischen Informationen hergestellt.

### **3.3 Automatisierte Bearbeitung und Analyse einer semantisch angereicherten Punktwolke**

Aufbauend auf dieser erzeugten Punktwolke erfolgen nun weitere automatisierte Analyse- und Bearbeitungsschritte. Da die Punktwolke voll automatisiert und ohne die Verwendung von Passpunkten und Marken generiert wurde, entsprechen weder ihre Position und Rotation im Raum noch ihr Maßstab den realen Gegebenheiten. Da dies jedoch für eine spätere Überführung in ein BIM-Modell benötigt wird, müssen diese Parameter nun ermittelt und mit ihnen die Punktwolke angepasst werden. Die dafür notwendigen Informationen lassen sich aus der Punktwolke unter Verwendung der in ihr enthaltenen und für diesen Schritt essentiell wichtigen semantischen Informationen extrahieren.

Als Ausgangsbasis werden die als Boden klassifizierten Punkte verwendet, da diese neben der Rotation auch eine sinnvolle Translation in die waagrechte XY-Achsebene ermöglichen. Zu diesem Zweck wird eine Best-Fit Ebene des Bodens bestimmt und die Punktwolke in einem iterativen Prozess rotiert, bis sich die Bodenebene im lokalen Koordinatensystem nahezu in der Waagerechten befindet und die Z-Achsen annähernd übereinstimmen. Nachdem die Punktwolke so ausgerichtet wurde, wird die Deckenebene durch die zu ihr gehörenden, segmentierten Punkte bestimmt.

Basierend auf der ausgerichteten Punktwolke lassen sich nun auch Wandebenen und Wandpunkte bestimmen, welche bei der Segmentierung durch Deep Learning in den Bildern nicht berücksichtigt werden konnten. Dies geschieht mit Hilfe einer aus den unbestimmten Punkten der Punktwolke erstellten, zweidimensionalen Heatmap, welche senkrecht von oben in einem feinen Raster die

Dichte der vorhandenen Punkte darstellt. Besonders an senkrechten Bauteilen ist bei dieser Betrachtung in zwei Dimensionen eine Häufung an Punkten zu erwarten, durch die vor allem Wände gut auszumachen sind.

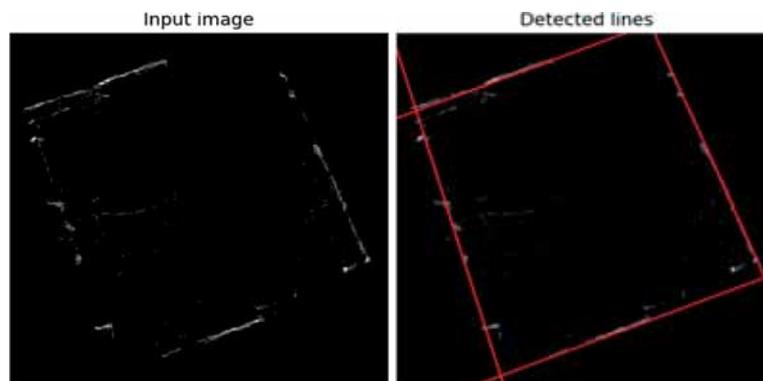


Abb. 1: Abgeleitete Heatmap des Raums (links) und die durch Hough-Linien-Transformation ermittelten Winkel der Wände (rechts)

Aus den erstellten Heatmaps werden mit einer Hough-Linien-Transformation die wahrscheinlichen Winkel der Wände ermittelt, welche im Anschluss wieder in die dritte Dimension übertragen werden. Bis dahin nicht kategorisierte Punkte, welche sich in geringem Abstand zu diesen befinden, werden ausgewählt, um aus ihnen die tatsächlichen Best-Fit Ebenen zu bestimmen. Die diesen Ebenen nahe gelegenen Punkte werden als Wandpunkte klassifiziert.

Zur Ermittlung eines Maßstabs für die Punktwolke wird das Verhältnis der Dimensionen eines Objekts in der Realität zum selben Objekt in der Punktwolke benötigt. Da sie als verbindendes Bauteil ein zentrales Element eines jeden Innenraums ist, wurde die Objektkategorie „Tür“ als Vergleichsobjekt ausgewählt. Zur Ermittlung des Maßstabs wird in der Punktwolke aus den als „Tür“ klassifizierten Punkten die größte zusammenhängende Türfläche ausgewählt. Als Vergleichsgröße wird die Höhe herangezogen. Sie ist sowohl in der Realität als auch in der Punktwolke ohne Schwierigkeiten zu ermitteln. So lässt sich die Höhe der Türfläche in der Punktwolke durch die Differenz der maximalen und minimalen Werte in der Z-Achse bestimmen, da ihre Ausrichtung zuvor korrigiert wurde. Durch den Vergleich mit der an der realen Tür mit einfachen Werkzeugen gemessenen Höhe lässt sich der Maßstab ermitteln. Basierend auf diesem wird eine Skalierung der Punktwolke durchgeführt.

Die Überführung der semantischen Informationen in die automatisiert und ohne Passpunkte generierte Punktwolke ermöglicht somit die Korrektur ihrer Rotation und Translation und erlaubt auch die Extraktion von Best-Fit Ebenen. Aufbauend auf diesen lassen sich weitere Informationen, wie Ebenen und dazugehörige Punkte extrahieren. Gemeinsam mit den vollständig aus den Daten abgeleiteten Wandebenen bildet die Boden- und Deckenebene die Raumgeometrie in ihren Grundzügen ab.

Eine Skalierung anhand eines Maßstabs durch den Vergleich der Höhe eines klassifizierten Objekts mit seinem realen Vorbild ist möglich, da die Ausrichtung der Punktwolke korrigiert wurde. Die Türe bietet sich hierbei als ein in jedem Raum vorhandenes Objekt an und hat auch auf Grund ihrer Größe gute Aussichten, in der Punktwolke in hoher Qualität vorhanden zu sein.

Die korrekt positionierte, rotierte und skalierte Punktwolke steht in Zukunft für weitergehende Untersuchungen hin zu einer Automatisierung der BIM-konformen Modellierung zur Verfügung.

## 4 Ergebnisse und Evaluation

### 4.1 Semantische Segmentierung

Aus Zeitgründen wurde darauf verzichtet, einen weiteren Datensatz für die Validierung des trainierten DeepLabv3+ Modells zu erstellen. So lassen sich für die segmentierten aktuell noch Bilder keine Aussagen über die statistische Genauigkeit treffen. Rückschlüsse auf diese lassen sich aber auch aus der in Kapitel 4.2 beschriebene Untersuchung der korrekten Farbwerte in der Punktwolke ziehen. Durch eine visuelle Bewertung der Bilder lässt sich dennoch ein Überblick über die Qualität der Segmentierungen schaffen.



Abb. 2: Vergleich zwischen Originalbildern und mit unserem DeepLabv3+ Modell segmentierten Bildern.

Die eigentlich beim Training angegebene mIoU (mean Intersection over Union) von ca. 90% wird in den vollständig unbekanntem Bildern des Büros nicht erreicht. Hierfür sind in den Bildern zu viele fehlerhafte Flächen vorhanden. Dies deutet auf ein Overfitting des Neuronalen Netzes auf die Trainingsdaten hin und lässt sich zu einem großen Teil auf die noch immer sehr kleine Datenbasis zurückführen.

Dennoch sind in den segmentierten Bildern die einzelnen Objekte zumeist korrekt erkannt worden. Hierbei wird die durch den Wechsel der Architektur des Neuronalen Netzes zu erwartende Verbesserung in der Segmentierung deutlich, da zumindest visuell deren Übereinstimmung eine höhere Qualität vermittelt als zuvor. Dies war wegen der gesteigerten Komplexität durch die Erweiterung der Kategorien nicht unbedingt zu erwarten gewesen.

Auffällig ist, dass Objekte mit relativ großer Fläche wie „Tische“ und „Stühle“, „Tür“, „Fenster“, „Boden“ und „Decke“ und besonders die „Heizung“ häufig gut segmentiert wurden. Bei ihnen ist erkennbar, dass Bereiche mit guter Segmentierung zumeist auch mit sehr präzisen Kanten ausgestattet sind. Allerdings scheinen Lücken innerhalb von segmentierten Flächen hin und wieder vorzukommen. Für diese sind besonders große, gleichförmige Bereiche wie die Decke anfällig.

Eine Ausnahme bildet die Kategorie „Regal“, bei der die Bereiche mit Ordnern häufig korrekt erkannt wurden, während Bereiche ohne diese seltener fehlerfrei segmentiert wurden. Diese Lücken werden oft durch andere Kategorien gefüllt. Kleinere Objekte wie „Lichtschalter“, „Steckdose“ und die mittelgroßen wie „Kabelkanal“, „Poster“ und „Lampe“ wurden in unterschiedlicher Qualität segmentiert. Auf ihre Segmentierungsgenauigkeit scheint der Blickwinkel einen starken Einfluss zu haben. Fehlklassifizierungen treten häufiger auch ohne ersichtlichen Grund auf. Komplet falsch klassifizierte Bereiche finden sich häufig an wiederkehrenden Stellen. So ist es für das Neuronale Netz ein Problem, Stuhl- und Tischbeine auseinanderzuhalten. Eine an der Wand hängende Tafel kennt es aus den Trainingsdaten zwar nicht, erkennt aber immer wieder Ähnlichkeiten zu Türen und segmentiert sie entsprechend. Außerdem wurden Vorhänge fälschlicherweise als „Fenster“ erkannt.

Insgesamt muss gesagt werden, dass in den Bildern große, voluminöse Bereiche besser segmentiert wurden als kleine und feine Flächen.

Aufbauend auf dieser mit DeepLabv3+ durchgeführten Segmentierung, welche viele der wichtigsten Kategorien von Innenräumen insgesamt qualitativ hochwertig extrahiert, ist es uns möglich, die semantischen Informationen in die erstellte Punktwolke zu projizieren.

## 4.2 Klassifizierte Punktwolke

Die automatisierte photogrammetrische Rekonstruktion einer Punktwolke ohne Passpunkte funktioniert aufbauend auf den Bildern des Büroraums ohne Probleme. Die in die Punktwolke übertragenen Farbwerte der Kategorien sind noch immer deutlich zu erkennen. So treten zwar Interpolationen auf, welche die eigentlichen Farben besonders an den Übergängen zwischen Objekten verändern und abdunkeln, aber insgesamt halten sie sich im Rahmen.

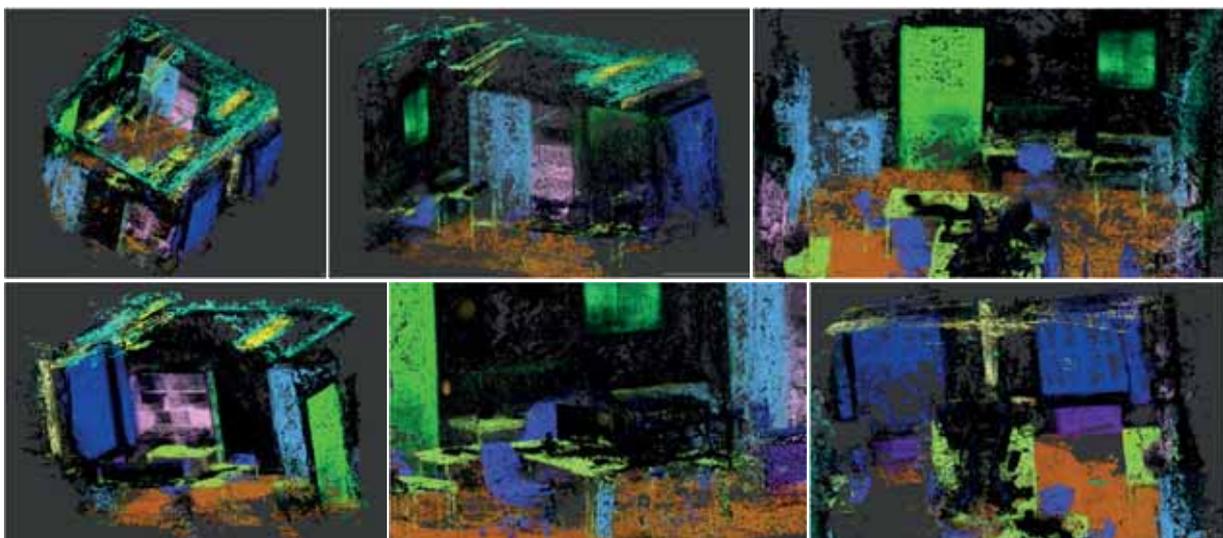


Abb. 3: Bilder der durch Photogrammetrie erzeugten Punktwolke. Aufgrund der durch Interpolation auftretenden Farbveränderungen sind die Kategorien zwar visuell unterscheidbar, aber nicht eindeutig klassifiziert.

Die erzeugte Punktwolke umfasst 16 Millionen Punkte, ist in großen Teilen vollständig und stellt ein einigermaßen gutes Abbild des Raums dar. So sind zwar verschiedene Lücken, durch eine

falsche Zuordnung entstandene falsche Punktcluster und auch einige Ausreißerpunkte vorhanden, doch ein großer Teil der Punktwolke bildet die Realität relativ genau ab.

Bei näherer Betrachtung wird deutlich, dass nicht für alle Bereiche ausreichend Überschneidungen gefunden wurden und Teile daher nur fehlerhaft rekonstruiert werden konnten. Besonders der Boden und die Decke sind hier auffällig. Beide sind großflächig und sehr gleichförmig. Die hierbei auftretenden Schwierigkeiten, Punkte in den Bildern zu extrahieren und mit den homologen Punkten anderer Bilder zu verknüpfen, führten dazu, dass Punkte der Decke fast nur in direkter Nachbarschaft zu anderen Objekten wie den Lampen oder Wänden liegen. Der Boden ist zwar vollständiger, jedoch nicht fehlerfrei rekonstruiert worden. Bei ihm handelt es sich nicht um eine zusammenhängende Fläche. Vielmehr besteht er aus verschiedenen, zusammengestückelten Bereichen, die teilweise unterschiedlich geneigt sind. Die ebenfalls großflächigen Wände sind deutlich besser erfasst worden, dennoch sind auch hier Lücken und nicht rekonstruierte Bereiche in den Ecken des Raums vorhanden. Viele der anderen in diesem Raum vorhandenen Objekte wurden gut, aber nicht immer vollständig und mit der höchsten Genauigkeit erfasst.

In dieser fehlerhaft klassifizierten Punktwolke wurden die in den einzelnen Punkten vorhandenen Farbwerte genutzt, um die durch Interpolation auftretenden Veränderungen zu eliminieren.

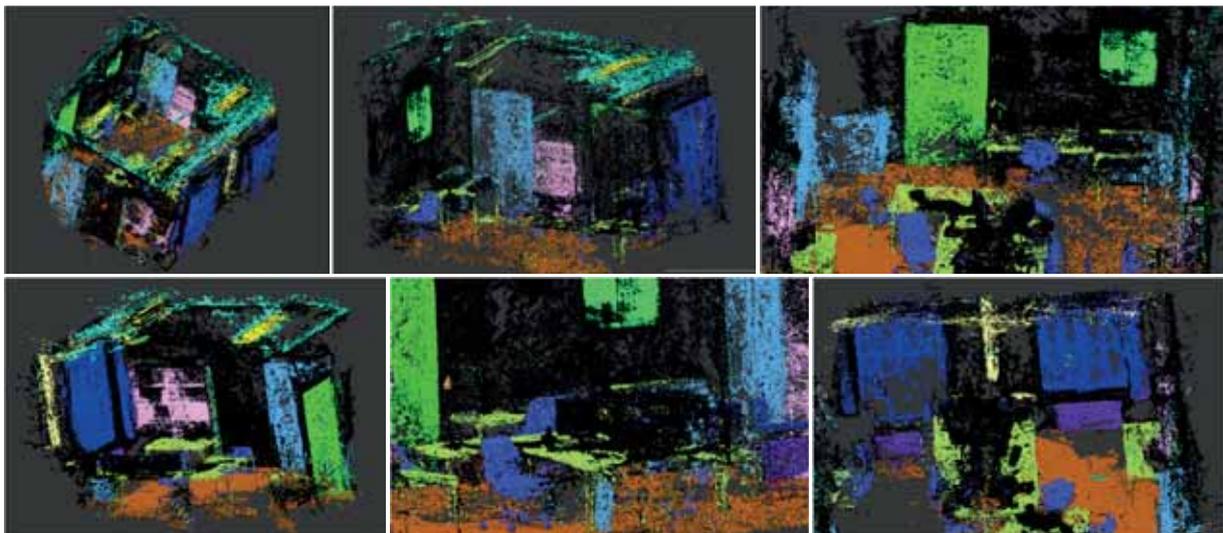


Abb. 4: Bilder der Punktwolke mit eindeutiger Klassifikation jedes einzelnen Punktes

Durch die extrahierten eindeutigen Kategorien sind diese viel besser zu unterscheiden. Es wird ersichtlich, wie gut die Übertragung der semantischen Informationen in vielen Fällen funktioniert. Durch die Überschneidung vieler segmentierter Bilder haben einzelne Fehler in diesen nur selten Einfluss auf die Klassifizierung der Punktwolke. Fehlerhafte Zuweisungen von Kategorien treten dann auf, wenn in den segmentierten Bildern regelmäßige Fehler vorhanden sind oder bei der Erstellung der Punktwolke fehlerhafte Kamerapositionen auftreten.

Basierend auf dieser eindeutig klassifizierten Punktwolke wurde eine Untersuchung zur Übereinstimmung der enthaltenen Kategorien mit den erwarteten Kategorien durchgeführt. Zu diesem Zweck wurde eine Vergleichspunktwolke mit den Soll-Kategorien erzeugt. So können für die in

dem untersuchten Büro vorhandenen Objektkategorien Aussagen über deren Genauigkeit und daraus abgeleitete Rückschlüsse auf die Segmentierung getroffen werden.

In den Genauigkeitsberechnungen wurden während der Interpolationskorrektur entfernte Ausreißerpunkte ebenso wenig berücksichtigt wie fehlerhafte Punkte aus der Vergleichspunktwolke.

Tab. 2: Genauigkeitsuntersuchung der in der Punkt wolke enthaltenen Kategorien.

	Mittel	Background	Boden	Tür	Fenster	Lichtschalter	Steckdose	Lampe	Heizung	Poster	Kabelkanal	Schrank	Decke	Regal	Tisch	Stuhl
IoU	51,9	60,8	75,1	77,9	75,6	6,3	10,8	59,2	66,4	43,6	41,6	71,4	53,5	38,4	43,3	54,9
Accuracy	60,4	93,8	77,1	84,4	88,8	48,2	11,3	62,2	70,5	45,3	43,0	74,2	54,7	38,8	50,0	63,4

Zur Einschätzung der Qualität wird erneut die mIoU herangezogen. Diese beträgt für die Punkt wolke gute 51,9 %. Außerdem wird zusätzlich die Genauigkeit der Kategorien, oft auch als Accuracy bezeichnet, berechnet. Ihr Durchschnitt für alle Kategorien beträgt sehr gute 60,4 % und deutet auf eine hohe Übereinstimmung hin. Betrachtet man die einzelnen Kategorien, zeigt sich ein weit gestecktes Feld an Werten. Die beste IoU erreicht die Kategorie „Tür“ mit 77,9 % dicht gefolgt von „Fenster“ mit 75,6 % und „Boden“ mit 75,1 %. Auch die knapp dahinterliegenden Kategorien sind noch immer sehr gut. Entsprechend sind viele Objekte in der Punkt wolke zu großen Teilen korrekt segmentiert.

Die Klassen, in denen Fehler in der Objektklassifizierung am deutlichsten sichtbar werden, sind die „Poster“, welche teilweise Lücken in der Segmentierung aufweisen, und besonders die „Regale“. Diese sind, wie ja bereits in den Bildern festzustellen war, nur unzureichend segmentiert. Die stattfindende Interpolation hat bei ihnen geholfen, verschieden auftretende falsche Segmentierungen auszugleichen. Dennoch sind die Regale nur unvollständig als solche klassifiziert. Einen positiven Einfluss hatte die Interpolation auch bei der vorhandenen Tafel. So wird diese nur in einem sehr kleinen Bereich als „Tür“ segmentiert.

Außerdem ist auffällig, dass besonders die Kategorien „Lichtschalter“ mit 6,3 % und „Steckdose“ mit 10,8 % nur eine sehr geringe IoU haben. Bei diesen beiden handelt es sich um die Kategorien mit der geringsten Anzahl an erwarteten Punkten (Lichtschalter: 5038, Steckdose: 6104). Entsprechend schnell haben falsch segmentierte Bereiche in den Bildern einen großen Einfluss auf die ermittelten Werte. So sind diese bei der „Steckdose“ oft als Background-Punkte klassifiziert worden. Betrachtet man bei den Lichtschaltern zusätzlich die errechnete Accuracy, wird deutlich, dass die eigentliche Klassifizierung der Vergleichsbereiche der Lichtschalter mit 48,2 % einen deutlich besseren Wert aufweist. Dies hat den Hintergrund, dass kleine Bereiche anderer Klassen, in diesem Fall „Boden“ mit 25741 Punkten, fälschlich als Lichtschalter klassifiziert wurden und einen sehr starken Einfluss auf die IoU dieser Kategorie haben. Die Kategorie „Boden“ weist eine ähnliche, leicht dunklere Farbe auf wie der „Lichtschalter“. So scheint für die schlechte IoU von 6,3% vor allem die bei der Überführung der semantischen Informationen auftretende Interpolation der Farben verantwortlich zu sein. Dies verdeutlicht ungefähr die Dimensionen, in denen die Interpolation nach ihrer Korrektur eine Rolle für die Klassifizierung der Punkt wolke spielt. Für die Berechnung

der Qualität des Lichtschalters ist das viel, im Zusammenhang mit der kompletten Punktwolke scheinen diese jedoch keine wesentliche Rolle zu spielen.

Insgesamt ist die Qualität der Punktwolke in Ordnung. Besonders die in ihr segmentierten Kategorien sind ein sehr gutes Abbild der realen Objekte. Die bei der Projektion der semantischen Informationen auftretenden Interpolationen haben wenige negative, im Gegenteil zumeist sogar positive Auswirkung auf die Klassifizierung der Punkte, da einzelne fehlerhafte Segmentierungen aus den Bildern nicht übernommen werden.

### 4.3 Post-Processing der Punktwolke

Die eindeutig klassifizierte Punktwolke wurde durch Rotation und Translation korrekt ausgerichtet.

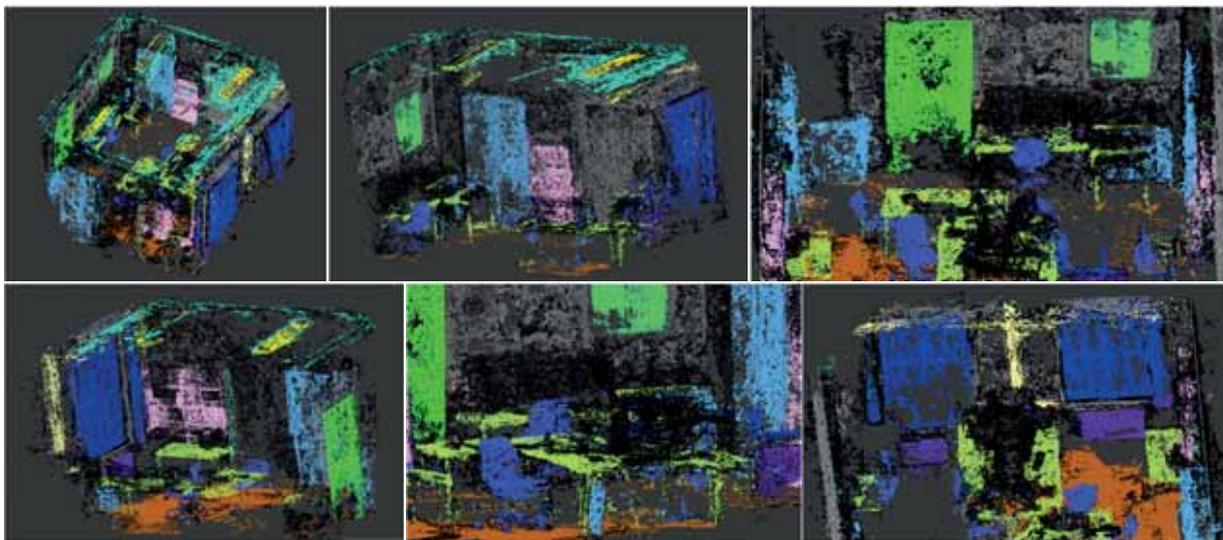


Abb. 5: Durch Rotation und Translation angepasste Punktwolke und zusätzlich in dieser segmentierte Wandpunkte.

Boden und Deckenpunkte wurden auf ihre jeweiligen Ebenen projiziert. Außerdem wurden weitere starke Ausreißerpunkte ebenso wie Punkte unterhalb der Bodenebene und oberhalb der Deckenebene entfernt. Dies hat zwar zur Folge, dass in schlecht rekonstruierten Bereichen mehr Lücken auftreten, die Punktwolke jedoch deutlich klarer aussieht und besser mit der Realität übereinstimmt. Die basierend auf den ermittelten Wandebenen bestimmten Wandpunkte fügen sich gut in das Gesamtbild dieser verbesserten Punktwolke ein und bilden die Wände korrekt ab.

Nach der Skalierung der Punktwolke mit dem ermittelten Maßstab wurden verschiedene Objekte zu Vergleichszwecken gemessen. Bei diesen wurde eine gute Übereinstimmung zwischen Soll- und Istwerten festgestellt. Die Höhe eines Schreibtisches wurde in der Realität auf 72,3 cm gemessen, während in der Punktwolke durchschnittlich 72,1 cm gemessen wurden. Die Breite eines Schrankes betrug in der Realität 110 cm, in der Punktwolke durchschnittlich 108,9 cm. Insgesamt zeigt sich, dass die Skalierung der Punktwolke gut funktioniert hat und deren Ausmaße sehr gut mit der Realität übereinstimmen.

Die automatisierte Nachbearbeitung der Punktwolke ermöglichte es uns somit, fehlerhafte Bereiche zu korrigieren, weitere Informationen aus dieser zu extrahieren und sie basierend auf diesen zu optimieren.

Dies verdeutlicht das Potential, welche eine Kombination von semantischen und dreidimensionalen Informationen für die automatisierte Auswertung bietet.

## 5 Fazit

Die kombinierte Extraktion von geometrischen und semantischen Informationen wurde aufbauend auf der Segmentierung mit DeepLabv3+ und der Projektion in die photogrammetrisch erstellte Punktwolke deutlich verbessert. So sind die gesuchten Objekte sehr gut in der Punktwolke differenzierbar. Diese hohe Genauigkeit wird auch durch die bei der Klassifikation der Punktwolke erreichte mIoU von 51,9 % unterstrichen.

Außerdem wurde gezeigt, dass durch die Analyse und Nachbearbeitung der Punktwolke wichtige Informationen extrahiert werden können, die für eine Automatisierung der Auswertung hin zu einem BIM-Modell von essentieller Bedeutung sind.

Wir sind zuversichtlich, die noch auftretenden Probleme durch eine Optimierung der verwendeten Verfahren minimieren zu können.

In Zukunft ist es angedacht, eine Ausweitung dieses Ansatzes auf die Verknüpfung und gemeinsame Auswertung mehrerer Innenräume herzustellen und dabei auch mobile Laserscanner mit einzubeziehen.

Mit den von uns vorgestellten Methoden haben wir eine solide Grundlage für die Erfassung und Modellierung von semantischen und geometrischen Informationen von Innenräumen für BIM-Modelle hin zu ihrer automatisierten Rekonstruktion geschaffen.

## 6 Danksagung

Das Projekt „i\_city: BIM-konforme Gebäudeerfassung“ wird vom Bundesministerium für Bildung und Forschung (BMBF) unter dem Förderkennzeichen 13FH9E01IA gefördert und vom Projektträger VDI Technologiezentrum GmbH für das BMBF betreut.

## 7 Literaturverzeichnis

- AGISOFT, 2020: Agisoft Metashape. <http://www.agisoft.com/>.
- BORRMANN, A., KÖNIG, M., KOCH, C. & BEETZ, J., 2015: Building Information Modeling - Technologische Grundlagen und industrielle Praxis. VDI-Buch, Springer Vieweg, Wiesbaden, 343-361.
- BOULCHA, A., GUERRY, J., LE SAUX, B. & AUDEBERT, N., 2017: SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks. *Computers & Graphics* **71**, 189-198 <https://doi.org/10.1016/j.cag.2017.11.010>.
- BRAMANN, H., MAY, I. & PLANEN-BAUEN 4.0 – GESELLSCHAFT ZUR DIGITALISIERUNG DES PLANENS, BAUENS UND BETREIBENS MBH, 2015a: Stufenplan Digitales Planen und Bauen.

- [https://www.bmvi.de/SharedDocs/DE/Publikationen/DG/stufenplan-digitales-bauen.pdf?\\_\\_blob=publicationFile](https://www.bmvi.de/SharedDocs/DE/Publikationen/DG/stufenplan-digitales-bauen.pdf?__blob=publicationFile).
- BRAMANN, H., MAY, I. & PLANEN-BAUEN 4.0 – GESELLSCHAFT ZUR DIGITALISIERUNG DES PLANENS, BAUENS UND BETREIBENS MBH, 2015b: Konzept zur schrittweisen Einführung moderner, IT-gestützter Prozesse und Technologien bei Planung, Bau und Betrieb von Bauwerken – Stufenplan zur Einführung von BIM. [https://www.bmvi.de/SharedDocs/DE/Anlage/Digitales/bim-stufenplan-endbericht.pdf?\\_\\_blob=publicationFile](https://www.bmvi.de/SharedDocs/DE/Anlage/Digitales/bim-stufenplan-endbericht.pdf?__blob=publicationFile).
- CHEN, L.-C., PAPANDREOU, G., KOKKINOS, I., MURPHY, K. & YUILLE, A. L., 2016: Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. arXiv:1412.7062.
- CHEN, L.-C., ZHU, Y., PAPANDREOU, G., SCHROFF, F. & ADAM, H., 2018: Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. arXiv:1802.02611v1.
- CLEMEN, C. & EHRICH, R., 2014: Geodesy goes BIM. In: allgemeine vermessungs-nachrichten. (avn) **121**(6), 231-237, <https://gispoint.de/artikelarchiv/avn/2014/avn-ausgabe-62014/2552-geodesy-goes-bim.html>.
- EGGER, M., HAUSKNECHT, K., LIEBICH, T. & PRZYBYLO, J., 2013: BIM-Leitfaden für Deutschland. [http://www.bmvi.de/SharedDocs/DE/Anlage/Digitales/bim-leitfaden-deu.pdf?\\_\\_blob=publicationFile](http://www.bmvi.de/SharedDocs/DE/Anlage/Digitales/bim-leitfaden-deu.pdf?__blob=publicationFile).
- ESCHENBRUCH, K., MALKWITZ, A., GRÜNER, J., POLOCZEK, A. & KARL, C. K., 2014: Maßnahmenkatalog zur Nutzung von BIM in der öffentlichen Bauverwaltung unter Berücksichtigung der rechtlichen und ordnungspolitischen Rahmenbedingungen – Gutachten zur BIM-Umsetzung. [https://www.bmvi.de/SharedDocs/DE/Anlage/Digitales/bim-massnahmenkatalog.pdf?\\_\\_blob=publicationFile](https://www.bmvi.de/SharedDocs/DE/Anlage/Digitales/bim-massnahmenkatalog.pdf?__blob=publicationFile).
- GIRSHICK, R., DONAHUE, J., DARRELL, T. & MALIK, J., 2014: Rich feature hierarchies for accurate object detection and semantic segmentation. arXiv:1311.2524v5.
- GIRSHICK, R., 2015: Fast R-CNN. <https://arxiv.org/pdf/1504.08083.pdf>.
- HE, K., ZHANG, X., REN, S. & SU, J., 2015: Deep Residual Learning for Image Recognition. arXiv:1512.03385.
- HUANG, G., LIU, Z., VAN DER MAATEN, L. & WEINBERGER, K. Q., 2018: Densely Connected Convolutional Networks. arXiv:1608.06993.
- JÉGOU, S., DROZDAL, M., VAZQUEZ, D., ROMERO, A. & BENGIO, Y., 2017: The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. arXiv:1611.09326.
- KADEN, R., CLEMEN, C., SEUB, R., BLANKENBACH, J., BECKER, R., EICHHORN, A., DONAUBAUER, A., KOLBE, T. H., GUBER, U., DVW – GESELLSCHAFT FÜR GEODÄSIE, GEOINFORMATION UND LANDMANAGEMENT E. V. & RUNDER TISCH GIS E.V., 2017: Leitfaden Geodäsie und BIM. [https://rundertischgis.de/images/2\\_publicationen/leitfaeden/GeoundBIM/Leitfaden%20Geod%C3%A4sie%20und%20BIM\\_Onlineversion.pdf](https://rundertischgis.de/images/2_publicationen/leitfaeden/GeoundBIM/Leitfaden%20Geod%C3%A4sie%20und%20BIM_Onlineversion.pdf).
- KRIZHEVSKY, A., SUTSKEVER, I. & HINTON, G. E., 2012: ImageNet Classification with Deep Convolutional Neural Networks. <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>.
- OBROCK, L. S. & GÜLCH, E., 2018: First Steps To Automated Interior Reconstruction From Semantically Enriched Point Clouds And Imagery. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. **42**(2), 781-787, <https://doi.org/10.5194/isprs-archives-XLII-2-781-2018>, 2018.

- REN, S., HE, K., GIRSHICK, R. & SUN, J., 2016: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. arXiv:1506.01497.
- SHELHAMER, E., LONG, J. & DARRELL, T., 2016: Fully Convolutional Networks for Semantic Segmentation. arXiv:1605.06211.
- SIMONYAN, K. & ZISSERMAN, A., 2015: Very Deep Convolutional Networks For Large-Scale Image Recognition. arXiv:1409.1556.
- SONG, S. & XIAO, J., 2014: Sliding Shapes for 3D Object Detection in Depth Images. In: Proceedings of the 13th European Conference on Computer Vision (ECCV2014). <http://slidingshapes.cs.princeton.edu/paper.pdf>.
- SONG, S. & XIAO, J., 2016: Deep Sliding Shapes for Amodal 3D Object Detection in RGB-D Images. 29th IEEE Conference on Computer Vision and Pattern Recognition, 808-81, <http://dss.cs.princeton.edu/paper.pdf>.
- SZEGEDY, C., LIU, W., JIA, Y., Sermanet, P., REED, S., ANGUELOV, D., ERHAN, D., VANHOUCHE, V. & RABINOVICH, A., 2014: Going deeper with convolutions. arXiv:1409.4842.
- XIE, S., GIRSHICK, R., DOLLÁR, P., TU, Z. & HE, K., 2017: Aggregated Residual Transformations for Deep Neural Networks. arXiv:1611.05431.
- ZHAO, H., SHI, J., QI, X., WANG, X. & JIA, J., 2017: Pyramid Scene Parsing Network. arXiv:1612.01105.

# Herausforderungen in der Fehlerfortpflanzung von Laserscandaten für multitemporale Analysen zur verbesserten Quantifizierung des Level of Detection

LUKAS WINIWARTER<sup>1</sup>, KATHARINA ANDERS<sup>1,2</sup> & BERNHARD HÖFLE<sup>1,2</sup>

*Zusammenfassung: Aussagen über die Signifikanz von quantifizierten Änderungswerten bei multitemporalen Analysen von 3D-Laserscandaten basieren typischerweise auf statistischen Tests. Um aus Sensor- und Prozessierungsfehlern auf ein lokales „Level of Detection“, also eine minimale quantifizierbare Änderung zu schließen, kann das allgemeine Kovarianzfortpflanzungsgesetz angewandt werden. Dabei kommt es durch die temporale Koregistrierung der Daten zu Korrelationen zwischen den einzelnen Laserpunkten, die berücksichtigt werden müssen. Eine naive Anwendung des Kovarianzfortpflanzungsgesetzes führt dabei zu Matrizen, die nicht mehr leicht im Speicher vorgehalten werden können. In diesem Beitrag wird daher aufgezeigt, wie durch analytische Lösung der Matrizenmultiplikation das Problem stark vereinfacht werden kann.*

*Die Analyse terrestrischer Laserscandaten eines Blockgletschers zeigt quantifizierte Änderungen, die mit bisherigen Berechnungsmethoden unter dem Level of Detection lagen, als signifikante Oberflächenänderungen auf. Dabei werden zusätzliche 20 % der Änderungswerte als signifikant erkannt.*

## 1 Einleitung

Bei der Prozessierung von Daten aus terrestrischem Laserscanning werden einzelne 3D-Punkte typischerweise als unkorrelierte Messwerte betrachtet und in dieser Weise in Analysen verwendet. Diese Betrachtung ist gültig, solange der Punktabstand größer als der Laser-Footprint ist, und die Punkte nicht gemeinsam verarbeitet werden. Bei multitemporalen Datensätzen werden jedoch Punktwolken mehrerer Epochen durch Koregistrierung zusammengeführt. Die im Zuge der Koregistrierung angebrachten Transformationen bzw. deren Unsicherheiten gelten aber für alle Daten eines Zeitpunktes und führen somit eine Korrelation zwischen den Punkten ein.

Die Koregistrierung mittels Iterative Closest Points (ICP, BESL & MCKAY 1992) in stabilen Bereichen ist notwendig, damit danach in den übrigen Bereichen Änderungen zwischen den beiden Epochen quantifiziert werden können. Fehler im resultierenden Punktwolkenabstand ergeben sich aus Mess- und Koregistrierungsfehlern. Diese können bei einer einzelpunktbasierter Analyse fortgepflanzt werden, um ein Maß für das Level of Detection, also die minimale statistisch quantifizierbare Änderung pro Punkt, zu erhalten.

Eine genaue Abschätzung dieses Level of Detection ist insbesondere dann wichtig, wenn die erwarteten Änderungen im Verhältnis zum Messrauschen nicht groß sind. Das ist zum Beispiel der Fall, wenn die Messfrequenz erhöht wird, um Daten mit einer höheren temporalen Auflösung zu

---

<sup>1</sup> Ruprecht-Karls-Universität Heidelberg, Geographisches Institut, Im Neuenheimer Feld 368, D-69120 Heidelberg, E-Mail: [lukas.winiwarter,katharina.anders,hoefle]@uni-heidelberg.de

<sup>2</sup> Ruprecht-Karls-Universität Heidelberg, Interdisziplinäres Zentrum für Wissenschaftliches Rechnen, Im Neuenheimer Feld 205, D-69120 Heidelberg

erhalten. Gerade aber diese Datensätze sind wichtig, da sie etwa geographische Analysen erlauben, die den Beitrag verschiedener geomorphologischer Prozesse auf eine kumulierte topographische Oberflächenveränderung untersuchen.

In unserem Beitrag wenden wir daher allgemeine Fehlerfortpflanzung an, um das Level of Detection durch Einbringen von Wissen über den Sensor quantifizieren zu können. Im Gegensatz dazu arbeiten herkömmliche Ansätze rein datengetrieben, wie etwa die Methoden von LAGUE et al. (2013) und JAMES et al. (2016). Aufgrund des Datenvolumens multitemporaler TLS-Punktwolken werden jedoch in der Prozessierung schnell technische Grenzen erreicht, die auf mathematischem Wege umgangen werden können.

## 2 Methoden

Das allgemeine Kovarianzfortpflanzungsgesetz ermöglicht es, Korrelationen in der Varianz-Kovarianzmatrix einzuführen und deren Einfluss auf die Zielgrößen über das linearisierte funktionale Modell zu berücksichtigen. Dabei müssen die Ableitungen des funktionalen Modells nach den Größen gebildet werden, die fehlerbehaftet sind. Aus der direkten Anwendung dieser Methode ergeben sich schnell Datenmengen, die nicht mehr leicht vorgehalten und prozessiert werden können, wie im Folgenden gezeigt wird. Im Anschluss präsentieren wir daher eine auf den Anwendungsfall zugeschnittene, alternative Berechnungsmethode, bei der die explizite Matrizenmultiplikation analytisch durchgeführt wird und sich aufgrund spärlich besetzter Matrizen vereinfacht.

In Gleichung 1 wird das funktionale Modell einer 12-Parameter-Transformation von Punkt  $i$  über reduzierte Koordinaten gezeigt, wie es beispielsweise in der ICP-Implementierung im OPALS-Softwarepaket (PFEIFER et al. 2014; GLIRA et al. 2015) verwendet wird. Die zu transformierenden Einzelpunkte werden dafür zunächst mit Gleichung 2 aus dem Sensorkoordinatensystem in ein kartesisches Koordinatensystem überführt.

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix}_{i,ICP} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}_i - \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} \quad (1)$$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix}_i = r_i \cdot \begin{pmatrix} \cos \varphi_i \sin \theta_i \\ \sin \varphi_i \sin \theta_i \\ \cos \theta_i \end{pmatrix} \quad (2)$$

Gleichung 1 muss nach den 15 fehlerbehafteten Größen der affinen Transformationsmatrix ( $a_{11}$ ,  $a_{12}$ ,  $a_{13}$ ,  $a_{21}$ ,  $a_{22}$ ,  $a_{23}$ ,  $a_{31}$ ,  $a_{32}$  und  $a_{33}$ ), des Translationvektors ( $t_x$ ,  $t_y$  und  $t_z$ ), und der Messungen ( $\varphi_i$ ,  $\theta_i$  und  $r_i$ ) abgeleitet werden, um den Einfluss des Mess- und des Koregistrierungsfehlers auf die Punktlage abzuschätzen. Die Parameter der Koregistrierung und deren Unsicherheit sind aber für den gesamten Datensatz ident, nur Einträge mit dem Index  $i$  sind für den einzelnen Punkt spezifisch. Daher sind die Ableitungen von Gleichung 1 für Punkt  $i$  nach den Einflussgrößen für Punkt  $j$  ungleich null, was im Gesamtmodell zu 12 Ableitungen nach den Transformationsparametern und  $3n$  Ableitungen nach den Messgrößen führt. Bei einer Punktwolke von  $n$  Punkten

sind die Dimensionen im Kovarianzfortpflanzungsgesetz dann wie in Gleichung 3 angegeben. Dabei stellt  $C_{aa}$  die Kovarianzmatrix der fehlerbehafteten Eingangsgrößen,  $F$  das linearisierte funktionale Modell und  $C_{xx}$  die (fiktive) Kovarianzmatrix der gesamten 3D-Laserpunkt wolke dar.

$$C_{xx} = F \cdot C_{aa} \cdot F^T \quad (3)$$

$$(3n \times 3n) \quad (3n \times (12 + 3n)) \quad ((12 + 3n) \times (12 + 3n)) \quad ((12 + 3n) \times 3n)$$

Bei einer Punkt wolke, die beispielsweise mit einem mittleren Punktabstand von 5 mm aufgenommen wurde und eine Fläche von 1 m<sup>2</sup> abdeckt und somit etwa 40 000 Punkte aufweist, hat die Matrix  $C_{xx}$   $(120\,000 \times 120\,000) = 14,4$  Mrd. Einträge. Wird jeder dieser Einträge auch nur als 32-bit Single Precision Float im Speicher vorgehalten, so ergibt sich ein roher Speicherbedarf von etwa 54 GB pro m<sup>2</sup>, ohne Berücksichtigung der Korrelationen, die über diesen einen Quadratmeter hinausgehen. Dies veranschaulicht, dass die Anwendung dieser Methode selbst auf modernen Computern nicht zielbringend sein kann, wenn ein gängiger Datensatz mit mehreren hundert Millionen Punkten verarbeitet werden soll.

Im Folgenden zeigen wir daher einen Ansatz, bei dem die Kovarianz in x, y und z zwischen den Punkten  $i$  und  $j$  über eine Formel explizit ausgedrückt wird. Eine Analyse der partiellen Ableitungen der Transformationsgleichung für den Punkt  $i$  nach den Einflussgrößen für Punkt  $j$  zeigt, dass diese nur mit dem Abstand in den jeweiligen Koordinatenrichtungen variieren. Sie lassen sich damit in vektorisierter Form für alle Punkte der Nachbarschaft berechnen lassen, ohne dass eine Matrix der Dimension  $((12 + 3n) \times (12 + 3n))$  aufgestellt werden muss. Dies ist am Beispiel der Varianz  $\sigma_{x_i x_j}$  der x-Koordinaten von Punkt  $i$  und Punkt  $j$  in Gleichung 4 beschrieben. Die Varianzen der y- und z-Koordinaten und die Kovarianzen der Koordinatenkombinationen lassen sich analog bestimmen.

$$\begin{aligned} \sigma_{x_i x_j} = & (x_i - x_0)^2 \sigma_{a_{11}}^2 + 2 (x_i - x_0)(y_i - y_0) \sigma_{a_{11} a_{21}} + \\ & (y_i - y_0)^2 \sigma_{a_{21}}^2 + 2 (y_i - y_0)(z_i - z_0) \sigma_{a_{21} a_{31}} + \\ & (z_i - z_0)^2 \sigma_{a_{31}}^2 + 2 (z_i - z_0)(x_i - x_0) \sigma_{a_{31} a_{11}} + \\ & 2 \left( (x_i - x_0) \sigma_{a_{11} t_x} + (y_i - y_0) \sigma_{a_{21} t_x} + (z_i - z_0) \sigma_{a_{31} t_x} \right) + \sigma_{t_x}^2 \end{aligned} \quad (4)$$

Für den Fall  $i = j$  werden noch die Terme aus Gleichung 5 addiert, die den Einfluss des Messfehlers beschreiben.

$$\begin{aligned} \sigma_{x_i}^2 = & \sigma_{x_i x_j} + (a_{11} \cos \varphi \sin \theta + a_{12} \sin \varphi \sin \theta + a_{13} \cos \theta)^2 \sigma_r^2 + \\ & (-a_{11} r \sin \varphi \sin \theta + a_{12} r \cos \varphi \sin \theta)^2 \sigma_\varphi^2 + \\ & (a_{11} r \cos \varphi \cos \theta + a_{12} r \sin \varphi \cos \theta - a_{13} r \sin \theta)^2 \sigma_\theta^2 \end{aligned} \quad (5)$$

Bei der Berechnung von punktwolkenbasierten Änderungen zwischen zwei Zeitpunkten mittels Multiscale Model-to-Model Cloud Comparison (M3C2, LAGUE et al. 2013) werden um vorgegebene 3D-Positionen (Core Points) die Laserpunkte des Gesamtdatensatzes pro Epoche auf einen lokalen Normalenvektor  $\vec{n}$  projiziert und dort gemittelt. Bei der Mittelung können für die Fehlerfortpflanzung die jeweiligen Einträge der Varianz-Kovarianz aufaddiert und in der Folge durch

die Anzahl der Punkte  $m$  dividiert werden. Damit erhält man für den gemittelten Punkt eine Varianz-Kovarianzmatrix  $C_{mm}$  nach Gleichung 6.

$$C_{mm} = \frac{1}{m} \begin{pmatrix} \sum_i \sum_j \sigma_{x_i x_j} & \sum_i \sum_j \sigma_{x_i y_j} & \sum_i \sum_j \sigma_{x_i z_j} \\ & \sum_i \sum_j \sigma_{y_i y_j} & \sum_i \sum_j \sigma_{y_i z_j} \\ \text{symm.} & & \sum_i \sum_j \sigma_{z_i z_j} \end{pmatrix} \quad (6)$$

Für die Koregistrierung multitemporaler Datensätze wird ein Datensatz festgehalten und nicht transformiert. Die Punkte dieses Datensatzes werden als untereinander unkorreliert angesehen, daher kann bei der Mittelung in diesem Datensatz die jeweils zweite Summe der Elemente (für alle Fälle  $i \neq j$ ) in Gleichung 6 ignoriert werden. Es macht für die Berechnung keinen Unterschied, welcher der beiden Datensätze als fest und welcher als bewegt angesehen wird.

Der Punktwolkenabstand selbst wird entlang der Zylinderachse als Distanz zwischen den beiden Mittelwerten gebildet. Damit lässt sich die Varianz der Distanz über Gleichung 7 bestimmen. Ein Signifikanztest zeigt auf, ob eine bestimmte Distanz  $d$  signifikant von Null abweicht oder nicht.

$$\sigma_{d,t_0,t_1}^2 = (\vec{n} \quad -\vec{n}) \begin{pmatrix} C_{mm,t_0} & 0 \\ 0 & C_{mm,t_1} \end{pmatrix} \begin{pmatrix} \vec{n} \\ -\vec{n} \end{pmatrix} \quad (7)$$

### 3 Daten und Ergebnisse

An einem Beispieldatensatz des Äußeren Hochebenkars, einem Blockgletscher im Ötztal, Österreich (ZAHS et al. 2019), zeigen wir die Anwendung der entwickelten Methode. Die Aufnahme des Untersuchungsgebiets mit terrestrischem Laserscanning fand seit 2015 jährlich, und im Sommer 2019 zusätzlich alle zwei Wochen zur genaueren Untersuchung der an der Oberflächenänderung beteiligten Prozesse statt. Dabei kamen bei jeder Epoche etwa 450 Millionen Einzelpunkte hinzu. Besonders für die Analyse der zweiwöchigen Zeitschritte ist die Berücksichtigung aller Einzelmessungen wichtig, um durch ein kleines Level of Detection auch kleine Änderungen gesichert quantifizieren und von Messrauschen unterscheiden zu können.

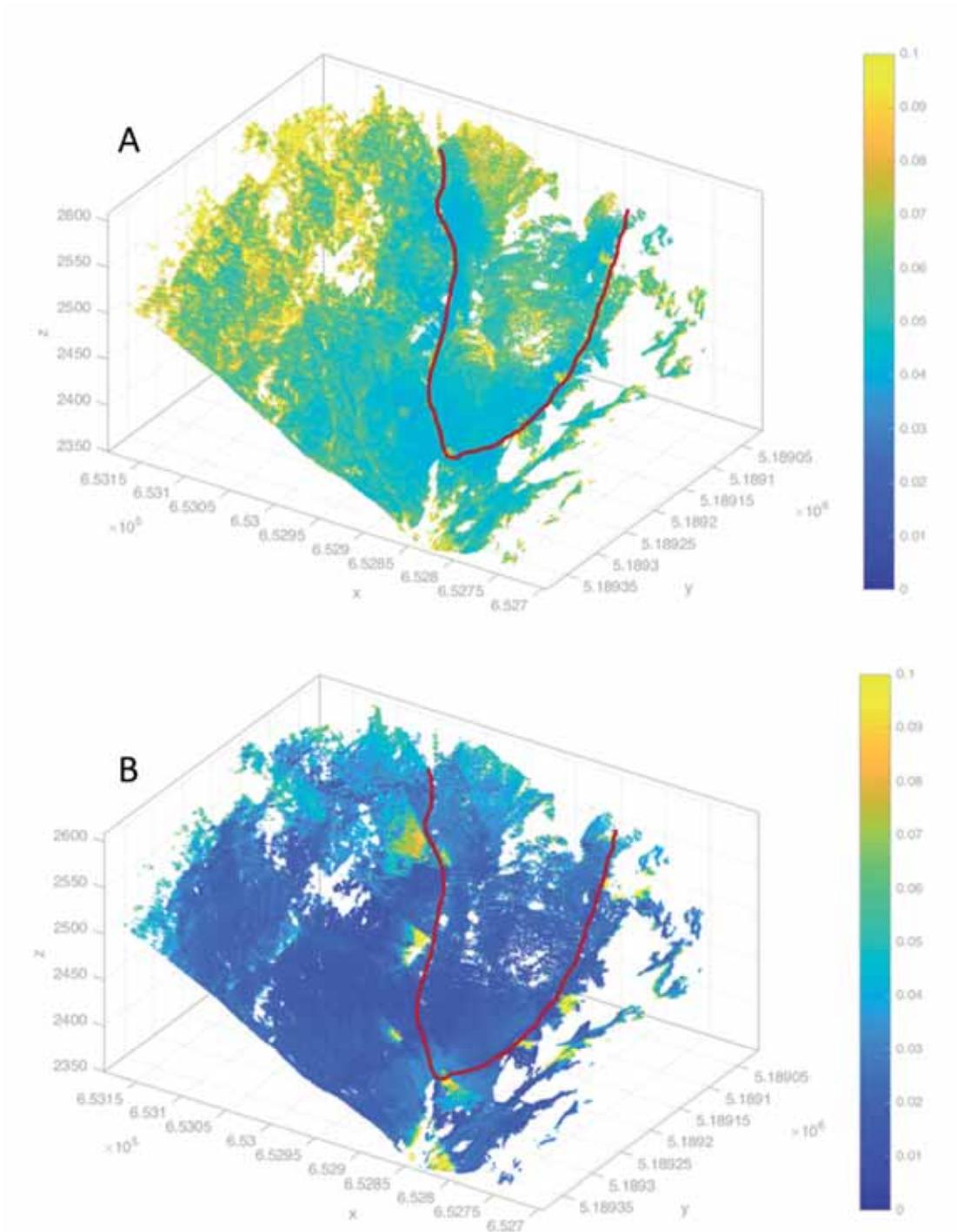


Abb.1: Level of Detection für eine Änderungsanalyse für Punktwolken der Jahre 2017 und 2018. Der Blockgletscher Äußeres Hochebenkar ist in der rechten Bildhälfte zu sehen (rote Markierung). Ergebnisse des (A) datengetriebenen Ansatzes nach LAGUE et al. (2013) mit Werten zwischen 1 und 10 cm und (B) aus dem mit Fehlerfortpflanzung berechneten Level of Detection, mit Werten zwischen 1 und 5 cm

Wir vergleichen unsere Ergebnisse mit denen der datengetriebenen Abschätzung des Level of Detection, und zeigen auf, in welchen Anwendungsfällen die genauere Betrachtung lohnt. Dieser Vergleich ist für eine Zeitdifferenz von einem Jahr (2017-2018) in Abbildung 1 dargestellt. Als Standardabweichung für die Distanzmessung entnehmen wir dem Datenblatt des Sensorherstellers  $\sigma_r = 0.005 \text{ m}$ , für die Winkelmessungen verwenden wir die Energieverteilung innerhalb des Laserstrahls als Näherung, also  $\frac{1}{4}$  der Strahlaufweitung, wenn diese an den  $1/e^2$ -Punkten gemessen wurde:  $\sigma_\varphi = \sigma_\theta = 0.0675 \text{ mrad}$  (RIEGL LMS 2017).

In Bereichen, in denen die Nullhypothese nicht verworfen werden kann, muss davon ausgegangen werden, dass keine signifikanten Änderungen aufgetreten sind. Dies ist in Abbildung 2 für den herkömmlichen Ansatz und den hier vorgestellten Ansatz gezeigt. Deutlich erkennbar sind Bereiche auch abseits des Blockgletschers, die mit der neuen Methode als nicht stabil angesehen werden müssen. Etwa 20 % der Core Points werden zusätzlich als signifikant markiert. Bei genauerer Betrachtung handelt es sich dabei insbesondere auch um Bereiche, bei denen Vegetationswachstum zu einer Oberflächenveränderung führt (in Abbildung 2 mit (1) markiert). Im Gegensatz dazu zeigen die Bereiche, bei denen keine signifikante Veränderung quantifiziert werden konnte, größtenteils anstehendes Gestein, das fest mit dem Untergrund verbunden ist (in Abbildung 2 mit (2) markiert). Zur Anwendung der präsentierten Methode, werden gegenüber den rein datengetriebenen Ansätzen zusätzliche Informationen benötigt. Diese sind insbesondere die a-priori Genauigkeitsinformation der eingesetzten Sensoren sowie die Koordinaten der einzelnen Standpunkte des Laserscanners in den jeweiligen Kampagnen. Darüber hinaus muss die Kovarianzmatrix der Koregistrierung vorliegen, wobei diese gegebenenfalls auch aus den Daten geschätzt werden kann – etwa, indem eine erneute ICP durchgeführt wird.

Die Laufzeit der Methode beträgt auf einem Computer mit einer Intel i9-CPU (7. Gen) und 64 GB RAM etwa 30 Minuten für 200.000 Core Points, wobei die Anzahl der in der Nachbarschaftssuche genutzten Laserpunkte um etwa einen Faktor 500 größer ist. Im Vergleich dazu steht die meistgenutzte Implementierung des M3C2 als herkömmliche Methode, die als Plugin in der Software CloudCompare für dieselben Daten nur wenige Sekunden benötigt.

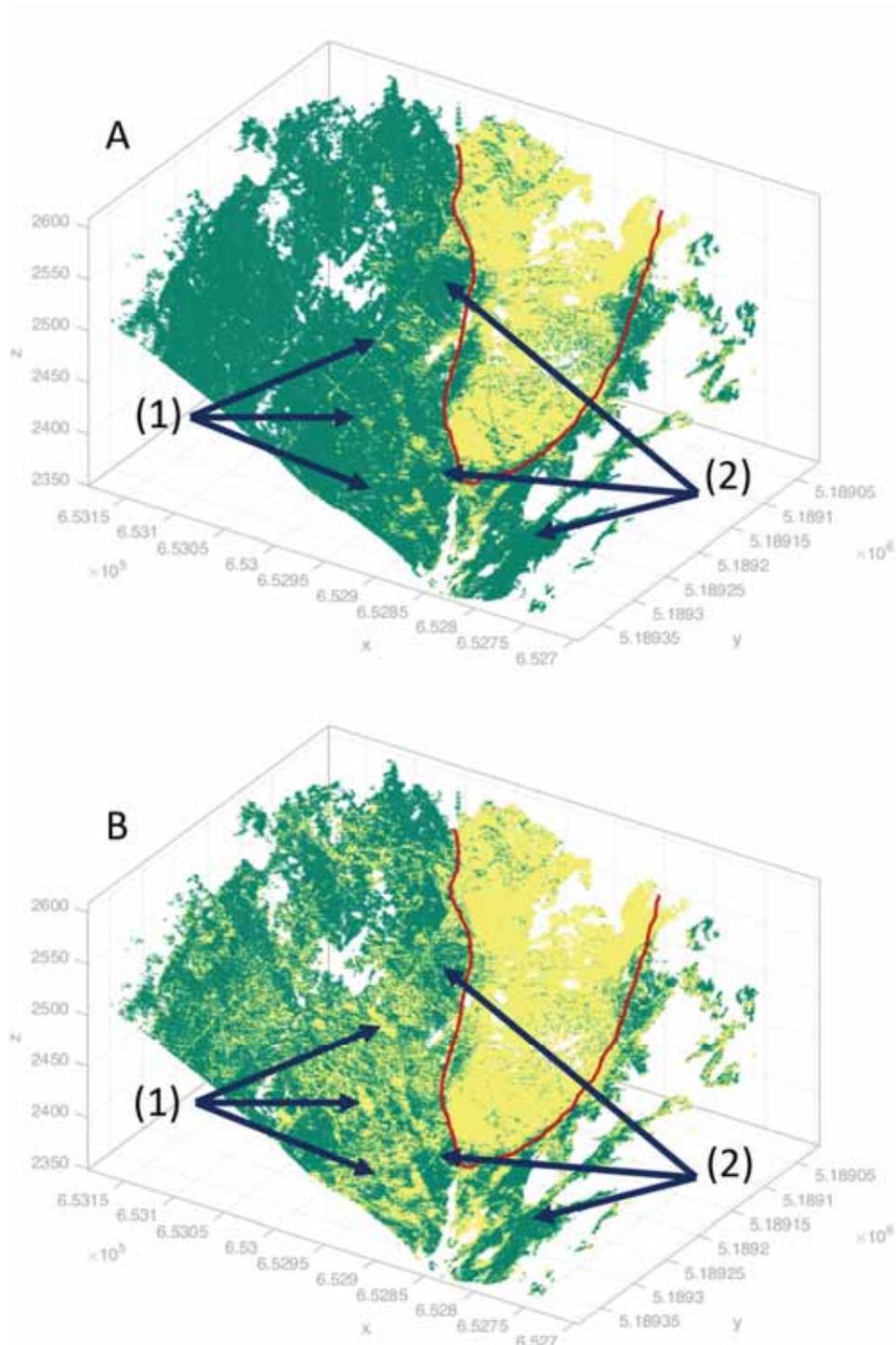


Abb.2: Darstellung signifikanter Veränderungen (gelb) für den Datensatz aus Abbildung 1. (A) datengetriebener Ansatz nach Lague et al. (2013), (B) aus Fehlerfortpflanzung bestimmte Signifikanz. Zusätzlich erkennbar sind die nun als signifikant ausgewiesenen Bereiche (1), in denen es zu Veränderungen durch Vegetationswachstum kommt. Felsbereiche (2) werden aber weiterhin als stabil erkannt

## 4 Fazit und Ausblick

In diesem Beitrag stellen wir einen neuen Ansatz vor, der Wissen über die Sensoren und die Datenprozessierung in die Genauigkeitsanalyse von Punktwolkenänderungsmaßen einbringt. Erste Ergebnisse zeigen, dass der Ansatz praxistauglich und relevant ist, auch wenn gegenüber den herkömmlichen Methoden ein vielfaches der Berechnungszeit benötigt wird.

Mögliche Erweiterungen der Methode sind die Implementierung unterschiedliche Sensormodelle, etwa ALS und ULS, wobei auch eine beliebige Kombination möglich sein soll. Damit können dann z.B. historische ALS-Modelle in Analysen verwendet werden, ohne, dass dabei die Qualität der aktuelleren Daten wie bislang schlechter angenommen werden muss, als sie es ist.

Zusätzlich kann bei einer zukünftigen Erweiterung des Ansatzes das hier diskutierte Fehlermodell insbesondere in Bezug auf den Distanzfehler kritisch betrachtet werden, und der Einfluss von Signalstärke, Einfallswinkel und weiteren Größen untersucht und im Modell inkludiert werden.

## 5 Literaturverzeichnis

- BESL, P. J. & MCKAY, N. D., 1992: Method for registration of 3-D shapes. *Sensor fusion IV: control paradigms and data structures* **1611**, 586-606, doi: 10.1117/12.57955.
- GLIRA, P., PFEIFER, N., BRIESE, C. & RESSL, C., 2015: A Correspondence Framework for ALS Strip Adjustments based on Variants of the ICP Algorithm, *Photogrammetrie-Fernerkundung-Geoinformation* **2015(4)**, 275-289, doi: 10.1127/pfg/2015/0270.
- JAMES, M. R., ROBSON, S. & SMITH, M.W., 2017: 3-D uncertainty-based topographic change detection with structure-from-motion photogrammetry: precision maps for ground control and directly georeferenced surveys: 3-D uncertainty-based change detection for SfM surveys. *Earth Surface Processes and Landforms* **42**, 1769-1788, doi: 10.1002/esp.4125.
- LAGUE, D., BRODU, N. & LEROUX, J., 2013: Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (N-Z). *ISPRS Journal of Photogrammetry and Remote Sensing* **82**, 10-26, doi: 10.1016/j.isprsjprs.2013.04.009.
- PFEIFER, N., MANDLBURGER, G., OTEPKA, J. & KAREL, W., 2014: OPALS—A framework for Airborne Laser Scanning data analysis. *Computers, Environment and Urban Systems* **45**, 125-136, doi: 10.1016/j.compenvurbsys.2013.11.002.
- RIEGL LMS, 2017: Riegl VZ-2000i (datasheet). [http://www.riegl.com/uploads/tx\\_pxriegl/downloads/RIEGL\\_VZ-2000i\\_Datasheet\\_2019-11-22.pdf](http://www.riegl.com/uploads/tx_pxriegl/downloads/RIEGL_VZ-2000i_Datasheet_2019-11-22.pdf).
- ZAHS, V., HÄMMERLE, M., ANDERS, K., HECHT, S., SAILER, R., RUTZINGER, M., WILLIAMS, J. G. & HÖFLE, B., 2019: Multi-temporal 3D point cloud-based quantification and analysis of geomorphological activity at an alpine rock glacier using airborne and terrestrial LiDAR. *Permafrost and Periglacial Processes* **30(3)**, 222-238, doi: 10.1002/ppp.2004.

# Änderungsdetektion in photogrammetrischen Punktwolken für das Monitoring hochalpiner, gravitativer Massenbewegungen – Beispiel Hochvogel

AGNES R. DINKEL<sup>1</sup>, LUDWIG HOEGNER<sup>1</sup>, ADRIAN EMMERT<sup>1</sup>, LUKAS RAFFL<sup>1</sup> & UWE STILLA<sup>1</sup>

*Zusammenfassung: Im Zentrum der Untersuchung steht die Evaluierung der photogrammetrischen Messmethode zur Änderungsdetektion von gravitativen Massenbewegungen in hochalpinen Gebieten am Beispiel des Hochvogel (2592 m, Allgäuer Alpen). Hierfür wurden mehrere Epochen des Gipfelbereichs aufgenommen und über ein Festpunktfeld orientiert. Zur besseren Verknüpfung wurden für jede Epoche zusätzliche Marken angebracht. Die im Rahmen einer Bündelausgleichung mit anschließendem Dense-Matching erzeugte photogrammetrische Punktwolke wird sowohl mit punktuellen Tachymetermessungen als auch einem für einen Teilabschnitt durchgeführten terrestrischen Laserscan verglichen. Anschließend wird über die M3C2-Distanz zwischen den Punktwolken der verschiedenen Epochen die Veränderung im Gipfelbereich bestimmt. Die hieraus abgeleiteten Bewegungen decken sich sowohl mit den Messungen von Tachymeter und Laserscanner als auch den Messdaten der Geologie.*

## 1 Einleitung

Der Klimawandel und die damit verbundenen Folgen sind im letzten Jahr stark ins öffentliche Interesse gerückt. So verursacht die Klimaerwärmung auch das Abtauen des Permafrosts in den hochalpinen Lagen. Dies, oft in Wechselwirkung mit den ebenfalls Klimawandel-bedingten extremen Wetterlagen v. a. Starkregenerereignissen, führt zu vermehrten gravitativen Massenbewegungen in den alpinen Regionen wie Bergstürze oder Hangrutschungen. Wesentlich für das Verständnis und die Beurteilung der Hangbewegungen ist das Feststellen des Bewegungstyps bzw. deren Klassifikation. Dabei werden fünf unterschiedliche Bewegungstypen unterschieden: Fallen, Kippen, Rutschen, Ausbreiten und Fließen (HUNGR et al. 2014). Um aber diese Klassifikation durchzuführen, ist eine genaue Beobachtung von Bewegungen notwendig. Dabei ist die alleinige Feststellung von Bewegungsraten bzw. -strecken eines einzelnen Abstands über feste Zeiträume nicht ausreichend. Vielmehr ist es notwendig unterschiedliche Distanzen im dreidimensionalen Raum über einen längeren Zeitraum zu überwachen, um die Klassifikation und die anschließende Analyse des Befunds vornehmen zu können. Zu diesem Zweck werden unterschiedliche Überwachungsmethoden verwendet. Neben komplexen Konfigurationen geologischer Messgeräte wie beispielsweise von Extensometern (MALET et al. 2002; MASSEY et al. 2013) finden auch geodätische Messungen mit Tachymeter, GNSS-Antennen (SQUARZONI et al. 2005), Radarinterferometrie oder Laserscanner Anwendung (ROBERTI et al. 2018; OPPIKOFER et al. 2009). Diese Systeme sind alle teuer in der Anschaffung und müssen durch fachlich versiertes Personal bedient werden. Eine flächendeckende Überwachung ist auf diese Art kaum möglich. Kameras auf der anderen Seite

---

<sup>1</sup> Technische Universität München, Arcisstraße 21, D-80333 München, E-Mail: [Agnes.Dinkel, Ludwig.Hoegner, Adrian.Emmert, Lukas.Raffl, Stilla]@tum.de

sind vergleichsweise günstig und mobil und können so leicht eingesetzt werden. Eine flächendeckende Überwachung möglicher Gefährdungen ist dann durch Laien möglich (ROMEO et al. 2019; MAYR et al. 2019; PEPPA et al. 2019; KROMER et al. 2019; ESPOSITO et al. 2017; WARRICK et al. 2016; PIERMATTEI et al. 2016; ELTNER et al. 2016; STUMPF et al. 2015). In diesem Beitrag werden die erreichten Genauigkeiten der Punktbestimmung sowie der Änderungsdetektion in photogrammetrischen Punktwolken im Rahmen des AlpSenseBench Modellprojekts (ALPSENSEBENCH 2018) im Gipfelbereich des Hochvogels mit den durch die Ingenieurvermessung erhobenen Daten verglichen.

## 2 Beschreibung der Mess- und Auswerteverfahren

Entscheidend für den Einsatz der Photogrammetrie, unter welchen Bedingungen photogrammetrisch erzeugte Punktwolken die erforderlichen Genauigkeiten und Zuverlässigkeiten insbesondere im Hinblick auf eine Änderungsdetektion erreichen können. Die Untersuchung gliedert sich daher in zwei Teile: i) Genauigkeit und Dichte der erfassten Punktwolke und ii) Genauigkeit und Zuverlässigkeit der Änderungsdetektion zwischen zwei Punktwolken.

Es wird für die späteren Versuche davon ausgegangen, dass ein lokales Festpunktfeld vorliegt, das über die Epochen hinweg stabil ist und für jede einzelne Epoche um weitere Messmarken ergänzt wird, die der Orientierung und Verknüpfung der Bilder dienen. Außerdem wird die Kamera als geometrisch kalibriert mit Festbrennweite und Fixfokus angenommen. Exemplarisch wird eine Aufnahmekonfiguration für die Überwachung einer Felsspalte modelliert.

### 2.1 Photogrammetrische 3D Rekonstruktion

Die photogrammetrischen Messungen beruhen auf einer Bündelblock-Ausgleichung zur Bestimmung der Bildorientierungen gefolgt von einem Semi-Global Matching (HIRSCHMÜLLER 2008) zur Erzeugung einer dichten Punktwolke. Für Messungen an breiten Spalten und Abbrüchen, die nur von einer Seite aus zugänglich sind, wird eine Bildüberdeckung von ca. 80% angestrebt. Dafür wird die Kamera in kurzen Abständen (ca. 0,5 m) parallel zur Spalte versetzt. Dabei werden für die Aufnahme der Bilder zwei Messkonfigurationen verwendet:

- Doppelkanten-Konfiguration: Entlang der Spalte werden Aufnahmen der Spalte angefertigt, die beide Außenkanten der Spalte enthalten. Die Aufnahmepositionen verschieben sich dabei parallel mit dem Verlauf der Spalte.
- Spalten-Konfiguration: Um mehr Daten aus dem Spalteninneren zu sammeln, wird in dieser Konfiguration die Kamera über die Spalte geneigt. Dabei enthalten die Bilder zumeist nicht die eigentlichen Oberkanten des Geländes. Auch diese Aufnahmen folgen dem Verlauf der Spalte.

Für kleinere, von beiden Seiten frei begehbar Spalten werden Bilder aus verschiedenen Perspektiven ebenfalls mit hohem Überlapp aufgenommen.

Bestandteil der Bündelausgleichung sind auch die Punkte des lokalen Festpunktnetzes sowie zusätzliche zum Teil ebenfalls eingemessene photogrammetrische Marken. Diese Passpunkte werden verwendet, um für Bilder die Näherungswerte der äußeren Orientierung zu bestimmen. Bei mindestens sechs sichtbaren Passpunkten wird eine Direkte Lineare Transformation (DLT) (LUHMANN 2018) durchgeführt. Mit mindestens drei bis maximal fünf sichtbaren Passpunkten

wird die äußere Orientierung nach dem "Minimale-Objekt-Information"-Verfahren (MOI-Verfahren) (LUHMANN 2018) geschätzt. Sind in einem Bild mehr als die drei nötigen Passpunkte vorhanden, so werden diejenigen ausgewählt, die die größte Dreiecksfläche im Bild aufspannen. Dies soll die Zahl an falsch zugeordneten homologen Bildpunkten reduzieren. Zusätzlich wird eine Varianzkomponentenschätzung verwendet, um die unterschiedlichen Genauigkeiten der gemessenen Passpunkte und der Bildpunkte abbilden zu können. Außerdem wird für die Ausgleichung das Ransac-Verfahren (FISCHLER & BOLLES 1984) eingesetzt, um fehlerhafte Punktzuordnungen zu entfernen. Als Ergebnis der Bündelausgleichung werden die verbliebenen Restfehler und geschätzten Genauigkeiten der Orientierungen, Bild- und Objektpunkte bestimmt.

## 2.2 Änderungsdetektion und Deformationsanalyse

Die Änderungsdetektion in Punktwolken lässt sich grob in drei Arbeitsschritte unterteilen: i) Koregistrierung, ii) Bestimmung des Rauschens, iii) Detektion signifikanter Änderungen. Auf Grund des vorhandenen Passpunktenetzes wird davon ausgegangen, dass eine erneute Koregistrierung nicht notwendig ist. Damit verbleiben die Bestimmung des Rauschniveaus und die eigentliche Änderungsdetektion.

Für die Bestimmung des Rauschens werden zwei Messungen mit zeitlichem Abstand am selben Tag wiederholt unter der Annahme, dass es in dieser kurzen Zeit keine geometrischen Veränderungen gegeben hat. Unterschiede in den beiden erzeugten Punktwolken beider Epochen desselben Tages rühren daher überwiegend aus Lichtveränderungen, Schatten, sowie der Szenenabtastung durch den Sensor und die Kameraorientierungen. Diese Unterschiede sollen als Rauschen ermittelt werden, um eine untere Grenze für die Detektion von Änderungen festzulegen. Als Basis für die Änderungen wird zu jedem Punkt von Punktwolke 1 der nächste korrespondierende Punkt aus Punktwolke 2 gesucht. Dies geschieht auf Basis des (Multiscale-Model-to-Model-Cloud Verfahrens (M3C2) (LAGUE et al. 2013). Aus dem Histogramm der kürzesten Abstände wird ein Schwellwert für das Rauschen ermittelt.

## 3 Untersuchung am Beispiel Hochvogel

Der auf der deutsch-österreichischen Grenze nahe Obersdorf gelegene Berg Hochvogel stellt mit seinen 2592 m den höchsten Berg der Allgäuer Alpen dar und zeichnet sich durch eine starke Erosion, eine intensive Verwitterung des Gesteins und große Spannungen durch übersteile, hohe Felswände in Kombination mit dem brüchigen Gestein aus (SCHOLZ & SCHOLZ 1981; JACOBSHAGEN et al. 2014). Diese geologischen Charakteristika bilden zugleich die Ursache für die umfangreichen Bewegungsprozesse und die zahlreichen Felsstürze in diesem Areal (DAV DONAUWÖRTH 2017). Der Gipfelbereich des Berges ist heutzutage geprägt von einer großen, den gesamten Berg durchziehenden Spalte geprägt, die knapp südlich des Gipfelkreuzes von Nordosten nach Südwesten verläuft (Abb. 1). Daneben haben sich noch weitere kleinere Risse vor allem im südlichen Bereich des Gipfels gebildet, die nahezu im rechten Winkel zu der Hauptspalte, von Südost nach Nordwesten, verlaufen. Der von der Spalte abgetrennte Teil des Gipfels befindet sich in einer gravitativen Rutschbewegung nach Süden (LEINAUER et al. 2019). Die genaue, dreidimensionale Beschreibung dieser Bewegung bildet das Ziel der Messungen.

Neben dem Festpunktfeld der Ingenieurgeodäsie werden für die photogrammetrischen Messungen zusätzliche Passmarken als Passpunkte und Verknüpfungspunkte verwendet. Die Aufnahme erfolgt während der drei Messkampagnen (09.07.2018, 27.09.2018 und 19.07.2019) mittels Stangenstativ, Kamera (Sony Alpha7 RII) mit Weitwinkelobjektiv und einem Endgerät (Laptop/ Smartphone), das mittels WiFi-Verbindung die Kamera steuert (Abb. 2). Die Aufnahme der Spalte erfolgt dem Verlauf der Spalte folgend in mehreren Bahnen mit hoher seitlicher Überlappung (ca. 80 %).



Abb. 1: Gipfelbereich des Hochvogel mit markanter von Nordosten nach Südwesten verlaufenden Spalte (Foto von HEIBEL & FIGL 2014)



Abb. 2: Photogrammetrische Vermessung der Spalten im Gipfelbereich des Hochvogels mit Stangenstativ, Kamera und WiFi-Steuerung

Wie die Daten in Tabelle 1 zeigen, sind eine hohe Zahl an Bildern und zahlreiche Passpunkte aufgenommen worden. Bei der Auswertung mit einer photogrammetrischen Standardsoftware (Pix4D) ergeben sich jedoch Probleme, die dazu führen, dass nur ein Teil der Bilder verwendet werden können, zusätzliche manuelle Verknüpfungspunkte angelegt und eine hohe Zahl an Bildmessungen durchgeführt werden müssen. Zum einen führen starke Licht-Schatten-Wechsel in Kombination mit dem Wandern der Schatten über die gesamte Messperiode, die sich durch das anspruchsvolle Gelände in die Länge zieht, zu falschen Verknüpfungen und Problemen beim Matching von Punkten. Starke Windböen und Schwierigkeiten beim ruhigen Positionieren des Stativs führen darüber hinaus zu unscharfen Bildern. Zudem kann sich auch das Gelände während der Messung selbst verändern, da sich aufgrund der witterungstechnisch stark zeitlich eingeschränkten und begrenzten Möglichkeiten, Messungen am Gipfel durchzuführen, meist mehrere Messtrupps zugleich im Gelände befinden und arbeiten (Abb. 3) oder Besucher den Gipfel frequentieren. Letztlich führen auch die räumlich eng begrenzten Bewegungsmöglichkeiten am Gipfel „um“ das beobachtete Objekt zu schlechten, sehr eingeschränkten Beobachtungsgeometrien und schleifenden Schnitten. Dies führt dazu, dass bei photogrammetrischen Vermessungen im hochalpinen Bereich nicht mit den gleichen Genauigkeiten zu rechnen ist, wie sie im Labor möglich sind.

Tab. 1: Übersicht über die photogrammetrischen Messungen am Hochvogel 2018/19

Kampagne	Bilder	Bilder verwendet	Passpunkte	manuelle Verknüpfungspunkte	Bildmessungen
09.07.2018	1057	407	32	20	2215
27.09.2018	1277	645	55	47	4910
19.07.2019	1243	510	74	55	4055



a)



b)

Abb. 3: Arbeiten am Hochvogel: Herausforderungen durch parallele Arbeiten verschiedener Messtrupps  
a) geringer Platz – b) Veränderungen während der Messarbeiten

## 4 Auswertung der Punktwolken

Die einzelnen, mittels des Festpunktfeldes georeferenzierten Punktwolken werden in CloudCompare zunächst von Störungen befreit (manuell oder mittels SOR-Algorithmus) und im Anschluss mittels M3C2-Algorithmus miteinander Verglichen. Es erfolgt keine Koregistrierung über den ICP-Algorithmus.

Um die Ergebnisse einordnen zu können, wird hierbei zunächst die Genauigkeit der photogrammetrischen Messungen untersucht. Dabei ist die innere Genauigkeit der Bündelausgleichung im Subpixelbereich nicht aussagekräftig für die absolute Genauigkeit einzelner 3D Punkte. Da die Aufnahmegeometrie mit konstanter Blickrichtung und Versatz in einer Richtung dem Stereo-Normalfall ähnelt, kann aus diesem für die vorhandenen Daten eine Genauigkeit in der Entfernung von ca. 2 mm für die photogrammetrische Messmethode angenommen werden. Diese Größenordnung wird auch durch den Vergleich der photogrammetrischen Punktwolke mit der TLS-Punktwolke (Messgenauigkeit ca. 2 mm) (WUNDERLICH et al. 2019) der gleichen Messkampagne bestätigt (Abb. 4), die beide über dieselben Festpunkte georeferenziert werden. Der Mittelwert der Distanzen zwischen den Punktwolken liegt bei 0,94 mm und einer Standardabweichung von 5,6 mm. Abweichung finden sich primär in Bereichen des Felsen, in denen sich lockeres Gestein befindet. In Bereichen aus zusammenhängenden, blanken Felsstrukturen ist dagegen die Deckungsgenauigkeit besonders ausgeprägt.

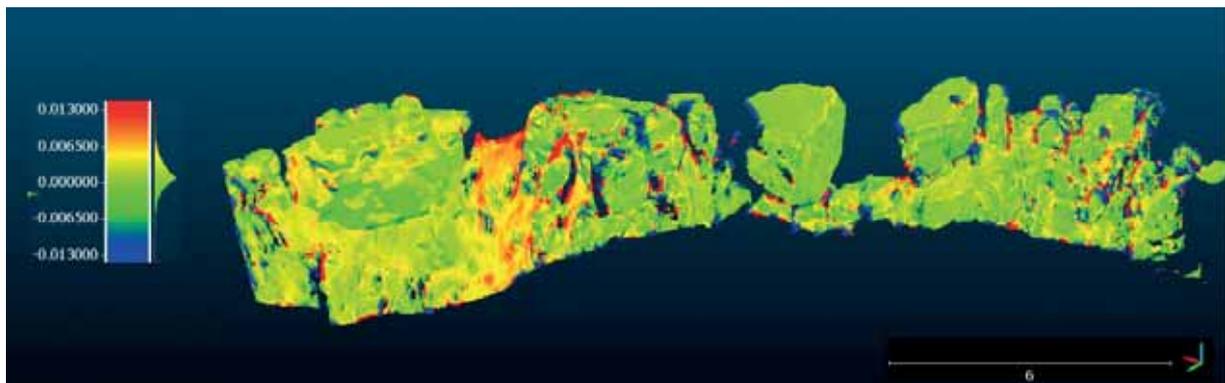


Abb. 4: Vergleich der photogrammetrischen Punktwolke der Rutschmasse am Gipfel des Hochvogel mit zeitgleicher TLS-Punktwolke (grün = höchste Deckungsgenauigkeit)

Daraus ergibt sich, dass Veränderungen signifikant festgestellt werden können, die eine Größenordnung von 8,2 bis 13,2 mm (95 % Zuverlässigkeit) bzw. 11,6 bis 18,6 mm (99 % Zuverlässigkeit) besitzen. Erst durch diese Abschätzung können die Ergebnisse des Vergleichs von Punktwolken unterschiedlicher Messepochen eingeordnet werden. Besonders aussagekräftig ist hierbei die Betrachtung des Vergleichs der georeferenzierten Punktwolken der Rutschmasse von September 2018 und Juli 2019 (Abb. 5). Wie die Verteilung im Histogramm der Distanzen zeigt, lassen sich in ihr mehrere Peaks bei 1,4, 8 und 1,6 mm ablesen. Trennt man die Vertreter der jeweiligen Peaks auf, so scheinen sie zu unterschiedlichen Bereichen mit klarer Charakteristik zu gehören: der erste Peak (-1,4 mm) tritt in Bereichen auf, die sich unten innen an der Rutschmasse befinden. Der Peak von 8 mm findet sich nahezu ausschließlich an horizontalen Fläche, wogegen der Peak von 1,6 mm sich bei vergleichsweise vertikal zum Beobachter stehenden Flächen findet. Damit umschreiben

diese Bewegungen, die sich an der Rutschmasse vollziehen. Es scheint Abschnitte zu geben, die sich zurück in die Spalte entwickeln (-1,4 mm), dagegen scheinen andere Bewegungen sich wiederum von der Spalte weg zu entwickeln, wobei 8 mm eine Abwärts- und 1,6 mm eine Wegbewegung zu beschreiben scheint.

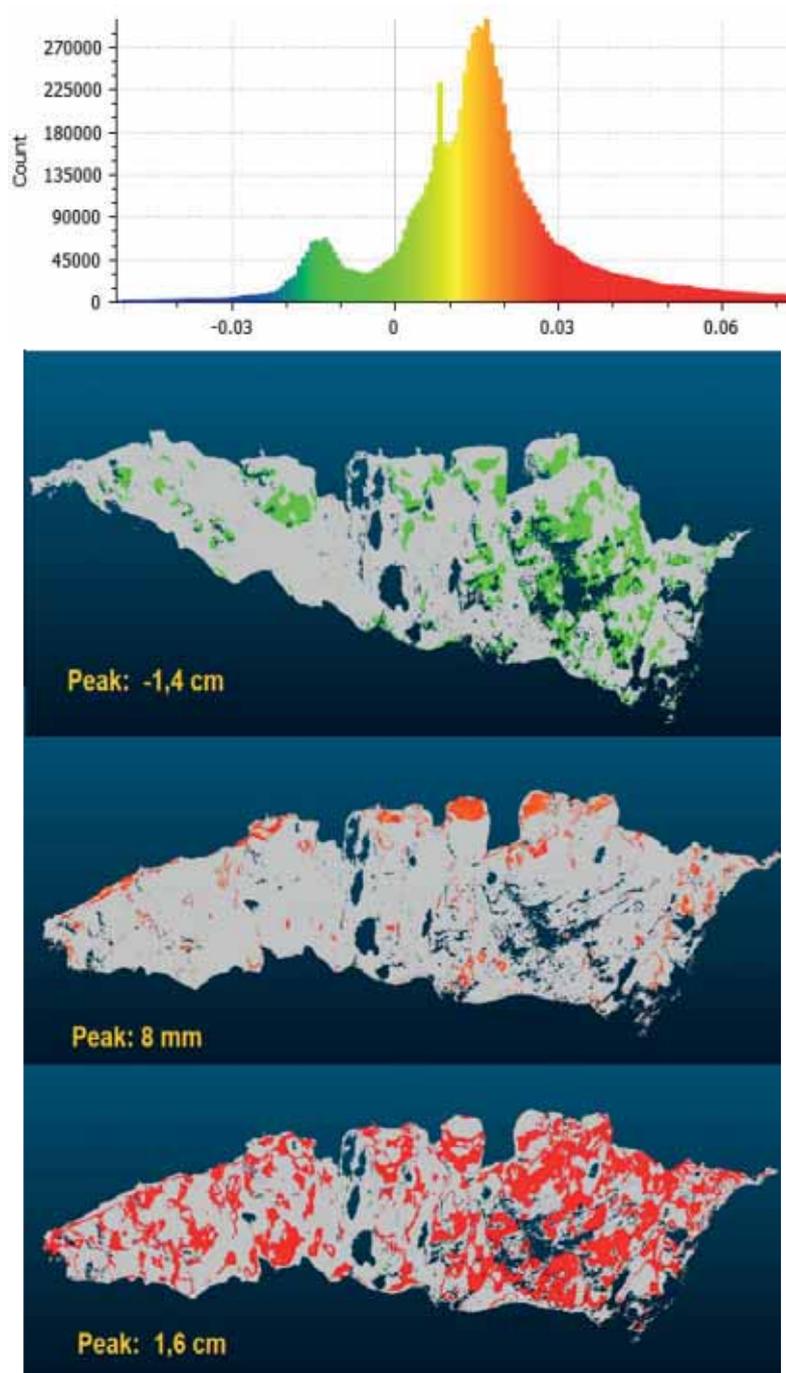


Abb. 5: Vergleich der Punktwolken von September 2018 und Juli 2019: Histogramm der Distanzen zwischen den Punktwolken (oben) und Auflösung der Distanzpunkte nach den Peaks im Histogramm (unten)

Es zeigt sich, dass die Masse keineswegs einheitlich als homogener Block zu rutschen scheint, sondern auch unterschiedliche Bereiche sich unterschiedlich stark zu bewegen scheinen. Darüber hinaus scheinen Flächen senkrecht zu Teilrichtungen der Bewegung und die dort ablesbaren Veränderungen als Zerlegung der Gesamtbewegung der Rutschmasse und somit als Indikatoren für die Rutschbewegung dienen zu können.

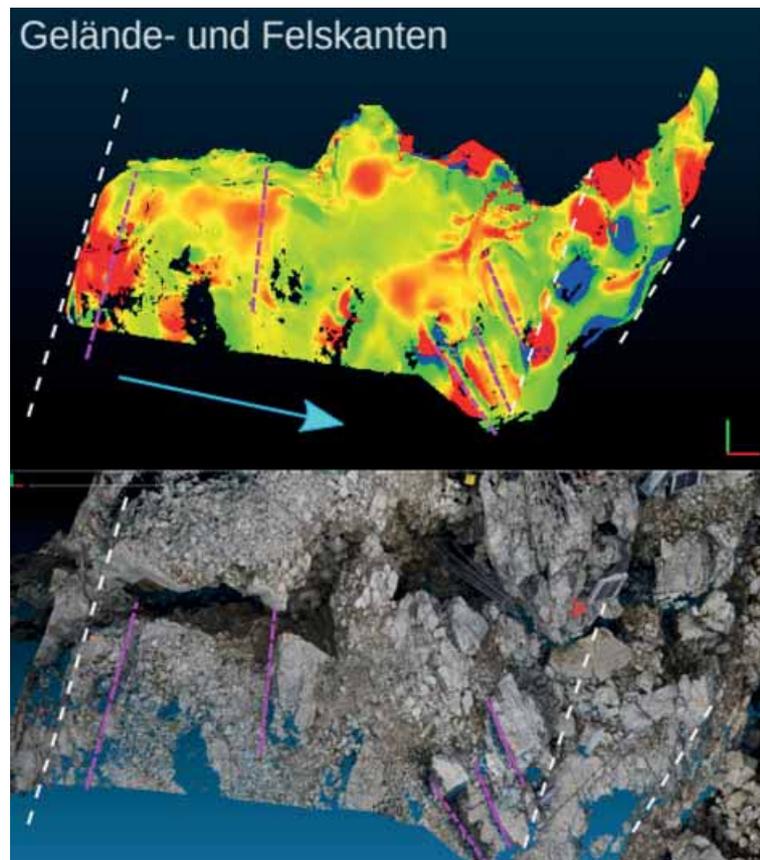


Abb. 6: Vergleich der Punktwolken von September 2018 und Juli 2019 im Bereich einer Nebenspalte am Gipfel des Hochvogel: Beobachtung von Kantenkonstellationen

Verfolgt man diese Beobachtung weiter, so muss um eine dreidimensionale Bewegung beschreiben zu können, nicht die gesamte Masse flächig beobachtet werden. Es müssen signifikante Kanten und Flächen oder noch besser Konstellationen von diesen und deren Veränderung über die verschiedenen Messepochen beobachtet werden. Aus der Summe dieser Teilbewegungen ergibt sich die Gesamtbewegung der Rutschmasse. Eine Kante ist dabei stets nur signifikant für die Richtung zu der sie orthogonal steht. Diese These lässt sich an den Nebenspalten des Gipfels überprüfen (Abb. 6) und scheint sich dort zu bestätigen. Kanten die orthogonal zu der Bewegungsrichtung stehen, fallen im Vergleich der Punktwolken besonders auf.

## 5 Fazit

Für den Hochvogel lässt sich auf diese Weise eine Bewegung für die Rutschmasse beschreiben, die sich aus einem weg- und abwärtsdriften, sowie einem zurücksacken zusammensetzt. Damit beschreiben die Richtungsvektoren eine klassische Rutschbewegung. Die Detektion von einzelnen Flächen und Kanten bzw. Kantenkonstellationen zur Beschreibung von dreidimensionalen Bewegungen scheint ein erfolgversprechender Ansatz zu sein, für den noch ein allgemein anwendbares Instrumentarium und Methodiken entwickelt werden müssen. Um die hier paraphrasierten Ergebnisse vom Hochvogel auf andere Rutschmassen zu übertragen und verallgemeinern zu können, ist es allerdings notwendig, diese Thesen an weiteren Beispielen zu überprüfen, zu verifizieren und zu verbessern.

## 6 Literaturverzeichnis

- ALPSENSEBENCH, 2018: AlpSenseBench Studie. Professur für Hangbewegung, TU München. <https://www.bgu.tum.de/landslides/alpsensebench/projekt/>, Stand 21.12.2019.
- DAV DONAUWÖRTH, 2017: Sperrung Bäumenheimer Weg. Gutachten der Landesgeologie Tirol 2014, 2015 und 2017. <http://www.dav-donauwoerth.de/index.php/3-gutachten-2017>, last visited: 2019-09-09.
- ELTNER, A., SCHNEIDER, D. & MAAS, H.-G., 2016: Integrated processing of high resolution topographic data for soil erosion assessment considering data acquisition schemes and surface properties. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **41**(B5), 813-819.
- ESPOSITO, G., MASTROROCCHO, G., SALVINI, R., OLIVETI, M. & STARITA, P., 2017: Application of UAV photogrammetry for the multi-temporal estimation of surface extent and volumetric excavation in the Sa Pigada Bianca open-pit mine, Sardinia, Italy. *Environmental Earth Sciences* **76**(3), 103.
- Fischler, M.A. & Bolles, R.C., 1981: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* **24**(6), 381-395.
- HEIBEL, G. & FIGL, T., 2014: Gemeinde Hinterhornbach – Hochvogel; Abklärung der aktuellen Steinschlag-, Blocksturz und Felssturzsituation; Stellungnahme der Landesgeologie. Amt der Tiroler Landesregierung.
- HIRSCHMÜLLER, H., 2008: Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**(2), 328-341.
- HUNGR, O., LEROUÉIL, S. & PICARELLI, L., 2014: The Varnes classification of landslide types, an update. *Landslides* **11**(2), 167-194.
- JACOBSSHAGEN, V., SCHWERD, K. & HORNING, T., 2014: Geologische Karte von Bayern 1:25000. Erläuterungen zum Blatt 8628 Hochvogel. Landesamt für Umwelt.
- KROMER, R., WALTON, G., GRAY, B. & HUTCHINSON, J., 2019: Lower cost fixed installation photogrammetry rockfall monitoring. In *Geophysical Research Abstracts* **21**.
- LAGUE, D., BRODU, N. & LEROUX, J., 2013: Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (NZ). *ISPRS journal of Photogrammetry and Remote Sensing* **82**, 10-26.

- LEINAUER, J., JACOBS, B. & FRIEDRICH, B., 2019: Process dynamics and early warning strategies for a preparing highmagnitude rock slope failure at the Hochvogel (Allgäu Alps). In: EGU General Assembly Conference Abstracts, **21**, 2196.
- LUHMANN, T., 2018: Nahbereichsphotogrammetrie: Grundlagen-Methoden-Beispiele. Wichmann, 2018.
- MALET, J. P., MAQUAIRE, O. & CALAIS, E., 2002: The use of Global Positioning System techniques for the continuous monitoring of landslides: application to the Super-Sauze earthflow (Alpes-de-Haute-Provence, France). *Geomorphology* **43**(1-2), 33-54.
- MASSEY, C. I., PETLEY, D. N. & MCSAVENEY, M. J., 2013: Patterns of movement in reactivated landslides. *Engineering Geology* **159**, 1-19.
- MAYR, A., RUTZINGER, M. & GEITNER, C., 2019: Object-Based Point Cloud Analysis for Landslide and Erosion Monitoring. *Photogrammetric Engineering & Remote Sensing* **85**(6), 455-462.
- OPPIKOFER, T., JABOYEDOFF, M., BLIKRA, L. & METZGER, R., 2009: Characterization and monitoring of the Åknes rockslide using terrestrial laser scanning. *Natural Hazards and Earth System Sciences* **9**(3), 1003.
- PEPPA, M. V., MILLS, J. P., MOORE, P., MILLER, P. E. & CHAMBERS, J. E., 2019: Automated co-registration and calibration in SfM photogrammetry for landslide change detection. *Earth Surface Processes and Landforms* **44**(1), 287-303.
- PIERMATTEI, L., CARTURAN, L. & GUARNIERI, A., 2015: Use of terrestrial photogrammetry based on structure-from-motion for mass balance estimation of a small glacier in the Italian alps. *Earth Surface Processes and Landforms* **40**(13), 1791-1802.
- ROBERTI, G., WARD, B., DE VRIES, B. V. W., FRIELE, P., PEROTTI, L., CLAGUE, J. J. & GIARDINO, M., 2018: Precursory slope distress prior to the 2010 Mount Meager landslide, British Columbia. *Landslides* **15**(4), 637-647.
- ROMEO, S., DI MATTEO, L., KIEFFER, D. S., TOSI, G., STOPPINI, A. & RADICIONI, F., 2019: The use of gigapixel photogrammetry for the understanding of landslide processes in alpine terrain. *Geosciences* **9**(2), 99.
- SCHOLZ, H. & SCHOLZ, U., 1981: Das Werden der Allgäuer Landschaft. Eine kleine Erdgeschichte des Allgäus. Verlag für Heimatpflege, Kempten.
- SQUARZONI, C., DELACOURT, C. & ALLEMAND, P., 2005: Differential single-frequency GPS monitoring of the La Valette landslide (French Alps). *Engineering Geology* **79**(3-4), 215-229.
- STUMPF, A., MALET, J. P., ALLEMAND, P., PIERROT-DESEILLIGNY, M. & SKUPINSKI, G., 2015: Ground-based multi-view photogrammetry for the monitoring of landslide deformation and erosion. *Geomorphology* **231**, 130-145.
- WARRICK, J. A., RITCHIE, A. C., ADELMAN, G., ADELMAN, K. & LIMBER, P. W., 2017: New techniques to measure cliff change from historical oblique aerial photographs and structure-from-motion photogrammetry. *Journal of Coastal Research* **33**(1), 39-55.
- WUNDERLICH, T., RAFFL, L. & WIEDEMANN, W., 2019: Wiedererkennung – zwei Lösungen für die strenge Deformationsanalyse flächenhafter Beobachtungen der Ingenieurgeodäsie. In: Hanke K, Weinold T (eds) 20. Internationale Geodätische Woche Obergurgl 2019, “Arbeitskreis Rutschung, Setzung, Deformation”, 264-273.

# Immersive VR-based Live Telepresence for Remote Collaboration and Teleoperation

MICHAEL WEINMANN<sup>1</sup>, PATRICK STOTKO<sup>1</sup>, STEFAN KRUMPEN<sup>1</sup> & REINHARD KLEIN<sup>1</sup>

*Abstract: We present a framework for sharing immersive live telepresence experiences to groups of remote users for arbitrary-sized environments. Our framework builds upon RGB-D data capture of the local environment (by a person or robot) and involves real-time 3D reconstruction, scalable data streaming and visualization to a multitude of remote users at modest bandwidth requirements and low latency while preserving the visual quality of current real-time reconstruction approaches. We evaluate our live-telepresence system regarding its performance, its visual quality and respective user experiences, and show its beneficial use in the context of robot teleoperation, where we achieve a higher situation awareness induced by the VR-based scene exploration in comparison to purely video-based teleoperation.*

## 1 Introduction

Sharing immersive live telepresence experiences has received increasing attention in recent years with applications in entertainment, teleconferencing, remote collaboration, site exploration, education, robotics, cultural heritage and medical rehabilitation. The impression of telepresence - defined as the subjective experience of being in an environment that may differ from the user's actual local physical surrounding - heavily relies on the ability of users to interactively explore the respective scene while avoiding any kind of discomfort such as motion sickness. This requires accurate scene visualization at high framerates and low latencies.

Purely video-based solutions, i.e. in terms of 360 degree videos, strongly restrict the scene exploration to views in the vicinity of the camera poses during scene capture and do not meet the strong real-time constraints required for live scenarios (LUO et al. 2018; SERRANO et al. 2019). Therefore, sharing live-captured scenes as needed for teleconferencing (ORTS-ESCOLANO et al. 2016) or remote collaboration (VASUDEVAN et al. 2011; FAIRCHILD et al. 2016; MOSSEL & KRÖTER 2016; STOTKO et al. 2019a; STOTKO et al. 2019b) typically involves data capture, 3D reconstruction, data streaming and visualization. All these steps have to be achieved within strong real-time constraints while preserving the visual quality of the scene, considering typically available network bandwidth and client-side hardware. Whereas capturing a small fixed-sized region of interest as typical for teleconferencing (ORTS-ESCOLANO et al. 2016) and small-scale remote collaboration (VASUDEVAN et al. 2011; FAIRCHILD et al. 2016), i.e. within a few square meters in single rooms, allows the exploitation of expensive well-calibrated setups with statically placed cameras, the efficient capturing, data representation and streaming become significantly more challenging when capturing scenes of arbitrary size with a moving camera.

---

<sup>1</sup> University of Bonn, Institute of Computer Science II – Visual Computing, Endenicher Allee 19A, D-53115 Bonn, E-Mail: [mw, stotko, krumpen, rk]@cs.uni-bonn.de

This paper addresses the task of sharing immersive live telepresence experiences for arbitrary-sized environments with groups of remote users based on efficient large-scale real-time 3D reconstruction, data streaming and visualization (see Figure 1). By design, the proposed system shifts the hardware requirements from the involved users towards the cloud. In addition, the system runs at low/modest bandwidth requirements with low latency and can handle network interruptions. As a result, this approach allows a re-thinking of well-established exploration processes towards (1) VR-based remote consulting/collaboration, where experts may save the travel time and costs, (2) VR-based remote exploration of contaminated scenes and thereby avoiding the exposure of people to danger, (3) VR-based education scenarios, where expensive excursions may be replaced with immersive group-scale telepresence in the respective environments or (4) VR-based robot teleoperation where remotely-located humans benefit from a higher situation awareness within the respective local environment of the robot. Example applications shown in this paper include remote collaboration as well as robot teleoperation.

## 2 Methodology

As illustrated in Fig. 1, the presented framework for immersive group-scale telepresence in live-captured scenes involves

- a **local user (or robot)** capturing the local environment based on RGB-D sensors as present in mobile phones or the Microsoft Kinect,
- a cloud-based **real-time reconstruction framework** for on-the-fly/online scene capture and camera localization based on volumetric fusion that reconstructs a dense 3D model in real-time and streams both the camera pose and the reconstructed scene parts to the server component,
- a cloud **server** to manage the global scene model and to control the data transmission according to the requests by remotely connected users, which requires reliable, efficient data representation and communication/management under concurrent memory operations while also handling possibly occurring network instabilities (for robot teleoperation, a further management of the coupling between user interactions and respective robot actions has to be taken into account),
- **visualization components** that update the locally generated meshes for the individual remote users according to the already transmitted data and provide the functionalities of interactive scene exploration with features such as measuring distances, marking objects and the collaboration with other connected experts (e.g. via VoIP), and
- remote users that may independently and immersively explore and interact with the live-captured scene while communicating with the person (or robot) capturing the scene.

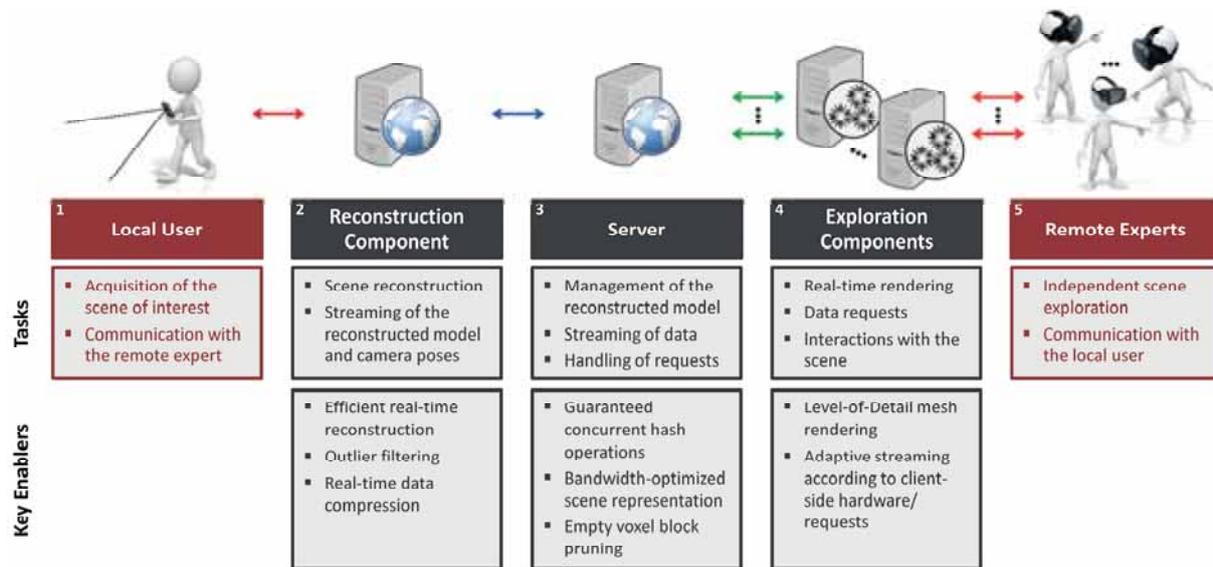


Fig. 1: Overview of the framework, its components and the challenges to be solved for sharing immersive live telepresence experiences for arbitrary-sized environments with groups of remote users. Images are partially provided by PresenterMedia.com

In the following, we describe the major steps towards an efficient live-telepresence system and its extension for robot teleoperation.

## 2.1 Efficient Immersive Live-Telepresence System for Remote Collaboration

The key to an immersive live-telepresence experience is given by efficient data capture, data management and visualization as discussed in detail in our respective investigations (STOTKO et al. 2019a; STOTKO et al. 2019b; STOTKO et al. 2019c). In the following, we briefly summarize respective optimizations separately for the respective components:

- Reconstruction component:** We use an RGB-D-based real-time reconstruction framework (NIEBNER et al. 2013). In a pre-processing step, we filter potentially unreliable data from the RGB-D frames provided by the camera to reduce the data amount and increase the robustness of the RGB-D data used during reconstruction. For this purpose, we discard samples at depth discontinuities and noisy regions within the depth data. Furthermore, we apply a filter to reduce the number of unnecessary allocations during the volumetric data fusion which leads to a significant speed-up of the reconstruction and also reduces the amount of data that are later queued for the streaming to the server. Finally, we also discard voxels that contain only very few and possibly unreliable observations.
- Server:** To efficiently maintain the global 3D scene model and control the streaming to connected remote users' exploration components, we assign each exploration component a stream set based on GPU hash data structures. Here, our data structure is designed to support guaranteed thread-level concurrent insertion, retrieval and removal of entries that are the prerequisite of efficient and reliable data updates. Furthermore, we convert the scene model to a highly bandwidth-efficient representation based on Marching Cubes (MC) indices (LORENSEN et al. 1987) to support pruning empty voxel blocks and enable efficient streaming to a large number of connected exploration components.

- **Visualization component:** To achieve efficient rendering, we group the received scene data into larger mesh blocks for which we asynchronously generate triangle mesh data as well as three levels of detail.

## 2.2 VR-based Immersive Teleoperation and Live Exploration with a Mobile Robot

The aforementioned live-telepresence system can also be used for robot teleoperation as shown by STOTKO et al. (2019d). We equipped a remotely operated ground robot (SCHWARZ et al. 2018) with a Microsoft Kinect v2 RGB-D camera that can be moved via the robot’s arm. As the poses of the robot’s individual components cannot be computed using only the camera pose, the robot client sends them to the live-telepresence framework where the current robot state is directly visualized (see Fig. 2). To allow an adequate immersive teleoperation experience, we additionally remove potential jittering effects due to imperfect camera poses by applying a temporal low-pass filter on the robot’s base pose.

In comparison to other VR-based teleoperation approaches, our technique overcomes the limitations regarding a sparse scene reconstruction in terms of a sparse point cloud (BRUDER et al. 2014) and the high latency induced by video-based VR approaches (KURUP & LIU 2016).

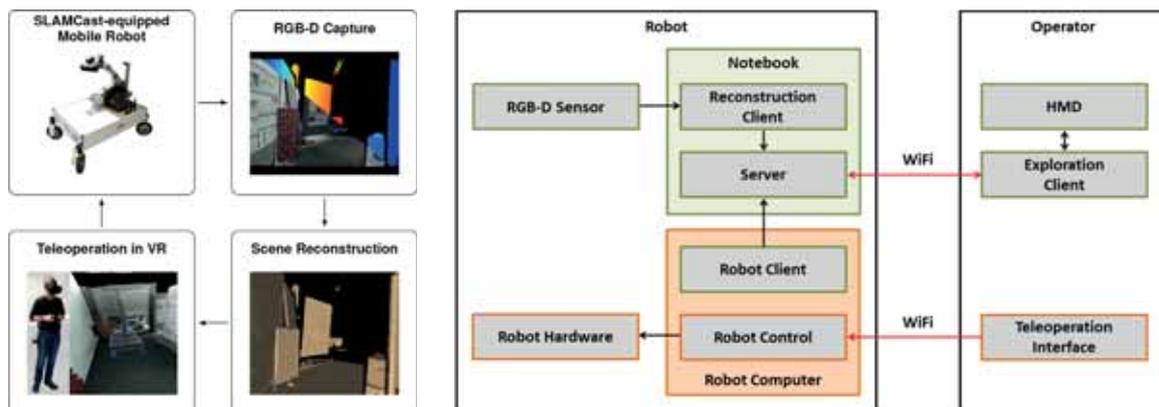


Fig. 2: (left) High-level overview of our VR-based robot teleoperation and scene exploration system: An operator controls a robot using a live-captured and reconstructed 3D model of the robot’s local environment. (right) Detailed implementation of the teleoperation system: The system allows the operator to enter the reconstructed scene without being limited to the specific view of the camera on the robot. The teleoperation of the robot is based on standard teleoperation devices (e.g. a gamepad). Components of the live-telepresence system are marked in green and orange components belong to the robotic system

## 3 Results

In this section, we provide an evaluation of the basic live-telepresence system as well as an evaluation of its use for robot teleoperation.

### 3.1 Evaluation of Live-Telepresence System

In the following, we evaluate our live-telepresence system regarding its performance, its visual quality and a study of respective user experiences.

**Performance Analysis** We measured the bandwidth for streaming the data to the exploration components as well as the streaming latency and component scalability using several datasets captured with off-the-shelf RGB-D cameras. Even with low streaming rates of 512 blocks/request at a request rate of 12Hz, a low latency was observed. Furthermore, the required mean (and maximum) bandwidth was around 13MBit/s (and 25 MBit/s respectively) while the server was able to handle 24 components simultaneously using standard consumer hardware without introducing further latency (see Fig. 3).

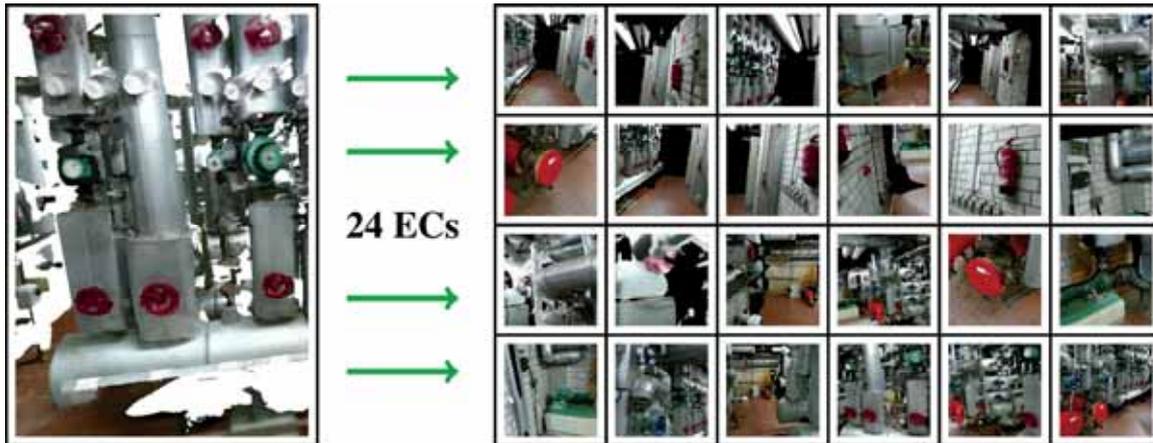


Fig. 3: Illustration of our scalable telepresence system which enables sharing live telepresence experiences of high-quality scene reconstructions to more than 24 exploration components

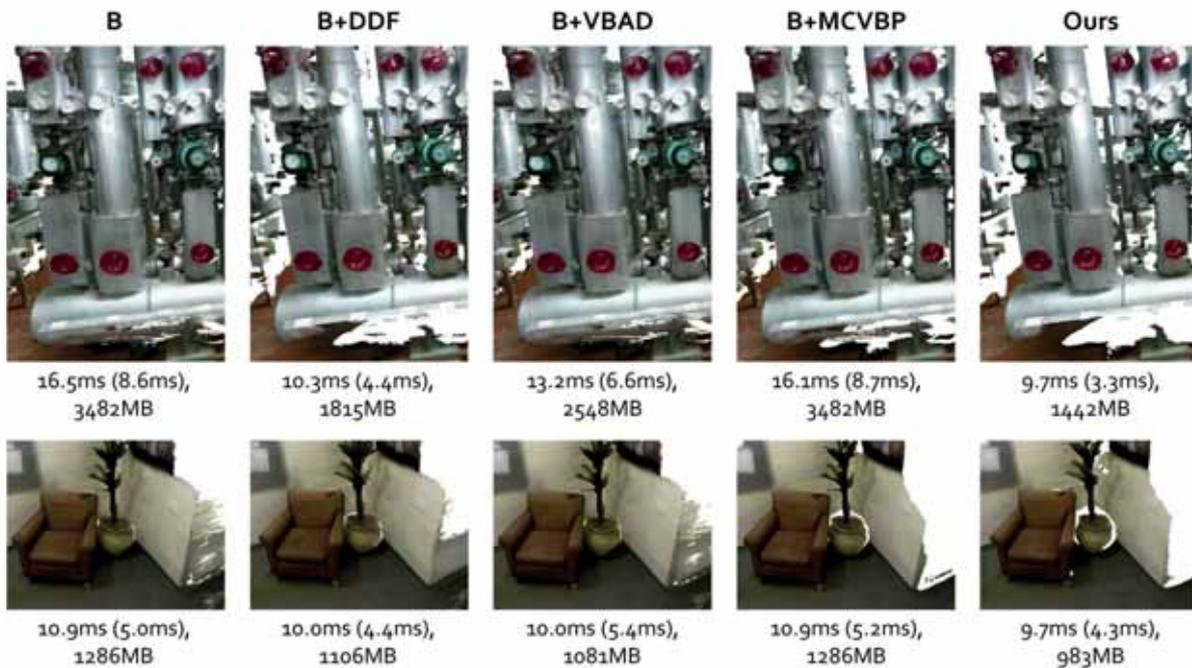


Fig. 4: Comparison of visual quality, mean runtime (and standard deviation) as well as memory requirements for different system variants. For more detailed descriptions, we refer to the discussion by STOTKO et al. (2019b)

**Visual Quality** To ensure a high degree of immersion, we also evaluated the visual quality of the reconstructed 3D models (see Fig. 4). As a result of filtering outliers in the input and model data, our reconstruction component generates smaller models with higher quality in comparison to standard 3D reconstruction which also benefits streaming bandwidth and scalability (see Figure 2). This may be explored for guiding the user to perform a more thorough scene acquisition resulting in more complete 3D models with higher accuracy.

**Evaluation of User Experience** To evaluate the practicality of our framework for telepresence in live-captured scenes, we immersed 18 subjects (mean age of 28.0 years) into an on-the-fly captured scene based on standard VR devices, where the current local scene model was visualized according to the participant’s current pose. This way, the users were able to interactively inspect the scene independent from the camera’s current pose. The user ratings (see Fig. 5) indicate that the users experienced a high degree of situation awareness and self-localization in the simultaneously captured scene and could easily assess the terrain for navigation purposes. Furthermore, they reported the controls for scene interaction (i.e. teleporting, performing distance measurements, etc.) to be intuitive. The ratings regarding the resolution of the reconstructed model and the speed of movements were slightly lower. Future improvements regarding texture resolution and regarding the overall model quality as well as the increasing availability of affordable VR devices, and with it the higher familiarity to the control mechanisms, may address these aspects.

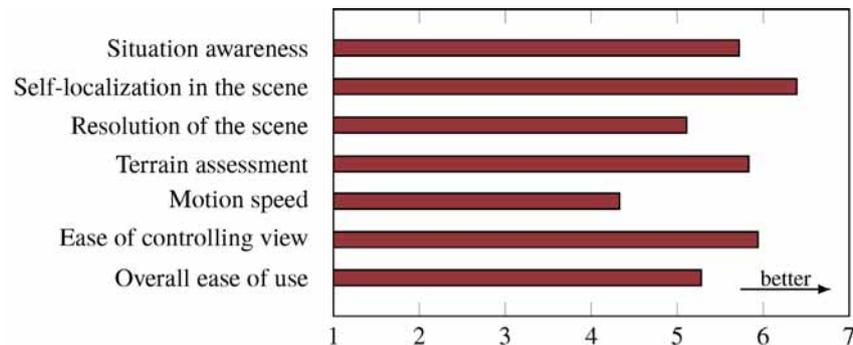


Fig. 5: Assessment of user experience based on mean ratings on a 7-point Likert scale

### 3.2 Evaluation of VR-based Robot Teleoperation

In the following, we evaluate our VR-based robot teleoperation system regarding its usability in comparison to video-based robot teleoperation as well as the quality of remote users’ interactions with the scene.

**Evaluation of User Experience** For the assessment of the benefits of our immersive VR-based teleoperation system in comparison to video-based approaches, we conducted a user study. The participants (20 participants with a mean age of 29.25 years) were asked to maneuver a robot through a course with challenges of different difficulties based on two different stimuli. During the VR mode, the participants were immersed into the local environment of the robot based on standard VR devices and were able to follow the robot through the course. Here, the HMD-based scene inspection corresponding to the already reconstructed scene parts is not limited to the cur-

rent view of the camera. Thus, we expect a higher degree of immersion and situation awareness as users obtain a better impression regarding 3D distances in the scene as well as occurring obstacles. During the video mode, the robot teleoperation was purely based on video data corresponding to the current view of the camera moved by the robot. As a consequence, there is no possibility of observing information outside the current camera view which we expect to result in a lower degree of situation awareness due to the reduced perception quality of distances between objects in the scene as well as occurring obstacles. To validate our expectations, we collected ratings of a variety of relevant aspects based on a 7-point Likert (see Figure 6), the number of collisions with the environment made in the different modes and the total execution time required to navigate the robot from the starting point to the target location. The ratings provided by the users indicate that our VR system is beneficial for self-localization in the scene, maneuvering around narrow corners, avoiding obstacles, assessing terrain for navigability as well as regarding the ease of controlling the view. For these aspects, the boxes defined by the medians and interquartile ranges do not overlap indicating a significant difference in favor of the VR-based teleoperation. In addition, the VR mode was judged to be well-suited for teleoperation and to allow moving the robot to target positions in an easier manner. These statistical results support the higher degree of situation awareness for the VR-based teleoperation.

However, it seems as if the higher degree of immersion for VR-based teleoperation results in a more time-consuming inspection that limits the speed of robot motion. The video-based mode allowed a faster completion of the course due to the lacking possibility to inspect the situation, e.g. by walking around the robot, at the cost of a higher number of collisions. Furthermore, the video-based mode was rated slightly better regarding the perceived latency as well as the resolution. The latter aspect may be addressed by future work on enhancing the texture resolution.

**Interaction of Remote Users with the Scene** In the context of disaster management, typical user interactions with the scene for exploring contaminated sites and establishing evacuation plans include the measurement of distances such as door widths to determine whether a different robot or the required equipment would fit through narrow spaces. For this reason, we implemented a tool for measuring 3D distances based on the controllers of the involved HMD device, where the resulting measurement accuracy is determined by the voxel resolution specified for the scene representation (see Fig. 6). We observed errors of up to 1 cm which we expect to be sufficient for disaster management scenarios. Furthermore, our system allows the user to label areas as interesting, suspicious or incomplete. This information is integrated into the overall map and may be used by the capturing robot to return to a particular position as well as to complete or to refine the scan.

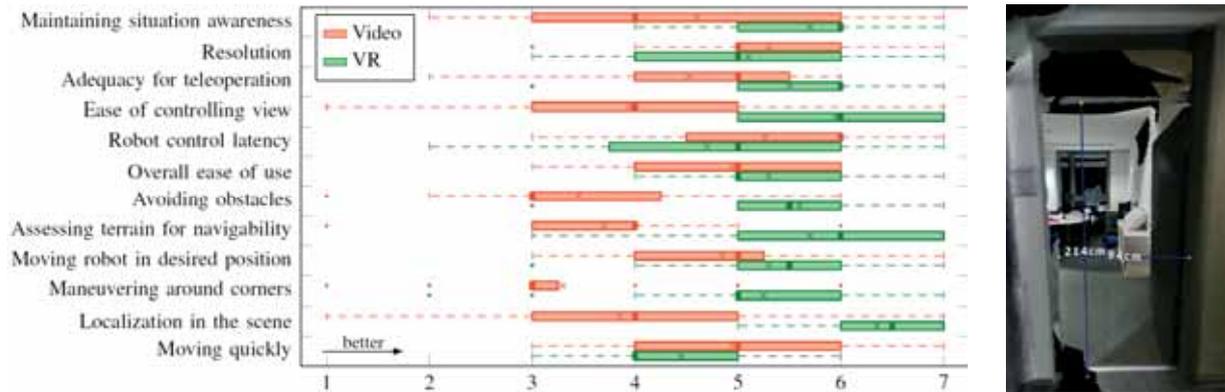


Fig. 6: (left) Statistical results of a user study, i.e. median, lower and upper quartile (includes interquartile range), lower and upper fence, outliers (marked with a dot) as well as the average value (marked with an x), for different aspects. Our VR-based system achieved higher ratings on the 7-point Likert scale than a video-based approach for most of the aspects. (right) Example of interactively taken measurements (94 cm and 214 cm) of the width and height of a door taken to guide the further management process. The ground truth values are 95 cm and 215 cm

## 4 Conclusions

We summarized our investigations towards scalable sharing of immersive live-telepresence experiences beyond room-scale based on efficient real-time 3D reconstruction and streaming. For this purpose, we (1) integrated several filters to discard unreliable data and thereby improve model compactness and robustness, (2) developed a bandwidth-optimized scene representation for the global scene model on the server, (3) introduced a novel hashing approach to allow concurrent thread-level insertion, removal and retrieval while preserving key uniqueness, and (4) use level-of-detail rendering of the mesh data for the exploration clients. While this system is suitable for remote collaboration purposes, we also described its extension towards VR-based robot teleoperation where we showed the benefits over purely 2D video-based teleoperation. Future challenges include the further improvement regarding model quality, streaming efficiency and collaborative capturing.

## 5 Acknowledgements

We thank Max Schwarz, Christian Lenz and Sven Behnke for their support in integrating our telepresence system on a robot platform for teleoperation.

## 6 References

- BRUDER, G., STEINICKE, F. & NÜCHTER, A., 2014: Poster: Immersive point cloud virtual environments. IEEE Symposium on 3D User Interfaces, 161-162.
- KURUP, P. & LIU, K., 2016: Telepresence Robot with Autonomous Navigation and Virtual Reality: Demo Abstract, ACM Conference on Embedded Network Sensor Systems, 316-317.

- FAIRCHILD, A.J., CHAMPION, S.P., GARCÍA, A.S., WOLFF, R., FERNANDO, T. & ROBERTS, D.J., 2016: A Mixed Reality Telepresence System for Collaborative Space Operation. *IEEE Transactions on Circuits and Systems for Video Technology* **27**(4), 814-827.
- LORENSEN, W.E. & CLINE, H.E., 1987: Marching Cubes: A High Resolution 3D Surface Construction Algorithm. 14th Annual Conference on Computer Graphics and Interactive Techniques, 163-169.
- LUO, B., XU, F., RICHARDT, C. & YONG, J., 2018: Parallax360: Stereoscopic 360 Scene Representation for Head-Motion Parallax. *IEEE Trans. on Visualization and Computer Graphics* **24**(4), 1545-1553.
- MOSSEL, A. & KRÖTER, M., 2016: Streaming and Exploration of Dynamically Changing Dense 3D Reconstructions in Immersive Virtual Reality. *IEEE International Symposium on Mixed and Augmented Reality*, 43-48.
- NIEBNER, M., ZOLLHÖFER, M., IZADI, S. & STAMMINGER, M., 2013: Real-time 3D Reconstruction at Scale Using Voxel Hashing. *ACM Transactions on Graphics* **32**(6), 169:1-169:11.
- ORTS-ESCOLANO, S., RHEMANN, C., FANELLO, S., CHANG, W., KOWDLE, A., DEGTYAREV, Y., KIM, D., DAVIDSON, P., KHAMIS, S., DOU, M., TANKOVICH, V., LOOP, C., CAI, Q., CHOU, P., MENNICKEN, S., VALENTIN, J. PRADEEP, V., WANG, S., KANG, S.B., KOHLI, P., LUTCHYN, Y., KESKIN, C. & IZADI, S., 2016: Holoportation: Virtual 3D Teleportation in Real-time. *Annual Symposium on User Interface Software and Technology*, 741-754.
- SCHWARZ, M., DROESCHEL, D., LENZ, C., PERIYASAMY, A.S., PUANG, E.Y., RAZLAW, J., RODRIGUEZ, D., SCHÜLLER, S., SCHREIBER, M. & BEHNKE, S., 2018: Team NimbRo at MBZIRC 2017: Autonomous valve stem turning using a wrench. *Journal of Field Robotics* **36**(1), 170-182.
- SERRANO, A., KIM, I., CHEN, Z., DIVERDI, S., GUTIERREZ, D., HERTZMANN, A. & MASIA, B., 2019: Motion parallax for 360 RGBD video. *IEEE Transactions on Visualization and Computer Graphics* **25**(5), 1817-1827.
- STOTKO, P., KRUMPEN, S., HULLIN, M.B., WEINMANN, M. & KLEIN, R., 2019a: SLAMCast: Large-Scale, Real-Time 3D Reconstruction and Streaming for Immersive Multi-Client Live Telepresence. *IEEE Transactions on Visualization and Computer Graphics* **25**(5), 2102-2112.
- STOTKO, P., KRUMPEN, S., WEINMANN, M. & KLEIN, R., 2019b: Efficient 3D Reconstruction and Streaming for Group-Scale Multi-Client Live Telepresence. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality*, 19-25.
- STOTKO, P., KRUMPEN, S., KLEIN, R., & WEINMANN, M., 2019c: Towards Scalable Sharing of Immersive Live Telepresence Experiences Beyond Room-scale based on Efficient Real-time 3D Reconstruction and Streaming. *CVPR Workshop on Computer Vision for Augmented and Virtual Reality*.
- STOTKO, P., KRUMPEN, S. SCHWARZ, M., LENZ, C., BEHNKE, S., KLEIN, R., & WEINMANN, M., 2019d: A VR System for Immersive Teleoperation and Live Exploration with a Mobile Robot. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, arXiv preprint arXiv:1908.02949.
- VASUDEVAN, R., KURILLO, G., LOBATON, E., BERNARDIN, T., KREYLOS, O., BAJCSY, R. & NAHRSTEDT, K., 2011: High-Quality Visualization for Geographically Distributed 3-D Teleimmersive Applications. *IEEE Transactions on Multimedia* **13**(3), 573-584.

# Gelingendes Zusammenspiel von Einzelmaßnahmen zur Weiterentwicklung des Lehr-Lernsystems im BSc-Studiengang „Geodäsie und Geoinformatik“ am Karlsruher Institut für Technologie (KIT)

MICHAEL MAYER<sup>1</sup>, KATRIN HESS<sup>2</sup> & JAN CERMAK<sup>3</sup>

*Zusammenfassung: Die Weiterentwicklung des Lehr-Lernsystems von Studiengängen stellt den organisationalen Transformationsprozess dar, mit dem Lehreinheiten auf veränderte Umgebungen (z.B. bildungspolitische Vorgaben, hochschulinterne Leitlinien, Ergebnisse hochschuldidaktischer Forschung) reagieren, um für die Fachdisziplin und den jeweiligen Standort Zukunftssicherheit gewährleisten zu können. Der Beitrag stellt repräsentativ ausgewählte, neu etablierte Einzelmaßnahmen vor, die auf Fakultäts- und Studieneingangsebene wirken, von unterschiedlichen Hierarchieebenen initiiert wurden und hohes Transferpotenzial aufweisen. Neben der Beschreibung der Einzelmaßnahmen werden Bedingungen für die gelingende Ausgestaltung des Veränderungsprozesses abgeleitet.*

## 1 Einleitung

Bildungspolitische Vorgaben (z.B. Bologna-Prozess, Kompetenzorientierung; STOLZ & SCHOLKMANN 2015), Hochschulleitungen (z.B. Leitbilder, Systemakkreditierung) und Fakultäten (z.B. Evaluationsergebnisse, administrative Prozesse) sowie Lehreinheiten (z.B. Gremien, Lehrende, Studierende) selbst (z.B. Curriculumentwicklung; NIETHAMMER et al. 2014) sind für die Weiterentwicklung von Studiengängen Treiber und Motivatoren. Durch die Veränderung bestehender und die Entwicklung neuer Studienangebote reagieren Lehreinheiten in kontinuierlichen Veränderungsprozessen auf dynamische und zunehmend globalisierte Anforderungen von Gesellschaft, Wissenschaft und Forschung sowie Berufswelt, die wiederum Megatrends (z.B. Digitalisierung, Individualisierung, Konnektivität; HORX 2011) rückkoppeln und dadurch Zukunftssicherheit gewährleisten. Diesen fachspezifisch individuellen Transformationsprozess gestalten einzelne organisationale Lehreinheiten unter besonderer Berücksichtigung von Historie und Identität sowie Rahmenbedingungen (z.B. personale, finanzielle, räumliche Ressourcen, Studierendenzahl); sie fokussieren dabei auf unterschiedliche Personengruppen (z.B. Lehrende, Studierende) sowie – bzgl. Curriculum, Fach, Modul, Lehrveranstaltung/-arrangement – auf inhaltliche und didaktische Aspekte.

Am Beispiel des deutschsprachigen Studiengangs „BSc Geodäsie und Geoinformatik“ (ILLNER et al. 2018; MAYER et al. 2019) am Karlsruher Institut für Technologie (KIT) wird im Folgenden

---

<sup>1</sup> Karlsruher Institut für Technologie (KIT), KIT-Fakultät Bauingenieur-, Geo- und Umweltwissenschaften, Kaiserstraße 12, D-76131 Karlsruhe, E-Mail: michael.mayer@kit.edu

<sup>2</sup> Karlsruher Institut für Technologie (KIT), Personalentwicklung und Berufliche Ausbildung (PEBA), Vincenz-Prießnitz-Str. 1, D-76131 Karlsruhe, E-Mail: katrin.hess@kit.edu

<sup>3</sup> Karlsruher Institut für Technologie (KIT), Lehreinheit Geodäsie und Geoinformatik, Englerstr. 7, D-76131 Karlsruhe, E-Mail: jan.cermak@kit.edu

aufgezeigt, wie durch das gelingende Zusammenspiel von Einzelmaßnahmen zur ganzheitlichen – den MSc-Studiengang „Geodäsie und Geoinformatik“ einschließenden – Weiterentwicklung des Lehr-Lernsystems innovativ, nachhaltig und profilbildend beigetragen werden kann (Kap. 3). Hierzu wird vorher in Kap. 2 auf weitere rahmengebende Veränderungsprozesse sowie Ressourcen eingegangen. Der Beitrag schließt in Kap. 4 mit einer wertenden Zusammenfassung und gibt einen Ausblick auf geplante Weiterentwicklungen des Lehr-Lernsystems.

## 2 Rahmgebung und Ressourcen

Neben den o.g. bildungspolitischen Vorgaben, die das gesamte deutsche Hochschulsystem betreffen, wirkt auf den Veränderungsprozess des Lehr-Lernsystems am KIT, der Forschungsuniversität in der Helmholtz-Gemeinschaft, insbesondere auch die Fusion zwischen einer Universität in Landeshoheit und einer Großforschungseinrichtung der Helmholtz-Gemeinschaft mit programmorientierter Vorsorgeforschung im Auftrag des Staates ein (KIT 2015). Das Ziel der Lehre am KIT besteht in der Qualifikation von Studierenden auf Basis (i) einer intensiven wissenschaftlichen und forschungsorientierten Ausbildung und (ii) des überfachlichen Kompetenzerwerbs. Als zentrale didaktische Leitidee wird zur Sicherstellung einer engen Verbindung von Forschung und Lehre – von Beginn des KIT-Studiums an und in allen Lehrveranstaltungen – das didaktische Konzept der forschungsorientierten Lehre verwendet, die an den disziplinspezifischen Standards ausgerichtet ist (BEYERLIN et al. 2014). Forschungsorientierte Lehre umfasst dabei Lehr-Lernkonzepte des forschenden Lernens (HUBER 2009), bei dem Studierende selbstständig forschend lernen, sowie Lehr-Lernkonzepte, die an aktuelle Forschungsergebnisse und -themen anknüpfen, wobei der Lehr-Lernprozess jedoch stärker lehrendenzentriert ausgestaltet ist (KIT 2015).

Darüber hinaus bewegen sich Lehrende an der „speziellen“ Organisation (KEHM 2012) Universität stetig im ressourcenbezogenen Spannungsfeld zwischen Forschung und Lehre, in dem bspw. in Berufungsverfahren Lehrleistungen und -qualifikationen im Vergleich zu Forschungsexpertise eine untergeordnete Rolle einnehmen (WISSENSCHAFTSRAT 2005), woraus ein forschungsbezogener Tätigkeitsfokus resultiert (ZQ & PROGNOSE AG 2016; HILBRICH & SCHUSTER 2014). Um einen Kulturwandel des Stellenwerts der Lehre zu initiieren, wurden unterschiedliche Förderprogramme bspw. der [Qualitätspakt Lehre](#) (Bundesministerium für Bildung und Forschung) ausgeschrieben. Auf der Ebene des Landes Baden-Württemberg fokussieren weitere Ausschreibungen auf die Verbesserungen der Studienbedingungen und der Lehrqualität.

In der Lehrereinheit Geodäsie und Geoinformatik (GuG) wurden die in Kapitel 3 beschriebenen Weiterentwicklungen des Lehr-Lernsystems insbesondere ermöglicht durch synergetische Kollaborationen zwischen den Drittmittel-Projekten

- Integratives Konzept zur Studiengangentwicklung mit Unterstützung von hochschuldidaktischen Fachtandems (HD-Fachtandem): Qualitätspakt Lehre, KIT-Lehre<sup>Forschung</sup>-PLUS (BMBF, FKZ: 01PL12004); Hauptziel: Erhöhung von forschungsorientierten Lehr-Lernanteilen;
- Lernen<sup>Coaching</sup>: Qualitätspakt Lehre, KIT-Lehre<sup>Forschung</sup>-PLUS (BMBF, FKZ: 01PL12004); Hauptziel: Fachkulturspezifische, individualisierte Ausgestaltung der Studieneingangsphase;

- Projekt MoWi-KIT: Fonds „Erfolgreich Studieren in Baden-Württemberg“ (FEST-BW), Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg (MWK) Baden-Württemberg, Programm „Wissenschaft lernen und lehren“; Hauptziel: Etablierung von modul- und semesterübergreifenden Lehr-Lernarrangements;
- Projekt GuG-Studienlotsen: Programm „Strukturmodelle in der Studieneingangsphase“ und „Studienstart – Studienlotsen, Mentoring und entschleunigte Studienpläne“, Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg (MWK); Hauptziel: Near Peer Mentoring zur verbesserten Ausgestaltung der Studieneingangsphase.

### 3 Weiterentwicklung des Lehr-Lernsystems

Während die Projekte Lernen<sup>Coaching</sup>, MoWi-KIT und GuG-Studienlotsen primär auf die Weiterentwicklung des GuG-Lehr-Lernsystems abzielen (siehe Kap. 3.2), ermöglicht das an der KIT-Fakultät für Bauingenieur-, Geo- und Umweltwissenschaften (BGU) angesiedelte Projekt HD-Fachtandem (HEß & KLINK 2018; MAYER & HEß 2018) einen wertvollen außenperspektivischen Blick auf das GuG-Lehr-Lernsystem. Im interdisziplinären HD-Fachtandem kooperieren eine Person aus der Fachdisziplin und eine Person aus der hochschuldidaktischen Abteilung des KIT. Das HD-Fachtandem verfügt somit über hochschuldidaktische Expertise, Außenperspektive und gewinnbringendes Systemwissen, bietet Chancen zur Perspektivenerweiterung und kennt die kulturellen disziplinspezifischen Besonderheiten (FLEISCHMANN et al. 2017), da ein akademischer Mitarbeiter aus GuG Teil des Tandems ist. Gemeinschaftlich können somit im Veränderungsprozess umfangreichere Chancen für positive Wirkungen auf das GuG-Lehr-Lernsystem mit erhöhter Akzeptanz kreiert werden.

In diesem Kontext entwickelt das HD-Fachtandem – teilweise gemeinsam mit GuG-Adressat\*innen – für diverse Zielgruppen (z.B. einzelne Lehrende, Gruppen von Studierenden und/oder Mitarbeitenden, Lehrstühle, Lehreinheit) bedarfsgerechte und passgenaue Angebote, die – gegenüber KIT-weiten Angeboten – erhöhte Akzeptanz und gesteigerte Selbstverpflichtung zur Partizipation bzw. Umsetzung aufweisen. Um eng an die Geschehnisse in der Lehreinheit angekoppelt sein zu können, ist das HD-Tandem Gast in der GuG-Studienkommission. Kap. 3.1 gibt durch ausgewählte Maßnahmen Einblick in das Spektrum neu entwickelter hochschuldidaktischer Angebote und Instrumente des HD-Fachtandems der KIT-Fakultät BGU.

#### 3.1 Chancen durch Angebote des HD-Fachtandems

In diesem Kapitel werden ausgewählte Maßnahmen beschrieben, die vom HD-Fachtandem der KIT-Fakultät BGU entwickelt und umgesetzt wurden und durch die das GuG-Lehr-Lernsystem kontinuierlich weiterentwickelt wird.

##### 3.1.1 Brown Bag Lectures

Seit 2019 werden alle zwei Monate in der KIT-Fakultät BGU im Rahmen der „Brown Bag Lectures“ Kurz-Inputs zu wechselnden Themen (z.B. Best Practice, Kommunikation, Methodenkoffer) durch das HD-Tandem oder eingeladene Referent\*innen gegeben. Das niederschwellige Setting ist dabei vom klassischen Workshop-Setting – aufgrund der kurzen Zeitspanne (13:10-13:50) und der Möglichkeit einen kleinen Mittagsimbiss während des Inputs einzunehmen – abweichend.

GuG-Lehrenden bietet diese Maßnahme die Möglichkeit sich weiterzubilden, sich mit anderen BGU-Lehrenden zu vernetzen oder als Referent\*in aktiv mitzugestalten; so informierten bspw. in diesem Rahmen zwei GuG-Lehrende über ihren didaktischen Ansatz, um Forschungsorientierung schon zu Beginn des Studiums sicherzustellen.

### 3.1.2 Workshopangebote und Zertifikatsprogramm Advanced Teaching Competencies

Das HD-Fachtandem bietet – ergänzend zu den an einzelne Mitarbeitende gerichteten Angeboten des [Hochschuldidaktikzentrum Baden-Württemberg](#) – zur Erhöhung der Lehrqualität auf Fakultätsebene u.a. für GuG-Mitarbeitende Workshopformate (unterschiedlicher zeitlicher Umfang: 1-8 Stunden) an; zu Beginn jedes Semesters werden hierbei bspw. halbtägige hochschuldidaktische Crashkurse durchgeführt, die an Neueinsteiger\*innen in der Lehre adressiert sind. Diese Kurse werden regelmäßig von GuG-Lehrenden wahrgenommen und dienen auch der individuellen Vernetzung innerhalb der BGU-Fakultät mit anderen Neueinsteiger\*innen.

Darüber hinaus kommt zeitlich umfangreichen (Dauer: ca. 1 Jahr), (lehr-)persönlichkeitsbildenden Maßnahmen eine bedeutsame Rolle bei der Weiterentwicklung von Lehre zu. Bspw. bildeten sich drei GuG-Mitarbeitende in 2019 durch das vom BGU-HD-Fachtandem in Kooperation mit der Fakultät für Maschinenbau neu entwickelte [Zertifikatsprogramms Advanced Teaching Competencies](#) (ATC) weiter, das Nachwuchswissenschaftler\*innen durch intensiven Diskurs mit der eigenen Lehre hochschuldidaktisch weiterqualifiziert und insbesondere lehrebezogen auf Berufungsverfahren vorbereitet. Gleichzeitig koppelt diese Maßnahme durch die weiterqualifizierten Mitarbeitenden in die Lehreinheit rück.

### 3.1.3 Prozessbegleitung Reakkreditierung

Als Reaktion auf die in Kap. 1 angeführten Veränderungen der Umwelten von Studiengängen und zur Sicherstellung einer qualitativ hochwertigen universitären Ausbildung durchlaufen Studiengänge regelmäßig Akkreditierungsverfahren. Im Rahmen der [KIT-Systemakkreditierung](#) begleitete das HD-Fachtandem im Wintersemester 2018/19 den Reakkreditierungsprozess der beiden deutschsprachigen GuG-Studiengänge. Dadurch konnten bedeutsame außenperspektivische Sichten auf das GuG-Lehr-Lernsystem berücksichtigt werden. Gleichzeitig wurde der Prozess so ausgestaltet, dass für die GuG-Studiengänge der Status der forschungsorientierten Lehre (vgl. Kap. 2) erfasst wurde, wodurch sich nahezu alle GuG-Lehrenden und interessierte Studierende mit diesem für die KIT-Lehre wichtigen Konzept auseinandersetzen.

## 3.2 Fokus auf GuG-interne Innovationen

In die in Kap. 3.1 beschriebenen Maßnahmen, die von übergeordneten Hierarchieebenen – teilweise gemeinsam mit GuG – initiiert wurden, fügen sich die in diesem Kapitel beschriebenen, in der Lehreinheit GuG entwickelten Maßnahmen ein.

### 3.2.1 Erprobung und Etablierung innovativer Lehr-Lernsettings

Eine hochbedeutsame Grundvoraussetzung für die Weiterentwicklung des GuG-Lehr-Lernsystems auf Lehrveranstaltungsebene – in unterschiedlichen Umfängen (z.B. einzelne Lehr-Lerneinheiten, Ausgestaltung der Selbstlernphase) und angeregt durch unterschiedliche Anlässe (z.B. Weiterqualifikation, Evaluationsergebnis, Fachartikel, informelles Gespräch) – sind intrinsisch

motivierter Lehrende, die neue Lehr-Lernsettings erproben oder für GuG adaptieren. Stellvertretend und repräsentativ ausgewählt können hier die Methoden „Leitfragen“ und „Gruppeninterview“ angeführt werden.

Durch das Instrument „Leitfragen“ (KING 1990) konnte ein niederschwelliges und multifunktionales Tool zur Nach- bzw. Vorbereitung von Präsenzlehre und zum Kompetenztraining in verschiedenen Lehrveranstaltungen etabliert werden. Dabei werden entweder vom Lehrenden oder von Studierenden auf unterschiedliche Taxonomie-Ebenen (z.B. Faktenwissen, Transfer, Reflexion; KRATHWOHL 2002) fokussierend im Nachgang von Präsenzlehre Fragen formuliert. Diese werden zur Sicherung der Qualität einem Peer-Review-Prozess unterzogen, bevor diese Fragen webbasiert den Teilnehmenden der Lehrveranstaltung zur Verfügung gestellt werden. Sie können damit orts- und zeitunabhängig im Selbststudium mit ausreichend Zeit zum individuellen Diskurs genutzt werden. Zu Beginn der folgenden Präsenzlehreinheit werden die Fragen genutzt, um an die Themen der vorhergehenden Einheit anzuschließen. Dabei tauschen sich die Studierenden zunächst in Kleingruppen aus, um anschließend im Plenum zu diskutieren. Leitfragen werden von den Studierenden als motivierend wahrgenommen und erzeugen gleichzeitig wichtiges Feedback für Lehrende. Zudem können die Fragen zur Prüfungsvorbereitung genutzt werden, wodurch ebenso Constructive Alignment (BIGGS & TANG 2011) unterstützt wird.

Durch Gruppeninterviews (KLINK et al. 2016) wurde ein alternatives Tool zur Lehrveranstaltungsevaluation erprobt und etabliert. Hierbei wird in Abstimmung mit den Verantwortlichen einer Lehrveranstaltung ein Fragenkatalog entwickelt, der die Basis eines Interviews darstellt, das von Mitarbeitenden der KIT-Personalentwicklung in studentischen Kleingruppen geführt, analysiert und anonymisiert in die Lehreinheit rückgespielt wird. Ergänzend zur schriftlichen standardisierten KIT-Lehrveranstaltungsevaluation bietet dieses Instrument höhere Flexibilität und ein engeres Anknüpfen an die Informationsbedarfe der Lehrenden.

### 3.2.2 Modul- und semesterübergreifende Lehr-Lernelemente

Durch die Ausgestaltung des BSc-Studiums (z.B. Modulprüfungen umfassen Lehrveranstaltungen aus maximal konsekutiven Semestern) benötigen Studierende zunehmend Unterstützungsstrukturen, die Zusammenhänge modul- bzw. semesterübergreifend kontextuell einbetten und individuell hochwertiges Lernen (BERENDT 2006) sicherstellen. In der Lehreinheit GuG wurden in enger Kooperation mit dem House of Competence (HoC), der zentralen, forschungsbasierten KIT-Einrichtung zur fachübergreifenden studentischen Kompetenzentwicklung Lehr-Lernarrangements zum integrativen Training von personalen Kompetenzen durch sog. Querschnittsthemen (HEUBLEIN et al. 2018a; 2018b) geschaffen. Aktuell wird hierbei von Beginn des BSc-Studiums an die mündliche Wissenschaftskommunikation (z.B. Präsentationstechnik) sowie das wissenschaftliche Schreiben durch regelmäßige individuelle Feedbackschleifen unter Verwendung von standardisierten Feedbackinstrumenten kontinuierlich weiterentwickelt.

Wird der Fokus auf die Lehrenden gelegt, so stellen modulübergreifende Formate wichtige Alternativperspektiven sicher. In diesem Kontext wurde im Wintersemester 2019/20 erstmalig ein Austausch aller GuG-Dozent\*innen des ersten BSc-Semesters mit dem Hauptziel durchgeführt, Wahrnehmungen und Beobachtungen abzugleichen und basierend darauf zu prüfen, ob gemeinsam Handlungsparadigmen zu schärfen sind.

### 3.2.3 Ausgestaltung der Studieneingangsphase

Die Studieneingangsphase des BSc-Studiengangs GuG ist geprägt durch importierte umfangreiche Grundlagenlehrveranstaltungen in MINT-Fächern (z.B. Höhere Mathematik, Physik, EDV und Informatik; Umfang: 60 von 180 Gesamt-ECTS). Gleichzeitig ermöglichen ausgewählte Lehrveranstaltungen das fachliche Ankommen in der Fachdisziplin. In GuG wird im Rahmen der Lehrveranstaltung „Fit für Studium und Beruf“ in den ersten beiden Semestern (3 ECTS; HEUBLEIN et al. 2018c) Studierenden – insbesondere durch berufsbildschärfende Maßnahmen (z.B. MAYER et al. 2020), Lehr-Lerneinheiten zur Aneignung von disziplinspezifischen Kompetenzen (z.B. Datenanalyse) und das Anbieten von Lehr-Lernräumen zum moderierten reflexiven Diskurs – der selbstkompetente Übergang ins KIT motivierend und an Begabungen und Vorwissen anknüpfend diversitätsgerecht ermöglicht (RABOLD et al. 2019a).

Dies ist insbesondere im Kontext der Ziele des Erkenntnisprozesses des forschungsorientierten und Bologna-konformen Lehrens bedeutsam; denn es gilt zu beachten, dass Studierenden- und Kompetenzorientierung – z.B. im Hinblick auf Forschungs- und Reflexionskompetenz im Referenzrahmen der Fachdisziplin (LÜBECK 2010) – nicht voraussetzungslos selbstständig erfolgen können. JENERT (2008) belegt dies am Beispiel der individuellen studentischen Lernkompetenz, die durch explizite Anlässe der lernkompetenzbezogenen Theorieaneignung und des Trainings weiterzuentwickeln ist.

### 3.2.4 Studierende als mündige Ausgestalter

Studierende agieren in GuG als mündige Ausgestalter (REINMANN & JENERT 2011) der Weiterentwicklung des Lehr-Lernsystems (RABOLD et al. 2019b). Sie verändern das Lehr-Lernsystem z.B. durch Impulse (z.B. „Leitbild gute Lehre“ von Studierenden für Lehrende), sind involviert in die Planung künftiger Lehrveranstaltungen (vgl. Lehrveranstaltung „Fit für Studium und Beruf“, Kap. 3.2.4) und gestalten als Near Peer Tutor\*innen (WHITMAN & FIFE 1988) Lehre aktiv aus. Damit diese verantwortungsvollen Aufgaben bestmöglich ausgeführt werden können, unterstützt die Lehreinheit GuG die individuelle Weiterqualifikation von Studierenden(-gruppen) und die Prozessausgestaltung bestmöglich (z.B. finanzielle Ressourcen, externe Moderator\*innen auf Klausurtagungen).

Hierdurch können in GuG bspw. im Projekt [GuG-Studienlotsen](#) während der Studieneingangsphase Near Peer Mentor\*innen aus unterschiedlichen Semestern eigenverantwortlich ihre Kommiliton\*innen der ersten beiden BSc-Semester unterstützen und dadurch auch die Mitarbeitenden entlasten. Erfolgreich etablierte Maßnahmen sind dabei wöchentliche Sprechstunden, BSc-weite Umfragen zur Studiensituation (z.B. Arbeitslast, Motivation), die anonymisiert über die Studienkommission eine wichtige Grundlage für die Weiterentwicklung des GuG-Lehr-Lernsystems darstellen und, „Lemonade Lectures“, die während der Mittagspause semesterspezifisch über wichtige Aspekte des Studiums informieren.

Die Bedeutung, die dem Beitrag von Studierenden im Kontext der Weiterentwicklung des Lehr-Lernsystems in GuG zukommt, wird auch in der gelebten Praxis sichtbar, dass Formate etabliert wurden, bei denen sich Lehrende und Studierende gemeinsam fortbilden (z.B. [Workshops im Rahmen von Lehrexport für die Fakultät Architektur](#)).

## 4 Zusammenfassung und Ausblick

Im Rahmen des Beitrags wurde ausgeführt, wie im deutschsprachigen BSc-Studiengang der KIT-Lehreinheit GuG das Lehr-Lernsystem weiterentwickelt wird. Hierzu wurden ausgewählte, repräsentative Maßnahmen beschrieben, die hohes Transferpotenzial für andere Studiengänge aufweisen und durch Tab. 1 klassifiziert werden können. Dabei wird jede Maßnahme hinsichtlich Initiatoren, Zielgruppe und -gruppengröße, notwendigem Aufwand für die Teilnehmenden, ausgewählten Zeitabschnitten im Verlauf des Studiums, der Möglichkeit sich zu vernetzen und der Außenperspektive, die in das GuG-Lehr-Lernsystem einwirkt, unterschieden. Gleichzeitig wurde im Beitrag aufgezeigt, wie die Einzelmaßnahmen synergetisch zusammenspielen und welche Rolle dabei dem HD-Fachtandem der KIT-Fakultät BGU zukommt.

Tab. 1: Maßnahmen zur Weiterentwicklung des GuG-Lehr-Lernsystems

Maßnahme	Initiatoren	Zielgruppe,- ngröße	Aufwand	Bereich im Studium	Chance zur Vernetzung	Außenper- spektive
Brown Bag Lectures	HD-Fachtandem	Einzelne Lehrende	gering	alle	mittel	hoch
ATC	HD-Fachtandem	Einzelne Lehrende	hoch	alle	hoch	hoch
Reakkreditierung	Hochschulleitung	gesamte Lehreinheit	mittel	alle	gering	mittel
Workshop	z.B. HD-Fachtandem	Einzelne Lehrende	gering	alle	mittel	mittel
Lehr-Lern-innovation	Lehrende	Lehrveranstaltungen	mittel	alle	gering	gering
Querschnittsthemen	Projekt MoWi-KIT	alle BSc-Studierende	mittel	alle	gering	gering
Fit für Studium & Beruf	Projekt Lernen <sup>Coaching</sup>	1.-2. BSc-Semester	hoch	Studieneingangsphase	gering	gering
Gestaltende Studierende	Studierende	gesamte Lehreinheit	hoch	alle	gering	gering

Bezüglich der in den ersten beiden Kapiteln angeführten Veränderungsprozessen entwickeln sich dadurch einerseits in vielen GuG-Gruppen (z.B. Professor\*innen, Mittelbau, Studierende) Change Agents (SPECHT et al. 2017), die als prototypische Vorbilder (z.B. erhöhte, selbsterprobte Methodenkompetenz) und als Ansprechpersonen fungieren oder neue Diskurse anstoßen können, und andererseits verändert sich in der Lehreinheit der Stellenwert der Lehre nachhaltig.

Da die Weiterentwicklung von universitären Lehr-Lernsystemen stets auch als Organisationsentwicklungsprozess betrachtet werden kann, leiten sich daran ankoppelnd auch die beobachtbaren Gelingensfaktoren ab. Sie bestehen in GuG insbesondere – nicht priorisierend gereiht – in

- der Berücksichtigung der Organisationskultur und damit implizit Würdigung des Autonomiebedürfnisses (DECI & RYAN 1985) von universitär forschenden und lehrenden Individuen (WEICK 1976),
- der ganzheitlichen Betrachtung des Lehr-Lernsystems z.B. durch Einbeziehen von Studierenden,
- agiler Prozessorientierung (HANFT et al. 2017),

- Erweiterung der verfügbaren Ressourcen (z.B. zusätzliche Mitarbeitende durch Drittmittelprojekte),
- der Berücksichtigung von Außenperspektive durch Einbeziehung von hochschuldidaktischen Expert\*innen,
- der Selbstverpflichtung von Entscheidungsträgern (z.B. Studiendekan\*in, Professor\*innen) z.B. durch Übernahme von aktiver Projektleitungsfunktion,
- Transparenz und Kommunikation im Prozess,
- der Neugier von universitär lehrenden Menschen und
- der Kenntnis, dass Veränderung Zeit benötigt.

Der Fokus für künftige Weiterentwicklungen des GuG-Lehr-Lernsystems liegt insbesondere in der nachhaltigen Vernetzung und Verstetigung von innovativen Lehr-Lernsettings und Maßnahmen, da Ende 2020 bedeutsame Finanzmittel auslaufen. Darüber hinaus arbeitet die Lehreinheit GuG aktuell an der Umsetzung der Rückmeldungen der in 2019 abgeschlossenen Reakkreditierung der deutschsprachigen Studiengänge und der Ausgestaltung des Übergangs in den neuen englischsprachigen MSc-Studiengang „Remote Sensing and Geoinformatics“ für nicht KIT-Studierende. Hier gilt es insbesondere im Vorfeld der Aufnahme des Studiums zur Gewährleistung von motivierenden Lehr-Lernsettings in Präsenzlehrveranstaltungen das Fachkompetenzniveau dieser Studierenden zu erfassen und Angebote zur individuellen Weiterqualifikation zu schaffen. Weiterhin möchte die Lehreinheit GuG in Kooperation mit dem DVW Baden-Württemberg e.V. die Wahrnehmbarkeit von BSc-Abschlussarbeiten erhöhen und auf der Basis des Fachspezifischen Qualifikationsrahmens Geodäsie und Geoinformatik (FQR\_GG 2018) in Kooperation mit dem Arbeitskreis 1 des DVW e.V. und der Konferenz der GeodäsieStudierenden KonGeoS die erworbenen Kompetenzen von BSc-Absolvent\*innen evaluieren.

## 5 Literaturverzeichnis

- BERENDT, B., 2006: Gut geplant ist halb gewonnen. Teilnehmerzentrierte Struktur- und Verlaufsplanung von Lehrveranstaltungen. Neues Handbuch Hochschullehre. Lehren und Lernen effizient gestalten, B 1.1.
- BEYERLIN, S., KLINK, K. & DIEZ, A., 2014: Forschungsorientierte Lehre: Konzept und Umsetzungsstrategien auf institutioneller und praktischer Ebene am Beispiel des Karlsruher Instituts für Technologie (KIT). Neues Handbuch Hochschullehre. Lehren und Lernen effizient gestalten, Berendt, B., Fleischmann, A., Wildt, J., Schaper, N. & Szczyrba, B. (Hrsg.), A3.13, 41-61.
- BIGGS, J. & TANG, C., 2011: Teaching for Quality Learning at University: What the Student does. 4. Aufl., Open University Press, Maidenhead.
- DECI, E. L. & RYAN, R. M., 1985: Intrinsic motivation and self-determination in human behavior. Plenum, New York.
- FLEISCHMANN, A., SCHROEDER, J. & TUSCHAK, J., 2017: Nichtweiterbildung. Vorbehalte, Hemmschwellen und Einwände gegen hochschuldidaktische Weiterbildung. Neues Handbuch Hochschullehre, L1.34.

- FQR\_GG, 2018: Fachspezifischer Qualifikationsrahmen Geodäsie und Geoinformation. [online verfügbar unter [https://www.dvw.de/sites/default/files/landesverein-bb/aktuelles/FQR\\_GG.pdf](https://www.dvw.de/sites/default/files/landesverein-bb/aktuelles/FQR_GG.pdf); letzter Zugriff: 10. Jan. 2020]
- HANFT, A., MASCHWITZ, A. & STÖTER, J., 2017: Agiles Projektmanagement an Hochschulen – get the things done. Synergie Fachmagazin für Digitalisierung in der Lehre **3**, 8-15.
- HEUBLEIN, M., RABOLD, J., HIRSCH-WEBER, A. & MAYER, M., 2018a: Feedbackregelkreise der Wissenschaftskommunikation im BSc-Studiengang „Geodäsie und Geoinformatik.“ Forum für Lehre – multiperspektivischer kollegialer Austausch über Lehre und Lernen am KIT, 6. November 2018, Karlsruhe. [online verfügbar unter: [http://www.gik.kit.edu/downloads/Poster\\_ForumFuerLehre\\_Wissenschaftskommunikation\\_final.pdf](http://www.gik.kit.edu/downloads/Poster_ForumFuerLehre_Wissenschaftskommunikation_final.pdf); letzter Zugriff: 10. Jan. 2020]
- HEUBLEIN, M., RABOLD, J. & MAYER, M., 2018b: Synergien@KIT durch „Fit für Studium und Beruf“ im BSc-Studiengang Geodäsie und Geoinformatik. Forum für Lehre – multiperspektivischer kollegialer Austausch über Lehre und Lernen am KIT, 6. November 2018, Karlsruhe. [online verfügbar unter: [http://www.gik.kit.edu/downloads/Poster\\_ForumFuerLehre\\_Synergien\\_FfSB\\_final.pdf](http://www.gik.kit.edu/downloads/Poster_ForumFuerLehre_Synergien_FfSB_final.pdf); letzter Zugriff: 10. Jan. 2020]
- HEUBLEIN, M., STÖVHASE, J.-M., MAYER, M. & HINZ, S., 2018c: Lernen<sup>Coaching</sup> im BSc-Studiengang „Geodäsie und Geoinformatik“ - Teilprojekt von KIT-Lehre<sup>ForschungPLUS</sup>. dghd 2018 „Hochschuldidaktik als professionelle Verbindung von Forschung, Politik und Praxis“, 28. Februar – 2. März 2018, Karlsruhe. [online verfügbar unter <https://www.lehre-hoch-forschung.kit.edu/img/Poster%202.3%20GuG.pdf>; letzter Zugriff: 10. Jan. 2020]
- HEß, K. & KLINK, K., 2018: Tandems for Educational Development: A novel approach to accelerate conceptual change in KIT's faculties. The International Consortium for Educational Development Conference, 5.-8. Juni 2018, Atlanta, Georgia, US. [online verfügbar unter: [https://www.iced2018.com/sites/default/files/slides/Poster\\_ICED\\_TandemsForEducationalDevelopment.pdf](https://www.iced2018.com/sites/default/files/slides/Poster_ICED_TandemsForEducationalDevelopment.pdf); letzter Zugriff: 10. Jan. 2020]
- HILBRICH, R. & R. SCHUSTER, 2014: Qualität durch Differenzierung? Lehrprofessuren, Lehrqualität und das Verhältnis von Forschung und Lehre. Beiträge zur Hochschulforschung **36**, 71-89.
- HORX, M., 2011: Das Megatrend-Prinzip – Wie die Welt von morgen entsteht. Deutsche Verlagsanstalt, München.
- HUBER, L., 2009: Warum Forschendes Lernen nötig und möglich ist. Forschendes Lernen im Studium: aktuelle Konzepte und Erfahrungen, Motivierendes Lehren und Lernen in Hochschulen, Huber, L., Hellmer, J. & Schneider, F. (Hrsg.), Universitätsverlag Webler, Bielefeld, **10**, 9-35.
- ILLNER, M., MAYER, M. & BREUNIG, M., 2018: Die universitäre Geodäsieausbildung in Karlsruhe – Historie, Gegenwart und Zukunft. In: Seitz, K. (Hrsg.): Festschrift zur 150-Jahr-Feier des Geodätischen Instituts (1868-2018). 2018-2, 275-290, Schriftenreihe des Studiengangs Geodäsie und Geoinformatik, KIT Scientific Publishing, Karlsruhe.
- JENERT, T., 2008: Ganzheitliche Reflexion auf dem Weg zu Selbstorganisiertem Lernen. In: Häcker, T., Hiltensauer, W., Reinmann, G.: Bildungsforschung, **5**(2).
- KEHM, B. M., 2012: Hochschulen als besondere und unvollständige Organisationen? - Neue Theorien zur ‚Organisation Hochschule‘. In: Wilkesmann, U., Schmid C.J. (Hrsg.): Hochschule als Organisation. Organisationssoziologie. VS Verlag für Sozialwissenschaften, Wiesbaden.

- KING, A., 1990: Reciprocal Peer-Questioning A Strategy for Teaching Students How to Learn from Lectures. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas* **64**(2), 131-135.
- KIT, 2015: KIT-Leitfaden: Forschungsorientierte Lehre. [online verfügbar unter: [https://www.peba.kit.edu/downloads/Leitfaden-Forschungsorientierte\\_Lehre\\_am\\_KIT.pdf](https://www.peba.kit.edu/downloads/Leitfaden-Forschungsorientierte_Lehre_am_KIT.pdf); letzter Zugriff: 10. Jan. 2020]
- KLINK, K., STRICKER, M., SCHOSTOK, P. & KLINK, K., 2016: Gruppeninterview als Methode zur Entwicklung der Hochschullehre – am Beispiel von Lehrveranstaltungsevaluationen des Karlsruher Instituts für Technologie (KIT). 45. Jahrestagung der Deutschen Gesellschaft für Hochschuldidaktik (dghd) „Gelingende Lehre: erkennen, entwickeln, etablieren“, 21.-23. Sept. 2016, Bochum.
- KRATHWOHL, D.R., 2002: A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice* **41**(4), 212-218.
- LÜBECK, D., 2010: Wird fachspezifisch unterschiedlich gelehrt? Empirische Befunde zu hochschulischen Lehransätzen in verschiedenen Fachdisziplinen. *Zeitschrift für Hochschulentwicklung* **5**(2), 7-24.
- MAYER, M. & HEß, K., 2018: Mit Lehrinnovationen einen Grundstein für Studiengangentwicklung legen am Beispiel des kleinen KIT-BSc-Studiengangs „Geodäsie und Geoinformatik“. dghd 2018 „Hochschuldidaktik als professionelle Verbindung von Forschung, Politik und Praxis“, 28. Februar – 2. März 2018, Karlsruhe. [online verfügbar unter: [https://www.gik.kit.edu/downloads/Poster\\_Lehrinnovationen\\_V4.pdf](https://www.gik.kit.edu/downloads/Poster_Lehrinnovationen_V4.pdf); letzter Zugriff: 10. Jan. 2020]
- MAYER, M., KUTTERER, H. & CERMAK, J., 2019: Forschungsorientiert und kompetent – Ausgestaltung von hochschulischen Veränderungsprozessen am Beispiel der Lehreinheit »Geodäsie und Geoinformatik« am Karlsruher Institut für Technologie (KIT). *zfv*, **144**(4), 206-222.
- MAYER, M., RABOLD, J., RAIBLE, B. & MUHLER, M., (2020): Geodäsie und Geoinformatik – Ein interkulturelles Berufsfeld mit hoher gesellschaftlicher Relevanz. *Mitteilungsheft, DVW Baden-Württemberg e.V.* [im Druck]
- NIETHAMMER, C., KOGLIN-HESS, I., DIGEL, S. & SCHRADER, J., 2014: Herausforderung Curriculumentwicklung: ein konzeptioneller Ansatz zur Professionalisierung. *Zeitschrift für Hochschulentwicklung* **9**(2), 27-39.
- RABOLD, J., HEUBLEIN, M., MAYER, M. & CERMAK, J., 2019a: Selbstkompetente Studieneingangsphase – Wirkungsanalyse prototypischer Unterstützungsinstrumente. Tagungsband zum 4. Symposium zur Hochschullehre in den MINTFächern, DiNa Sonderausgabe, 165-173.
- RABOLD, J., HEUBLEIN, M. & MAYER, M., 2019b: Weiterentwicklung von Lehre als implizite Kompetenz Studierender. dghd 2019 „(Re-)Generation Hochschullehre. Kontinuität von Bildung, Qualitätsentwicklung und hochschuldidaktischer Praxis“, 5. – 8. März 2019, Leipzig. [online verfügbar unter: [https://www.gik.kit.edu/downloads/DGHD19\\_Poster\\_final.pdf](https://www.gik.kit.edu/downloads/DGHD19_Poster_final.pdf); letzter Zugriff: 10. Jan. 2020]
- REINMANN, G. & JENERT, T., 2011: Studierendenorientierung: Wege und Irrwege eines Begriffs mit vielen Facetten. *Zeitschrift für Hochschulentwicklung* **6**(2), 106-122.
- SPECHT, J., KUONATH, A., PACHLER, D., WEISWEILER, S. & FREY, D., 2017: How change agents' motivation facilitates organizational change: Pathways through meaning and organizational identification. *Journal of Change Management* **18**(3), 198-217.

- STOLZ, K. & SCHOLKMANN, A., 2014: Modulübergreifende Studiengangsentwicklung – das Beispiel hochschuldidaktischer Planungskompetenz. *die hochschullehre* **1**(1), 1-14.
- WEICK, K.E., 1976: Educational organizations as loosely coupled systems. *Administrative Science Quarterly* **21**(1), 1-19.
- WHITMAN, N. A. & FIFE, J. D., 1988: Peer Teaching: To Teach Is To Learn Twice. ASHE-ERIC Higher Education Report No. 4.
- WISSENSCHAFTSRAT, 2005: Empfehlungen zur Ausgestaltung von Berufungsverfahren. [online verfügbar unter: <https://www.wissenschaftsrat.de/download/archiv/6709-05.pdf>; letzter Zugriff: 10. Jan. 2020]
- ZQ [Zentrum für Qualitätssicherung und -entwicklung] & PROGNOSE AG, 2016: Evaluation des Bund-Länder-Programms für bessere Studienbedingungen in der Lehre (Qualitätspakt Lehre). Abschlussbericht über die 1. Förderperiode 2011-2016 im Auftrag des BMBWF. [online verfügbar unter: [https://www.bmbwf.de/files/Abschlussbericht\\_Evaluation\\_barrierefrei.pdf](https://www.bmbwf.de/files/Abschlussbericht_Evaluation_barrierefrei.pdf); letzter Zugriff: 10. Jan. 2020]

# Erfahrungen mit einem Regelbetrieb der Erstellung von E-Learning-Angeboten für Photogrammetrie und Mobile 3D-Vermessung

JÜRGEN SELENSKI<sup>1</sup>, VALENTIN WICH<sup>1</sup> & ANSGAR BRUNN<sup>1</sup>

*Zusammenfassung: Seit einigen Jahren setzt die Hochschule für angewandte Wissenschaften Würzburg-Schweinfurt (FHWS) Blended Learning und Blended Networking erfolgreich ein. Die Methodik findet in der Ausbildung im Projekt „SMART vhb“ sowie in der Weiterbildung im Projekt „Mobile 3D Vermessung für kleine und mittlere Unternehmen“ (MV-3D4KMU) Anwendung. Die Erstellung der Online-Lerneinheiten im Regelbetrieb erfordert die Realisierung einer Prozesskette, die Schaffung klarer Rahmenbedingungen, die Auswahl von Themen und Inhalten der Lektionen, das Schreiben von Drehbüchern, die Umsetzung der Lektionen, eine Qualitätssicherung in der Herstellung und Überführung der Lektionen in den Regellehrbetrieb durch ein Repositorium. Die Intention einer Netzwerkbildung zwischen den Lehrenden erfordert zusätzlich Interaktionen in Präsenzveranstaltungen und eine Dokumentations- und Austauschplattform. Dieser Beitrag behandelt die Ziele und Motivationen der Projekte und gibt methodische Einblicke in didaktische, technische und organisatorische Konzepte. Praktische Einblicke stellen bereits erprobte Einsatzfelder von Blended Learning und Blended Networking vor und ermöglichen Bewertungen des Erreichten.*

## 1 Einführung

### 1.1 Motivation

Internetbasiertes Lernen und internetbasierte Weiterbildung werden in einer zunehmend digitalisierten Welt immer wichtiger. Begriffe wie „E-Learning“ und „Blended Learning“ sind in aller Munde. Der Bedarf an Lehrangeboten wird von den Lernenden formuliert. Der Bedarf zusätzlicher Bildung ist notwendig und weitgehend erkannt. Das moderne Lernumfeld an Hochschulen und just-in-time und just-in-sequenz Arbeitsprozesse erfordern von allen Beteiligten eine hohe Flexibilität, der sich die persönliche Weiterbildung vielfach unterordnen muss. Die Lehrenden versuchen gleichzeitig motivierende Angebote der Wissensvermittlung zu finden. Interbasierte Lernangebote bieten hier eine weite Palette von Möglichkeiten, die das Labor für Photogrammetrie und Fernerkundung (LPF@FHWS) der Hochschule für angewandte Wissenschaften (FHWS) in zwei Projekten untersucht hat.

„Blended Learning“ wird in der Literatur auf unterschiedliche Arten definiert (z.B. HORN & STAKER 2011; OSGUTHORPE & GRAHAM 2003). Im Projekt „Digitale Photogrammetrie für Bildverarbeitung für kleine und mittlere Unternehmen (DiPhoBi4KMU)“ ist Blended Learning als Methode bereits eingesetzt worden (MEYER & BRUNN 2016). Für das Projekt Smart vhb werden die Rahmenbedingungen im Hinblick auf das Hochschulumfeld modifiziert. Die Zielgruppe besteht hier aus den Mitgliedern der bayerischen Hochschulen, vorwiegend Studierenden, die im Rahmen von Lehrveranstaltungen oder auch selbstmotiviert einzelne Lehreinheiten zur per-

---

<sup>1</sup> Hochschule für angewandte Wissenschaften Würzburg-Schweinfurt, Röntgenring 8, D-97070 Würzburg, E-Mail: [juergen.selenski,valentin.wich,ansgar.brunn]@fhws.de

sönlichen Aus- und Weiterbildung bearbeiten. Seitens des Projektträgers sind beide Projektlinien eng miteinander verbunden und sollen gegenseitig nutzbare Lehrmaterialien generieren (VIRTUELLE HOCHSCHULE BAYERN 2020).

WICH & BRUNN (2019) beschreiben den Einsatz von Blended Learning im Rahmen von Blended Networking. Blended Networking setzt auf die verstärkte Netzwerkbildung zwischen den Teilnehmern und der Hochschule. Angepasst an die Projektziele des jeweiligen Projektes sind eine eigene Organisationsstruktur und ein eigenes methodisches Vorgehen notwendig. Praktische Umsetzungsbeispiele demonstrieren den Einsatz von Autorensoftware zur Erstellung von Bildungsinhalten und Lernmanagementpaketen zur Bildungsweitergabe an eine Onlineplattform (FHWS 2020).

## 1.2 Einordnung der Arbeiten

Jedes Projekt besitzt eine eigene Organisationsstruktur. Dieser Beitrag zeigt Lehrenden die Möglichkeiten und Erfahrungen der Umsetzung von Blended Learning und Blended Networking auf.

Mit SMART vhb möchte die Virtuelle Hochschule Bayern (vhb) ihr Kursangebot erweitern und gleichzeitig ein attraktives Modell zur Nutzung des Blended-Learning Konzeptes schaffen. Im Rahmen des Projektes entstehen daher flexibel einsetzbare digitale Lerneinheiten, welche den Lehrenden der Trägerhochschulen zur Ergänzung der Präsenzlehre bereitgestellt werden. Der Einsatz von Autorenwerkzeugen ermöglicht es, anschauliche Kurse in hoher Qualität zu produzieren. Die hochschulübergreifende Nutzung erfolgt über ein zentrales Repositorium (VIRTUELLE HOCHSCHULE BAYERN 2019).

In MV-3D4KMU erfolgt eine Vernetzung zwischen Hochschule und bayerischen KMU. Der Wissenstransfer zwischen Hochschule und Firmen erlaubt den regen Austausch von aktuellen Forschungsergebnissen und Erfahrungen aus bewährten praktischen Anwendungen der Unternehmen. Die Präsenztreffen haben einen Workshop-Charakter und werden mit freier und Open Source Software (FOSS) (STALLMAN 2020) umgesetzt. Synergieeffekte ergeben sich in der Arbeit in Workshops im Projektnetzwerk und in der Durchführung von gemeinsamen Projekten und Abschlussarbeiten mit Studierenden der Hochschule.

Die FHWS strebt die Digitalisierung der Lehre an und erweitert kontinuierlich das hauseigene online Kursangebot (FHWS, 2018). Insbesondere das SMART vhb Projekt schafft viele neue online Lektionen. MV-3D4KMU bietet Veranstaltungen zur Aus- und Weiterbildung für Projektteilnehmer an und Perspektiven für Studierende.

Im folgenden Kapitel werden zunächst die Grobprozesse beschrieben. Im Rahmen des dritten Kapitels folgenden Details der Realisierung, die eng mit Bewertungen und Erfahrungen verbunden sind. Das letzte Kapitel zieht ein Fazit und zeigt Perspektiven auf zukünftige Entwicklungen.

## 2 Prozessdefinition

### 2.1 Einordnung

Blended Learning ist eine Kombination aus Präsenzveranstaltung und online stattfindender Lehre, welche sich sowohl sozialen, als auch inhaltlich-effektiven Faktoren auf neue Art und Weise zuwendet (HOLTEN & NITTEL 2010). Die vhb beschreibt ihr Blended Learning Konzept als die Bereitstellung eines hochschulübergreifend nutzbaren Repositoriums zur Einbindung digitaler Lerneinheiten in Präsenzveranstaltungen. Lerneinheiten werden mit inhaltlich in sich

geschlossenen Lernzielen definiert, dürfen aber dennoch aufeinander aufbauen. Variierende Lehrmaterialien, wie Videos, Texte oder Simulationen, garantieren eine mediendidaktisch abwechslungsreiche Darstellung der jeweiligen Thematik (VIRTUELLE HOCHSCHULE BAYERN 2020).

In DiPhoBi4KMU ist das Bedürfnis der Netzbildung zwischen allen Teilnehmern des E-Learning-Angebots entstanden, um den technisch-fachlichen Austausch als auch das didaktische Konzept, welches vom freien Wissensaustausch lebt, besser umzusetzen. Diese Methodik wird in MV-3D4KMU weiterentwickelt. Gesammeltes Wissen wird in Best-Practice-Leitfäden aufbereitet (BRUNN 2018). Präsenztreffen dienen der Stärkung des sozialen Aspektes. Aktuelle technische Lösungen für spezifische Anliegen können im Netzwerk einfach und effektiv diskutiert und entwickelt werden.

## 2.2 E-Learning Angebot in SMART vhb

Viele der Rahmenbedingungen sind durch die vhb vorgegeben (VIRTUELLE HOCHSCHULE BAYERN 2019), um einen einheitlichen Auftritt des bayernweiten Bildungsangebots zu realisieren. Die Vorgaben beziehen sich auf Aspekte des methodischen Aufbaus und des Bildungsformats der online Kurse. Lerneinheiten sollen folgende vier Kriterien erfüllen:

- eine Bearbeitungsdauer von ca. 45 Minuten aufweisen,
- multimediale Lernmaterialien beinhalten,
- in sich geschlossene Lektionen ergeben und
- über definierte Lernziele mit Selbstkontrollmechanismen verfügen.

Aufgrund der großen Anzahl von Lerneinheiten -an der FHWS wurden mehr als 100, im Studienbereich Geo der FHWS mehr als 60 Lerneinheiten und 30 davon im LPF@FHWS erstellt-sind weitergehende zentrale Festlegungen erforderlich. So wurde z.B. das Design und das Layout auf Hochschulebene abgestimmt.

Innerhalb des LPF@FHWS wurde eine Prozesskette definiert (siehe Abb. 1). Der Lehrende schlägt zunächst die Titel der Kurse mit Kurzbeschreibungen der vhb vor. Diese gibt die eingereichten Lerneinheiten zur Produktion frei, wenn Doppelungen im Repositorium nicht zu erwarten sind. Nach der Definition der einzelnen fachlichen Inhalte wird ein Drehbuch erstellt. In diesem Schritt erfolgen eine Literaturrecherche und eine Zusammenstellung verfügbarer und zu erstellenden multimedialer Materialien. Die Prüfung der festgehaltenen Lehrinhalte durch den Lehrenden markiert die Freigabe zur Umsetzung des Kurses. Studentische Hilfskräfte bearbeiten nach der ersten Realisierung erstellte Lerneinheiten als Teil der Qualitätssicherung und evaluieren die Einheiten anhand vorgefertigter Bewertungsbögen, nach der dann die Kurse ein weiteres Mal bearbeitet werden.

Die Abgabe der Lerneinheiten in das Repositorium erfolgt über das Content-Management-System (CMS) edu-sharing (METAVENTIS GMBH 2020) der vhb. Die Kurse werden um Metadaten und die Kurzbeschreibung mit dem Titel, einigen stichpunktartigen Kursinhalten, einem Vorschaubild und Hinweisen zur Nutzung ergänzt. Die vhb prüft jede Lerneinheit nach den projektspezifischen Programmkriterien und weiteren formalen Vorgaben. Nach der Freigabe durch die vhb erscheinen die Lerneinheiten in einem (hochschul-bayernweit) öffentlichen Bereich des Repositoriums. Lehrende und Lernende aller bayerischer Trägerhochschulen haben die Möglichkeit auf alle freigegebenen Lektionen zuzugreifen. Lehrende können die Lektionen ins hauseigene Learning-Management-System (LMS), z.B. Moodle (HILGENSTOCK 2020) verlinken und in ihre Kurse integrieren.

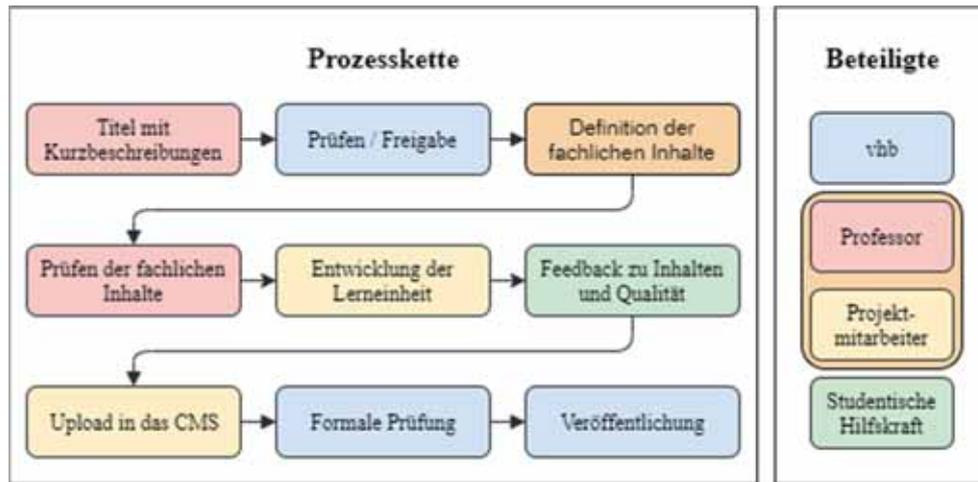


Abb. 1: Prozesskette der Erstellung des e-Learning Angebotes im Projekt SMART vhb mit Zuständigkeiten der Beteiligten.

### 2.3 Blended Networking Angebot in MV-3D4KMU

Die Rahmenbedingungen in MV-3D4KMU legen fest, dass innovative fachspezifische Inhalte an die beteiligten Unternehmen von der Hochschule weitergegeben werden sollen. Durch eine Netzwerkbildung sollen zwischen den KMU Synergieeffekte erzielt werden. Das aus diesen eher didaktischen Anforderungen entwickelte strukturelle Konzept ist in Abb. 2 dargestellt.

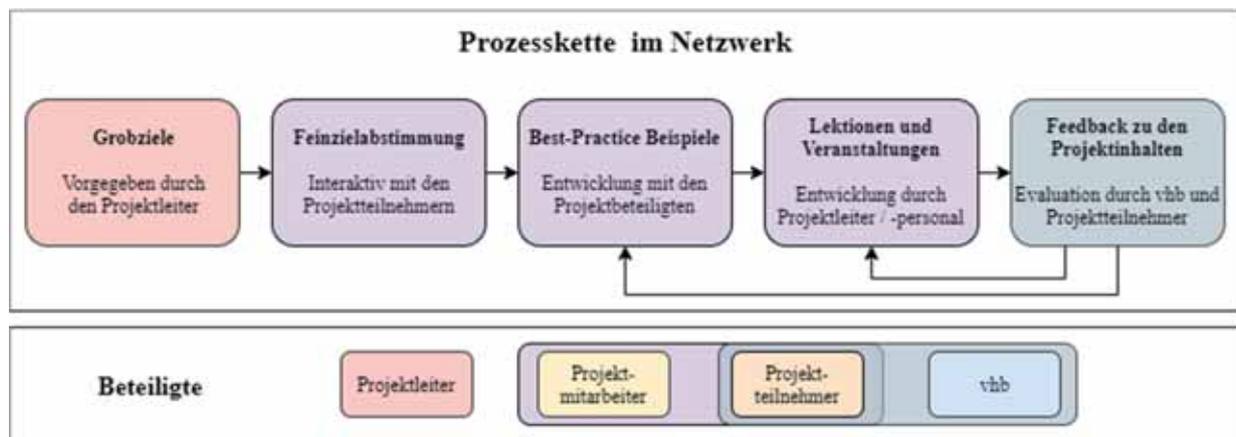


Abb. 2: Prozesskette zur Gestaltung der Projektkontente mit Verantwortlichkeiten und Rückkopplung. Optimierungsmechanismen helfen die Inhalte auf die Bedürfnisse der Teilnehmer abzustimmen.

Die inhaltliche Rahmenvorgabe erfolgt durch die Hochschule. Mit den KMUs erfolgt die detaillierte Ausarbeitung der Themenliste und die gemeinsame Definition von Themen für Best-Practice-Beispiele. Die Umsetzung der Ziele und der Best-Practice-Beispielen findet im Netzwerk statt.

Grobziele definieren die Teilschritte der Umsetzung im Projekt. Die Projektteilnehmer lernen z.B. mobile Aufnahmetechniken und -systeme, deren Anwendungsfelder und Marktsegmente, kennen und stimmen in einem interaktiven Webinar über die geplanten Feinziele im Projekt ab.

Präsenzveranstaltungen dienen neben der Kommunikation unter den Teilnehmern der Vermittlung technischer Aspekte der mobilen 3D Vermessung. Praktische Projekte dienen der Vermittlung von Knowhow in der Planung und Anwendung der Techniken. Daraus entstehen Best-

Practice-Anwendungen im Umgang mit Hard- und Software. In Newslettern, Videos und Online-Lektionen werden diese weitergegeben. Ein Feedback gibt Auskunft über die Qualität aus Sicht der Projektteilnehmer. Das E-Learning dient ebenso als Repositorium und zentrale Kommunikationsplattform.

### 3 Realisierung

#### 3.1 E-Learning Angebot in SMART vhb

Die *Realisierung* der einzelnen Lerneinheiten ist von zwei Phasen geprägt, der Drehbuchphase und der eigentlichen technischen Entwicklung (Umsetzung) der Lerneinheit (Abb. 3).

Ausgangspunkt der Drehbuchentwicklung sind die festgelegten Lernziele. In einem tabellarischen Textdokument (Drehbuch, Storyboard) erfolgt die Dokumentation der Recherche. Alle genutzten Materialien sollen ohne fremde Urheberrechtsansprüche verwertbar sein, da die Einholung von Nutzungsrechten in dem engen Zeitrahmen nicht umsetzbar wäre. Die Inhalte sind im Drehbuch chronologisch strukturiert. Der Medieneinsatz und die Art der Interaktionen sind bewusst abwechslungsreich gehalten. Dies erzeugt ein mediendidaktisch wirksames Erscheinungsbild. Zur Selbstkontrolle sind an geeigneten Stellen Rechenaufgaben, Quizze oder Zuordnungsaufgaben konzipiert. Das vorgesehene Bildschirmlayout ist anhand von Gestaltungsvorlagen festgehalten. Durch kurz gehaltene Bildschirmtexte erfolgt eine in sich geschlossene Verbindung der Lehrinhalte. Der Lehrende prüft das fertige Drehbuch, kann bei Bedarf redaktionelle Änderungen einpflegen und gibt das Drehbuch zur technischen Umsetzung frei.

Die Umsetzung des Drehbuchs erfolgt mit dem Autorentool Lectora (CHEMMEDIA AG 2019). Der Projektmitarbeiter arbeitet die Texte, Videos, Bilder, Interaktionen und Selbstkontrollaufgaben aus dem Drehbuch in hochschuleigene Gestaltungsvorlagen ein. Diese sind vom Zentrum für digitale Lehre der FHWS zur Verfügung gestellt. Dies begünstigt zum einen den einheitlichen Auftritt der Lektionen, zum anderen wird dadurch der Entwicklungsaufwand verringert. Am Ende der Autorenphase entsteht ein SCORM-Paket (RUSTICI SOFTWARE LLC 2020). Dieses ist in ein LMS integrierbar

Innerhalb der Umsetzungsphase prüfen Studierende die erstellten Lektionen auf Interaktions-, Rechtsschreib- und Formatierungsfehler. Vorgefertigte Evaluationsbögen dienen dem Sammeln von Feedback zur Qualität und der Verbesserung der Lerneinheiten. Die Lektionen werden entsprechend der Qualitätsbewertung überarbeitet. In einzelnen Fällen erfolgt eine zweite Qualitätsbewertung.

Der Upload der fertigen Lektionen ist als SCORM<sup>2</sup>-Paket in das CMS der vhb realisiert. Lektionen erhalten Metadaten und ein beschreibendes PDF-Dokument. Nach formaler Prüfung durch die vhb erfolgt die Freigabe zur hochschulübergreifenden Nutzung.

*Erfahrungen:*

Im Rahmen der ersten Förderrunde sind 30 Lerneinheiten des LPF@FHWS im SMART vhb Repositorium entstanden. Davon behandeln 15 Lerneinheiten Themen aus der Photogrammetrie (Abb. 4 und 5) und 15 weitere Themen aus den Bereichen Laserscanning und Punktwolkenverarbeitung. Die Drehbuchphase ist mit dem größten Zeitaufwand verbunden. Hierbei sind fünf bis acht Arbeitstage je Lektion zu veranschlagen. Die Herausforderung liegt hauptsächlich in der inhaltlichen und medialen Aufbereitung der Wissensinhalte. Bei den Themen variiert die Fülle und Zugänglichkeit geeigneter Materialien teilweise sehr.

---

<sup>2</sup> SCORM (Sharable Content Object Reference Model)

Der effiziente Umgang mit dem Autorentool steigt mit zunehmendem Projektfortschritt. Die grundlegende Einarbeitung in das Autorentool Lectora legt den Grundstein für den Umgang mit der komplexen Software. Die Entwicklungszeit konnte durch die gewonnene Erfahrung von ca. fünf, auf zwei bis drei Tage zum Projektende reduziert werden.

Für die Optimierung sind ca. zwei Tage inkl. der studentischen Prüfung anzusetzen. Insgesamt verringerte sich der Arbeitsaufwand während der Projektumsetzung von ca. 15 auf annähernd zehn Arbeitstage pro Lerneinheit.

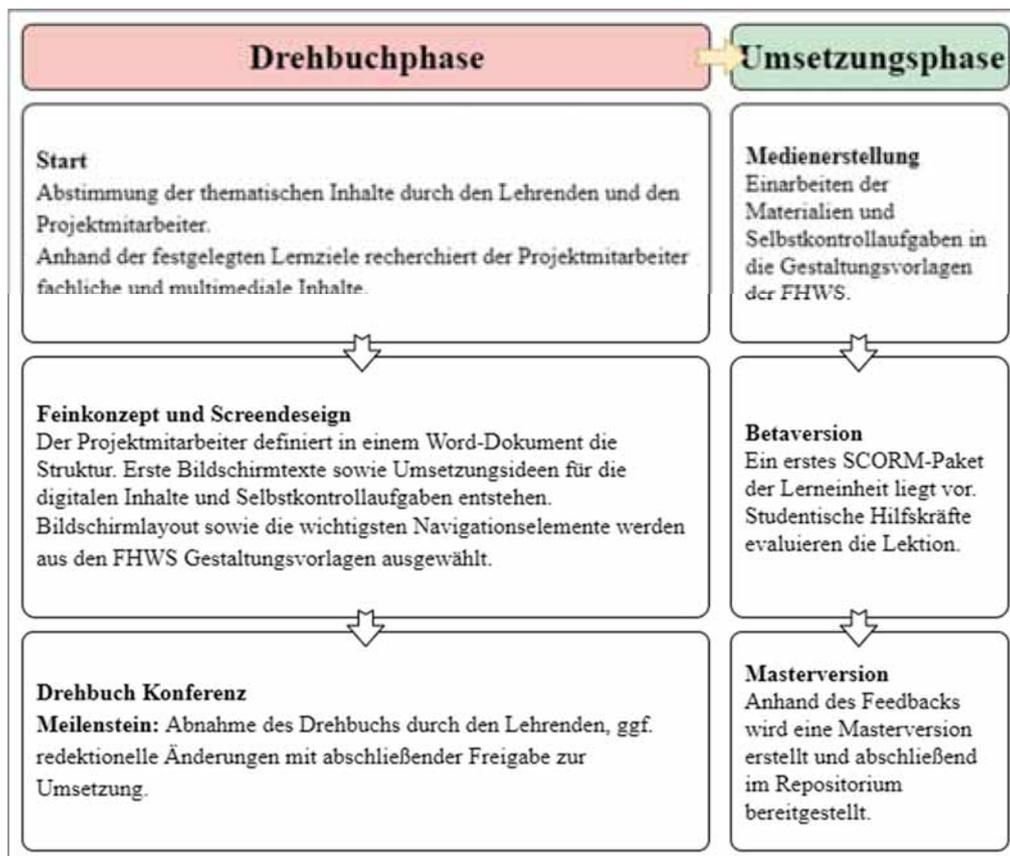


Abb. 3: Konzeption und Produktion einer Lerneinheit mit Arbeitsschritten und Verantwortlichkeiten

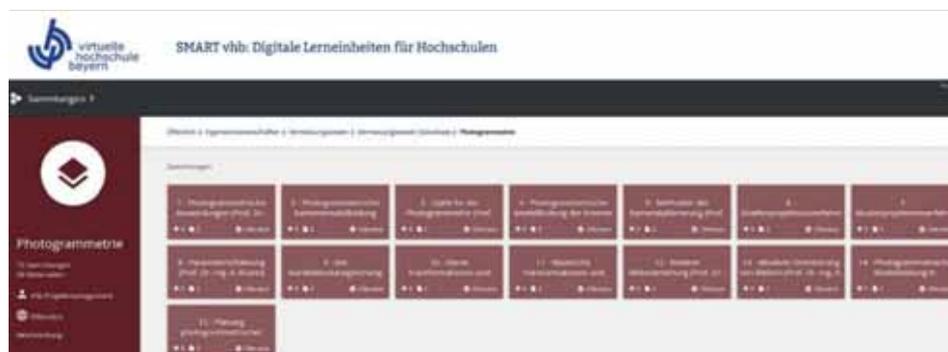


Abb. 4: Ausschnitt aus dem SMART vhb Repository mit Sammlungen der erstellten Lerneinheiten. Jede rote Schaltfläche repräsentiert eine Lerneinheit

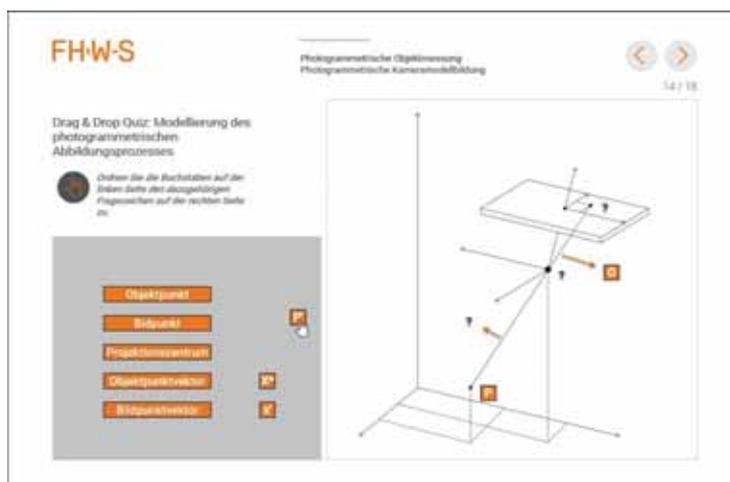


Abb. 5: Seite einer Lerneinheit „Photogrammetrische Modellbildung“, Seite 14, Drag & Drop Quiz zur Modellierung des photogrammetrischen Abbildungsprozesses

Die Qualitätsprüfung bringt eine wesentliche Qualitätssteigerung mit sich. Die Prüfung der Rechtschreibung und Formatierung sorgt für formale Richtigkeit. Feedback zu Bearbeitungszeiten, Komplexität der Kontrollaufgaben sowie der mediendidaktischen Gestaltung der Kurse erlaubt die Sicherstellung des Bearbeitungsflusses der Lektionen.

Das Feedback ermöglicht auch die zielgerichtete Weiterentwicklung der Lektionen. In SMART vhb ist zusätzlich ein Evaluationskonzept etabliert, welches auf dem Feedback der Lehrenden und Lernenden aufbaut. Bei Bedarf veranlasst die vhb Evaluationen einzelner Lerneinheiten. Dafür sind Fachexperten vorgesehen, deren Feedback in Berichtform dokumentiert wird (VIRTUELLE HOCHSCHULE BAYERN 2019).

### 3.2 Blended Networking in MV-3D4KMU

In der *Projektrealisierung*, in gemeinsamer Abstimmung ist innerhalb des Projekts eine Übersicht über die Projektinhalte entstanden. Diese Themen werden sequentiell bearbeitet und durch wöchentliche Newsletter im Projekt publiziert. Weitere Bildungsinhalte sind Lernvideos oder interaktive Webinare. Praxistreffen sind als Workshops realisiert. Die Ergebnisse aus den Workshops, Messen und Veranstaltungen sind im E-Learning verfügbar und können flexibel abgerufen werden. Die Ergebnisse werden bei Bedarf als Best-Practice-Lektionen umgesetzt. In der technischen Umsetzung für Webinare, aber auch Präsenzveranstaltungen hat sich Adobe Connect vom deutschen Forschungsnetzwerk (DFN) bewährt (VEREIN ZUR FÖRDERUNG EINES DEUTSCHEN FORSCHUNGSNETZES E. V. 2019). Es eignet sich zum Mitschnitt von Präsentationen, zur interaktiven Zusammenarbeit im Chatraum, Videochat oder zur moderierten Abstimmung auf einem interaktiven virtuellen Whiteboard. In Adobe Connect sind unterschiedliche Layouts mit unterschiedlichen vordefinierten Arten der Zusammenarbeit möglich. Dies ermöglicht die Präsentation einer Person zur Gruppe, die Bildung von Gruppen zur Zusammenarbeit oder die freie Zusammenarbeit aller Beteiligten.

Lernvideos bieten ein zeiteffizientes Format der Weiterbildung. In weniger als zehn Minuten lassen sich komplexe Teilbausteine anschaulich in Bild-, Ton- und Beschreibungstexten, realisiert als Videopräsentation, umsetzen. Die Lernvideos sind in die E-Learning Plattform integriert. Durch eine modulare Organisation der Videos in Folien sind diese mit wenig Zeitaufwand pflegbar.

Netzwerktreffen werden für die interaktive Bearbeitung von komplexen technischen Themen genutzt. Die Projektteilnehmer nehmen das Format zum Wissensaustausch positiv an, da das erworbene Fachwissen in der Veranstaltung angewendet wird. Durch Zeit für eigene Anwendungen und Fragenstellungen besteht die Möglichkeit, die Verfahren an eigenen Daten zu erproben und Lösungen im Netzwerk zu erarbeiten. Pausen dienen der Stärkung der sozialen Komponente und damit der Förderung der Netzwerkbildung.

#### *Erfahrungen:*

Durch die Online-Plattform lassen sich Kurznachrichten in kurzer Zeit verfassen und verbreiten. Einsatz finden diese zur Ankündigung von Ereignissen wie: Öffentlichen Tagungen und Veranstaltungen, Projekttreffen, Erscheinen von Newslettern und bei der Veröffentlichung neuer Videolektionen.

Die Bearbeitungsdauer von Newslettern ist abhängig von Art- und Umfang der vorzustellenden Ergebnisse. Diese schwankt zwischen mehreren Stunden hin bis zu zwei Wochen bei größeren Projekten, z.B. aus dem Mobile Mapping. Lehrvideos mit einer Laufzeit von weniger als zehn Minuten sind in etwa an einem Arbeitstag umsetzbar. Das Video ist ähnlich wie bei einer Präsentation mit Folien organisiert, sodass spätere Änderungen und Erweiterungen in möglich sind. Videos besitzen das Potential durch Audio- und Videoinhalte anschaulicher zu sein, als klassische Präsentationsfolien.

Die praktischen Workshops sind am Effizientesten in der Bildungsvermittlung, bedürfen aber auch einer sorgfältigen Vorbereitung. Je nach Aufwand und Vorbereitungsaufwand sind mehrere Tage bis ca. zwei Wochen Vorbereitungszeit einzuplanen. Anwendungen aus Hard- und Softwarethemen sorgen für einen Praxisbezug. Projektteilnehmer arbeiten interaktiv an Anwendungen am eigenen Notebook, um diese selbst nachzuvollziehen. Bei Fragen der Teilnehmer mit Relevanz für mehrere Projektteilnehmer werden diese für alle Teilnehmer in Workshops erklärt. Jeder Teilnehmer profitiert von den Gedanken und Anregungen der anderen Teilnehmer. Die Nachbereitung einer Veranstaltung dauert je nach Art der Veranstaltung einen halben Tag, bei Workshops, bis zwei Tage, z.B. bei Messebesuchen.

In Präsenztreffen werden Newsletter und Videolektionen besprochen. Dort können Projektteilnehmer Feedback geben. Über Gamification „Level-Up“ (BRANCH UP PTY LTD 2020) Elemente besteht eine Möglichkeit im LMS nachzuvollziehen, in welcher Art- und Umfang die Teilnehmer Inhalte im E-Learning nutzen. Das erlaubt eine Vereinfachung der Inhalte, eine Erweiterung durch zusätzliche Erklärungen oder auch die Verbesserung einer intuitiven Gestaltung von Interaktionen.

Die vhb selbst führt durch Gutachter eine Qualitätssicherung der Onlineangebote durch. Prüfkriterien sind hier der Umfang der Inhalte gemäß Projektantrag, der eingesetzte Medienmix, die formale Richtigkeit, das eingesetzte didaktisches Konzept und die Verknüpfung und Einordnung der Lehrinhalte.

## **4 Fazit**

Die Erstellung von online Lehrinhalten bietet für alle Beteiligten einen Mehrwert, durch die Erweiterung der Fachkenntnisse, didaktischen Methodenwissens, den Erwerb von Hard- und Softskills mit Werkzeugen wie Autorentools und in der praktischen Anwendung.

Damit eine nachhaltige Nutzung der Lektionen möglich ist, sind Inhalte ohne fremde Urheberrechtsansprüche einzusetzen. Das Lizenzmodell von den Creative Commons beschreibt mit

seiner CC0-Lizenz (CREATIVE COMMONS DEUTSCHLAND 2009) ein gemeinfreies Format, das eine uneingeschränkte, selbst kommerzielle, Nutzung ohne Namensnennung erlaubt.

Barrierefreiheit gewinnt weiter an Bedeutung und lässt sich in einfachster Form durch ein Drehbuch mit Bildbeschreibungen als strukturiertes PDF-Dokument durch die Nutzung von HTML Inhalten umsetzen. In beiden Varianten ist es möglich, Screenreader (z.B. den freien Screenreader NVDA (BECKMANN 2019)) zu nutzen.

Die Gestaltung von Lernpaketen erfolgt in einer endgeräteübergreifenden Form. Für Autorentools sind responsive Layoutvorlagen zu nutzen, welche sich auf das Bildschirmformat der Endgeräte anpassen. Lernende haben dadurch einen einfacheren Zugang zu den Lehrmaterialien und bekommen die Inhalte für das Endgerät entsprechend angezeigt.

Die Modularisierung, wie im E-Learning Angebot der SMART vhb, erlaubt die einfache Wiederverwendbarkeit von Kursen ohne Mehraufwand. Dabei sollte das Lernpaket im originalen Autorensoftwareformat sowie die Autorensoftware vorgehalten werden, damit spätere Änderungen einfach erfolgen können. Alle Lehrenden können durch das SMART vhb Repositoryum die Qualität der eigenen Präsenzveranstaltung steigern und die gewonnene Zeit zur intensiveren Betreuung der Lernenden einsetzen.

Lernende erhalten multimediale Lektionen auf Höhe der Zeit. Computer und Smartphone Nutzung sind selbstverständlich, der Konsum dieser beiden Medien hat im Vergleich zu den vorherigen Generationen stark zugenommen. Die Zeit- und Ortsunabhängigkeit begünstigt die Flexibilität in der Weiterbildung und schafft die Grundlage zur freien Wissensaneignung.

Die Lernenden erhalten in Selbstkontrollen innerhalb der Lektionen direktes Feedback über das erworbene Wissen. Es ist möglich, den Lernfortschritt an das LMS zu übergeben. Dadurch besteht sowohl für den Lehrenden, als auch dem Lernenden die Möglichkeit den Lernfortschritt zu messen. Lehrende erkennen frühzeitig, wie Lernende das Angebot nutzen und welche Aufgaben in Präsenzveranstaltungen näher zu besprechen sind.

Langfristig betrachtet eignet sich das Blended Learning Format besonders für die Erstellung von in sich abgeschlossenen Bildungsinhalten. Als Ergänzung der Präsenzlehre oder der Erweiterung des Lektionsangebots spart ein Repositoryum mit wiederverwendbaren Einheiten nachhaltig Zeit. Durch die Messbarkeit der Lernfortschritte im LMS erlauben diese eine zielgerichtete Form der Weiterbildung und geben mehr Raum für eine individuelle Betreuung (Binnendifferenzierung). Weiterhin ist ein Trend zu Massive Open Online Kursen (MOOC) (WULF et al. 2014) zu erkennen. Diese haben das Potential eine hochwertige Umsetzung und eine einfachere Finanzierung durch den größeren Nutzerkreis sicherzustellen. Feedback dient der kontinuierlichen Verbesserung der Inhalte. Die einfache Zugänglichkeit ermöglicht eine hochfrequente Nutzung, besonders auch von „digital Natives“ (Günther, 2007) auf unterschiedlichen Endgeräten. Komplexe Inhalte sind in mehrstufig gegliederten Lektionen mit stringenter didaktischem Konzept implementierbar.

Blended Networking eignet sich für den Einsatz in der weiterführenden Erwachsenenbildung und stärkt den freien Austausch über die Präsenzveranstaltungen hinaus, da in den Gruppen ein persönliches Kennenlernen erfolgt. Das fördert die nachhaltige Netzwerkbildung mit einem Kennenlernen aller Teilnehmer. Die soziale Komponente wird gestärkt. Synergien zwischen den Firmen und zwischen Firmen und Hochschule entstehen. Im Projektnetzwerk sind mehrere Projekt- und Bachelorarbeiten mit Studierenden und Netzwerkpartnern entstanden und berufliche Karrieren initiiert worden. Weiterführend ist die Aus- und Weiterbildungsmethodik auch für Firmen intern zur Fortbildung, in Schulungen mit Kunden, oder für den Wissenstransfer und der Vernetzung der Teilnehmer von Verbänden geeignet.

*Danksagung:* Die Entwicklungen dieses Beitrags wurden teilweise in dem Projekt „Mobile 3D-Vermessung für kleine und mittlere Unternehmen (MV-3D4KMU)“ durch die EU im Rahmen des Europäischen Sozialfonds und durch den Freistaat Bayern sowie teilweise durch die Virtuelle Hochschule Bayern (vhb) im Rahmen des Projekts „SMART vhb“ gefördert.

## 5 Literaturverzeichnis

- BECKMANN, U., 2019: NVDA - der freie ScreenReader für blinde Computernutzer. <http://www.bhvd.de/nvda/index.html> (letzter Zugriff 14.1.2020).
- BRANCH UP PTY LTD, 2020: Level up! Documentation. <https://levelup.plus/docs/> (letzter Zugriff 14.1.2020).
- BRUNN, A., 2018: Blended Learning in Photogrammetrie und Bildverarbeitung für regionale KMUs – Ein Erfahrungsbericht; Publikationen der DGPF, Band 27. 450-460.
- CHEMMEDIA AG, 2019: Lectora E-Learning Authoring. <https://chemmedia.lpages.co/lectora-landingpage/> (letzter Zugriff 8.1.2020).
- CREATIVE COMMONS DEUTSCHLAND, 2009: Neu im Programm: CC0 – Creative Commons Deutschland. <https://de.creativecommons.org/index.php/2009/03/17/neu-im-programm-cc0/> (letzter Zugriff 14.1.2020).
- GÜNTHER, J., 2007: Digital natives & digital immigrants. StudienVerlag, Innsbruck.
- Hilgenstock, R., 2020: Lernerfolg mit Moodle. <https://moodle.de/> (letzter Zugriff 8.1.2020).
- FHWS, 2020: E-Learning@FHWS. <https://elearning.fhws.de/> (letzter Zugriff 14.1.2020).
- FHWS, 2018: Digitalisierungsstrategie Lehre und Studium, FHWS intern.
- HOLTEN, R. & NITTEL, D., 2010: E-Learning in Hochschule und Weiterbildung: Einsatzchancen und Erfahrungen, Erwachsenenbildung und lebensbegleitendes Lernen. Bertelsmann, Bielefeld.
- HORN, M. B. & STAKER, H., 2011: The Rise of K–12 Blended Learning, Innosight Institute.
- METAVENTIS GMBH, 2020: Bildungscloud- & E-Learning-Lösungen. [edu-sharing.com](http://edu-sharing.com). <https://edu-sharing.com/> (letzter Zugriff 8.1.2020).
- MEYER, T. & BRUNN, A., 2016: Digitale Photogrammetrie und Bildverarbeitung für kleine und mittlere Unternehmen durch „Blended Learning“. Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V., Band 25, 436-448.
- OSGUTHORPE, R. T. & GRAHAM, C. R., 2003: Blended Learning Environments: Definitions and Directions. *Quarterly Review of Distance Education* 4, 277.
- RUSTICI SOFTWARE LLC, 2020: SCORM.com HomePage: What is SCORM and How it Works. SCORM.com. <https://scorm.com/> (letzter Zugriff 8.1.2020).
- STALLMAN, VON R., 2020: Why Open Source misses the point of Free Software. <https://www.gnu.org/philosophy/open-source-misses-the-point.de.html> (letzter Zugriff 14.1.2020).
- VEREIN ZUR FÖRDERUNG EINES DEUTSCHEN FORSCHUNGSNETZES E. V., 2019: DFNconf: Webkonferenzdienst über Adobe Connect. <https://www.conf.dfn.de/webkonferenzdienst-ueber-adobe-connect/> (letzter Zugriff 14.1.2020).
- VIRTUELLE HOCHSCHULE BAYERN, 2020: SMART vhb: Blended Learning. vhb. <https://www.vhb.org/lehrende/smart-vhb-blended-learning/> (letzter Zugriff 14.1.2020).
- Virtuelle Hochschule Bayern, 2019. Merkblatt SMART vhb.

- WICH, V., BRUNN, A., 2019. Methodik des Blended Learnings im Bereich der Aus- und Weiterbildung der mobilen 3D Vermessung; Publikationen der DGPF, Band 28, 270-281.
- WULF, J., BLOHM, I., LEIMEISTER, J.M. & BRENNER, W., 2014. Massive Open Online Courses. WIRTSCHAFTSINFORMATIK 56. <https://doi.org/10.1007/s11576-014-0405-7>. Springer Fachmedien, Wiesbaden, 127-130.

# Analyse von multi-temporalen Sentinel-2-Daten zur Klassifizierung von Landbedeckung und Landnutzung am Kaiserstuhl

MARTIN WEINMANN<sup>1</sup>, FELIX EHMER<sup>1</sup> & UWE WEIDNER<sup>1</sup>

*Zusammenfassung: Mit der Verfügbarkeit von zeitlich hochauflösenden Daten wie denen des Sentinel-2-Systems ergibt sich eine Verbesserung der Datenlage hinsichtlich einer multi-temporalen Auswertung durch Klassifizierung. Im Rahmen dieser Arbeit wurden Sentinel-2-Daten anhand des Untersuchungsgebietes des Kaiserstuhls evaluiert. Hierzu wurden zunächst die verfügbaren Sentinel-2-Daten der Jahre 2017 und 2018 hinsichtlich einer Landbedeckungs- und Landnutzungsklassifizierung gesichtet und anschließend erfolgte eine Auswahl von geeigneten Zeitpunkten. Die Daten dieser ausgewählten Zeitpunkte wurden aufbereitet und anschließend klassifiziert, wobei die Wichtigkeit von einzelnen Zeitpunkten und/oder einzelnen Spektralbändern genauer untersucht wurde.*

## 1 Einleitung

Die Sentinel-2-Satelliten werden zur systematischen Erfassung der Erdoberfläche genutzt (SPOTO et al. 2012; KUKAWSKA et al. 2017). Sie bieten im Vergleich zu vorherigen Erdbeobachtungssystemen, die entweder durch eine relativ hohe räumliche Auflösung und eine niedrige Aufnahmezeit (z.B. Landsat-7) oder durch eine niedrige räumliche Auflösung und eine hohe Aufnahmezeit (z.B. MODIS) charakterisiert sind, sowohl hinsichtlich der Auflösung als auch bezüglich der Aufnahmezeit eine verbesserte Verfügbarkeit flächendeckender Multispektraldaten. Dies lässt sich gewinnbringend für die Szenenanalyse anhand von Satellitendaten einsetzen. Während die multi-spektrale Analyse einer Szene mit relativ hoher geometrischer und spektraler Auflösung beispielsweise zur genaueren Kartierung und besseren Unterscheidung von diversen Landbedeckungs- und Landnutzungsklassen genutzt werden kann, erlaubt die multi-temporale Analyse mit hohen Aufnahmezeiten eine genauere Beschreibung von Veränderungen und von dynamischen Prozessen. Durch eine Kombination kann jedoch die multi-temporale Analyse auch zur Verbesserung der mit der multi-spektralen Analyse erzielten Ergebnisse genutzt werden. Dies ist insbesondere dadurch bedingt, dass das spektrale Reflektanzverhalten verschiedener Objekte von zahlreichen Einflussfaktoren abhängt (z.B. vom Wettergeschehen, von der Jahreszeit, von Wachstumszyklen oder von der Oberflächenbeschaffenheit von Objekten) und dadurch systematische Änderungen auftreten können. Durch die Betrachtung des spektralen Reflektanzverhaltens über mehrere Zeitpunkte hinweg werden zuverlässigere Aussagen über die jeweilige Klassenzugehörigkeit möglich.

Im Rahmen dieses Beitrags wird die multi-temporale Analyse von multi-spektralen Sentinel-2-Daten genauer beleuchtet. Dabei wird für das Untersuchungsgebiet des Kaiserstuhls anhand von verfügbaren Sentinel-2-Daten der Jahre 2017 und 2018 eine Analyse hinsichtlich diverser Land-

---

<sup>1</sup> Karlsruher Institut für Technologie, Institut für Photogrammetrie und Fernerkundung, Englerstraße 7, D-76131 Karlsruhe, E-Mail: [Martin.Weinmann, Uwe.Weidner]@kit.edu, Felix.Ehmer@student.kit.edu

bedeckungs- und Landnutzungsklassen durchgeführt. Nach der Aufbereitung der Rohdaten im Sinne einer atmosphärischen Korrektur sowie einer Analyse der Wolkenbedeckung und korrespondierender Schatten erfolgt eine Merkmalsextraktion, bei welcher anhand der Reflektanzwerte ausgewählter Spektralkanäle sowohl pro Zeitpunkt als auch für die Kombination bezüglich verschiedener Zeitpunkte Merkmalsvektoren abgeleitet werden. Diese Merkmalsvektoren dienen wiederum als Eingabe in einen Random Forest, der für jeden dieser Fälle eine Klassifizierung hinsichtlich diverser Landbedeckungs- und Landnutzungsklassen durchführt. Als Grundlage für das Training des Random Forest und für die Evaluierung der Qualität der erzielten Klassifizierungsergebnisse werden geeignete Trainings- und Kontrollgebiete herangezogen, welche über eine manuelle Annotation gewonnen wurden. Anhand der Trainingsdaten lässt sich auch auf die Wichtigkeit der einzelnen Spektralkanäle hinsichtlich der betrachteten Klassifizierungsaufgabe schließen, wozu im Rahmen dieser Arbeit verschiedene Verfahren verwendet werden. Die für die Klassifizierung erzielten Ergebnisse zeigen das Potential einer multi-temporalen Analyse deutlich auf, wobei für die betrachtete Szene eine signifikante Steigerung der Gesamtgenauigkeit mit bis zu >10% erreicht wurde. Ferner wird deutlich, dass die Qualität der erzielten Klassifizierungsergebnisse durch das Ignorieren der Reflektanzwerte einzelner Spektralkanäle kaum beeinträchtigt wird.

Nach einer kurzen Zusammenfassung des aktuellen Stands der Technik (Kapitel 2) folgt eine detaillierte Erläuterung der verwendeten Methodik zur Analyse von multi-temporalen Sentinel-2-Daten hinsichtlich diverser Landbedeckungs- und Landnutzungsklassen (Kapitel 3). Daraufhin werden die mit dieser Methodik für ein Gebiet um den Kaiserstuhl erzielten Ergebnisse präsentiert (Kapitel 4). Abschließend folgen eine Zusammenfassung sowie ein Ausblick auf zukünftige Arbeiten (Kapitel 5).

## 2 Stand der Technik

Prinzipiell kann die Klassifizierung von Sentinel-2-Daten hinsichtlich diverser Landbedeckungs- und Landnutzungsklassen in Analogie zu Standardverfahren der Bildauswertung erfolgen. Wurden Sentinel-2-Daten für einen bestimmten Zeitpunkt erfasst, können sie als Bildstapel interpretiert werden und die Reflektanzwerte der einzelnen Spektralkanäle pro Pixel als Einträge von Merkmalsvektoren definiert werden. Ferner können auch spektrale Indizes wie z.B. der Normalized Difference Vegetation Index (NDVI) oder der Normalized Difference Water Index (NDWI) für die Definition von Merkmalsvektoren genutzt werden (VALERO et al. 2016) sowie eine bestimmte Auswahl von relevanten Kanälen (ZHANG et al. 2017; WEINMANN & WEIDNER 2018; WEINMANN & WEIDNER 2019). Die entsprechend definierten Merkmalsvektoren können wiederum als Eingangswerte für diverse Klassifizierungsverfahren genutzt werden, wobei in der Regel Standardklassifizierungsverfahren wie eine Support Vector Machine (CORTES & VAPNIK 1995) oder ein Random Forest (BREIMAN 2001) genutzt werden, welche mittlerweile auch in vielen Softwarepaketen vorhanden sind. Mittlerweile werden aber auch oftmals leistungsfähigere Deep-Learning-Ansätze zur Klassifizierung von Sentinel-2-Daten hinsichtlich diverser Landbedeckungs- und Landnutzungsklassen genutzt (ZHU et al. 2017). Da derartige Verfahren darauf basieren, eine komplexe Optimierungsaufgabe unter Verwendung vieler Millionen Parameter durchzuführen, wird in der Regel eine große Menge an repräsentativen Trainingsdaten benötigt

bis eine angemessene Generalisierungsfähigkeit auf neue Daten erreicht wird. Um eine aufwändige manuelle Annotation zu vermeiden, können beispielsweise auch OpenStreetMap-Daten für das Training einer Standard-CNN-Architektur genutzt werden (LIEBEL 2018).

Um zusätzlich den Einfluss verschiedener Jahreszeiten und den Einfluss von Wachstumszyklen sowie der damit jeweils verbundenen Änderung des spektralen Reflektanzverhaltens vieler Objekttypen (insbesondere bezüglich diverser Vegetationsklassen) bei der Klassifizierung zu berücksichtigen, können beispielsweise für jedes Pixel die Reflektanzwerte der einzelnen Spektralkanäle pro Zeitpunkt zur Definition von Merkmalsvektoren genutzt werden. Die für die kombinierte Betrachtung von Zeitpunkten resultierenden Merkmalsvektoren können dann wiederum mit Standardklassifizierungsverfahren wie z.B. einer Support Vector Machine oder einem Random Forest den jeweiligen Klassen zugeordnet werden (GÓMEZ et al. 2016). Dies lässt sich auch direkt auf die Schätzung der Verteilung von Baumarten in bewaldeten Ökosystemen anhand von multi-temporalen Sentinel-2-Daten übertragen (KARASIAK et al. 2017). Im Gegensatz zu derartigen Standardverfahren lässt sich eine multi-temporale Analyse von Sentinel-2-Daten auch über eine Rekonstruktion der Zeitreihe realisieren, was beispielsweise zur genaueren Analyse von landwirtschaftlichen Nutzflächen genutzt werden kann (BRUZZONE et al. 2017). Hierbei kann mittels eines nicht-parametrischen Regressionsmodells in Form eines Multi-Layer Perceptrons eine Sentinel-2-Zeitreihe in Tagesschritten rekonstruiert und als Eingabe in eine Support Vector Machine gegeben werden. Neben der spektralen und temporalen Information in den Daten können auch räumliche Informationen in den Satellitenbildern mittels Zusammenhangskomponenten einbezogen werden (SOLANO-CORREA et al. 2017), was die Separierung von einzelnen Feldern und deren Analyse mit einer hohen zeitlichen Auflösung ermöglicht. Alternativ kann die Modellierung von dynamischen Prozessen auch über die Nutzung von Hidden Markov Models (HMMs) (SIACHALOU et al. 2015) oder über eine Erweiterung eines Conditional Random Fields (CRFs) durch den zusätzlichen Einbezug temporaler Interaktionen (HOBERG et al. 2015) erfolgen. Ferner ist auch eine Anwendung von modernen Deep-Learning-Ansätzen für die Klassifizierung von multi-temporalen Sentinel-2-Daten möglich. Diesbezüglich bietet sich insbesondere die Nutzung von „Long Short-Term Memory (LSTM)“ Netzwerken (HOCHREITER & SCHMIDHUBER 1997) an, deren wesentliche Idee im Gegensatz zu herkömmlichen rekurrenten neuronalen Netzen auf dem Einbringen eines „langen Kurzzeitgedächtnisses“ basiert. Die Nutzung derartiger Ansätze zur Landbedeckungs- und Landnutzungsklassifizierung ermöglicht eine Berücksichtigung von zeitlichen Charakteristika und führt dadurch im Vergleich zu Verfahren, bei denen lediglich Daten einzelner Zeitpunkte genutzt werden, zu besseren Klassifizierungsergebnissen (RUBWURM & KÖRNER 2017), erfordert jedoch wiederum eine große Menge an annotierten Daten als Grundlage für das Training.

### 3 Methodik

Als Eingangsdaten erhält die verwendete Methodik Rohdaten in Form von multi-temporalen Sentinel-2-Daten. Zunächst erfolgt dann eine Datenvorverarbeitung, um aus den verfügbaren Rohdaten über eine geeignete Korrektur auf die entsprechenden Reflektanzwerte der einzelnen Spektralkanäle zu schließen (Kapitel 3.1). Auf der Basis dieser Reflektanzwerte werden anschließend Merkmalsvektoren definiert (Kapitel 3.2), welche wiederum als Eingangsgrößen für

eine Klassifizierung dienen (Kapitel 3.3). Als Ergebnis liefert die verwendete Methodik schließlich eine thematische Karte, welche die Klassenzugehörigkeiten einzelner Pixel des betrachteten Satellitenbildes widerspiegelt. In einigen Fällen bietet es sich an, die Klassifizierung lediglich auf der Grundlage einer bestimmten Auswahl an Spektralkanälen durchzuführen und dabei diese Auswahl anhand von automatisierten Verfahren zu treffen (Kapitel 3.4).

### 3.1 Datenvorverarbeitung

Zunächst werden die Reflektanzwerte, die zu den gegebenen Spektralkanälen korrespondieren und später zur Definition von Merkmalsvektoren verwendet werden, aus den verfügbaren Rohdaten über eine atmosphärische Korrektur mittels „Sen2Cor“ (<https://step.esa.int/main/third-party-plugins-2/sen2cor/>) ermittelt, wobei verschiedene Parameter eingestellt werden können. Im Rahmen dieser Arbeit werden Effekte von Cirrus-Wolken korrigiert, ein digitales Höhenmodell eingebunden und die bidirektionale Reflexionsverteilungsfunktion (BRDF, von engl. „Bidirectional Reflectance Distribution Function“) berücksichtigt. Zudem wird die Bodenauflösung („Ground Sampling Distance“) für alle Bänder auf eine Auflösung von 20 m angepasst. Um bei der Klassifizierung Effekte durch Wolkenbedeckung oder den Schatten von Wolken zu vermeiden, wird abschließend noch über alle verfügbaren Zeitpunkte geprüft, inwiefern für die entsprechenden Datenpunkte mindestens einmal eine derartige Störung vorliegt (EHMER 2019). Das Verschneiden dieser Information über alle betrachteten Zeitpunkte liefert eine Wolken- und Schattenmaske, über die entsprechende Datenpunkte für alle nachfolgenden Datenverarbeitungsschritte verworfen werden können.

### 3.2 Merkmalsextraktion

Basierend auf den abgeleiteten Reflektanzwerten erfolgt pro Zeitpunkt eine Auswahl von bestimmten Spektralkanälen für die weiteren Betrachtungen (WEINMANN et al. 2018). Hierbei werden nur die Reflektanzwerte ausgewählt, welche zu den Spektralkanälen B2 (zentrale Wellenlänge: 490 nm), B3 (560 nm), B4 (665 nm), B5 (705 nm), B6 (740 nm), B7 (783 nm), B8a (865 nm), B11 (1610 nm) und B12 (2190 nm) korrespondieren. Reflektanzwerte, welche zu den Spektralkanälen B1 (443 nm), B9 (945 nm) und B10 (1375 nm) korrespondieren, werden nicht berücksichtigt, da sie Bereiche im elektromagnetischen Spektrum abdecken, in denen die atmosphärische Durchlässigkeit nur gering ist oder der Wasserdampf das Verhalten dominiert. Der Spektralkanal B8 (842 nm) wird nicht berücksichtigt, da er mit dem Spektralkanal B8a überlappt, aber deutlich breiter und daher weniger aussagekräftig ist (siehe Abb. 1). Die Reflektanzwerte der ausgewählten Spektralkanäle werden weiterhin zur Definition von Merkmalsvektoren genutzt, wobei eine Konkatenation dieser Reflektanzwerte für jeden Zeitpunkt erfolgt. Liegen entsprechende Reflektanzwerte für mehrere Zeitpunkte vor, erfolgt eine Konkatenation unter Berücksichtigung aller Zeitpunkte.

### 3.3 Klassifizierung

Die definierten Merkmalsvektoren dienen wiederum als Eingangswerte für ein Klassifizierungsverfahren, welches es erlaubt, die jeweiligen Klassenzugehörigkeiten zu ermitteln. Im Rahmen

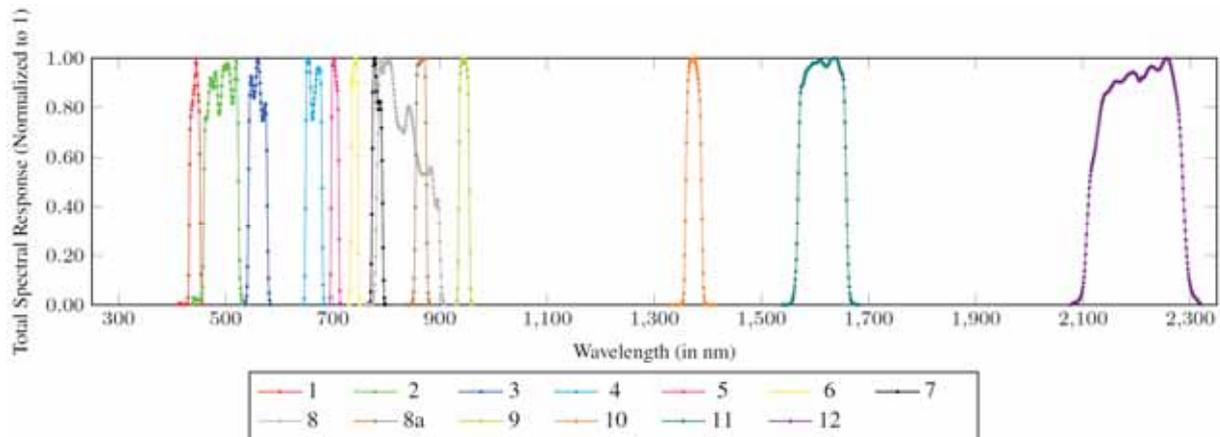


Abb. 1: Eigenschaften der einzelnen Spektralkanäle bei multi-spektralen Sentinel-2-Daten (Weinmann et al. 2018): die Darstellung zeigt die sog. „Sentinel-2 Spectral Response Functions (S2-SRFs)“

dieser Arbeit wird für die Klassifizierung ein überwachtes Lernverfahren in Form eines Random Forest (BREIMAN 2001) verwendet.

In der Trainingsphase wird anhand von Trainingsbeispielen unter Verwendung eines Bootstrap Aggregating („Bagging“) ein Ensemble von randomisierten Entscheidungsbäumen erzeugt, d.h. jeder Entscheidungsbaum wird auf einer zufällig anderen Untermenge der Trainingsdaten trainiert. Beim Training jedes Entscheidungsbaumes erfolgt eine sukzessive Aufteilung der jeweils betrachteten Trainingsdaten in immer kleinere Untermengen, so dass jede resultierende Untermenge „so rein wie möglich“ ist. Diese Aufteilung wird in Form einer Sequenz von Entscheidungen formalisiert.

In der Testphase, d.h. bei der Betrachtung der zu klassifizierenden Merkmalsvektoren, wird dann bei jedem Entscheidungsbaum die entsprechende, durch das Training festgelegte Sequenz von Entscheidungen durchlaufen, bis ein Blatt des Entscheidungsbaumes mit einer entsprechenden Hypothese bezüglich einer Klassenzugehörigkeit erreicht wird. Eine Betrachtung der Hypothesen („Votes“) über alle Entscheidungsbäume führt zu Klassenwahrscheinlichkeiten und der Mehrheitsentscheid („Majority Vote“) liefert schließlich eine robuste Hypothese für eine geeignete Zuweisung einer Klassenzugehörigkeit.

### 3.4 Selektion von relevanten Spektralkanälen

Um als Grundlage für die Klassifizierung eine sinnvolle Auswahl von Spektralkanälen zu treffen, können verschiedene Strategien der Merkmalsselektion herangezogen werden (GUYON & ELISSEFF 2003; SAEYS et al. 2007; ZHAO et al. 2010; WEINMANN & WEIDNER 2019). Am Beispiel der Analyse von multi- und hyper-spektralen Daten lassen sich relevante Spektralkanäle durch 1) eine intensive Interaktion mit einem Klassifizierungsverfahren, 2) interne Metriken eines Klassifizierungsverfahrens (welche während des Trainings ermittelt werden) oder 3) unabhängig von Klassifizierungsverfahren durch aufgestellte Relevanzmetriken ableiten (WEINMANN & WEIDNER 2019). Im Rahmen dieser Arbeit wurde aus Effizienzgründen ein Verfahren aus der letzteren Kategorie verwendet, welches auf der kombinierten Betrachtung verschiedener Relevanzmetriken (wie z.B. Distanz-, Informations-, Abhängigkeits- oder Korrelationsmetriken) basiert und somit eine Abschätzung der Wichtigkeit einzelner Spektralkanäle hinsichtlich der be-

trachteten Klassifizierungsaufgabe erlaubt (WEINMANN & WEIDNER 2019). Zudem wird im Random Forest eine interne Metrik herangezogen.

## 4 Ergebnisse

Im Folgenden werden zunächst das Untersuchungsgebiet und die verwendete Datengrundlage beschrieben (Kapitel 4.1). Anschließend wird die Bestimmung von Trainings- und Kontrollgebieten als Grundlage für das Training von Klassifizierungsverfahren und die Evaluierung von Klassifizierungsergebnissen erläutert (Kapitel 4.2). Daraufhin werden die zur Bewertung der Leistungsfähigkeit der vorgestellten Methodik verwendeten Evaluierungs-metriken beschrieben (Kapitel 4.3). Abschließend erfolgt die Präsentation der erzielten Ergebnisse (Kapitel 4.4).



Abb. 2: Falschfarbendarstellung des Untersuchungsgebietes um den Kaiserstuhl für den Zeitpunkt des 19.06.2017 unter Verwendung der Kanalrepräsentation B4-B3-B2 der entsprechenden multispektralen Sentinel-2-Daten (links) sowie die aus 12 verschiedenen Zeitpunkten abgeleitete Wolken- und Schattenmaske (rechts)

### 4.1 Untersuchungsgebiet

Als Untersuchungsgebiet wurde die Region des Kaiserstuhls gewählt (siehe Abb. 2, links). Der Kaiserstuhl und sein Umland zeichnen sich durch seine vielfältige Vegetation aus. Besonders das Weinanbaugebiet am Kaiserstuhl sowie die hohe landwirtschaftliche Nutzung im Umland prägen diese Region. Aber auch der Rhein westlich des Kaiserstuhls und die vielen Siedlungen haben

ihren Anteil an der facettenreichen Landbedeckung. Insgesamt weist das Untersuchungsgebiet Höhendifferenzen von bis zu knapp 400 m auf.

Für die durchgeführten Untersuchungen wurden Sentinel-2-Daten für 12 verschiedene Zeitpunkte herangezogen, von denen sechs im Jahr 2017 (10.04., 10.05., 19.06., 19.07., 23.08., 22.09.) und sechs im Jahr 2018 (20.04., 05.05., 09.07., 24.07., 03.08., 12.09.) liegen und jeweils die wesentlichen Wachstumszyklen abdecken. Anhand dieser Daten wurde eine Wolken- und Schattenmaske erstellt (siehe Abb. 2, rechts), über die entsprechend fehlerbehaftete Datenpunkte für alle nachfolgenden Datenverarbeitungsschritte verworfen werden.

## 4.2 Auswahl von Trainings- und Kontrollgebieten

Als Grundlage für das Training eines Klassifizierungsverfahrens und die Evaluierung der Qualität der erzielten Klassifizierungsergebnisse diente eine hinreichende Anzahl von repräsentativen Trainings- und Kontrollgebieten, welche den Klassen *Siedlung*, *Industrie*, *Wasser*, *Wald*, *Wein*, *Wiese* und *Feld* zugeordnet werden können. Diese Trainings- und Kontrollgebiete wurden aus den vorliegenden Satellitenbildern über eine manuelle Annotation von Polygonen mit möglichst homogenen Charakteristika der eingeschlossenen Pixel abgeleitet. Zunächst wurden 29 Trainingsgebiete ausgewählt, welche den definierten Klassen entsprechen. Diese Trainingsgebiete umfassen insgesamt 6298 Pixel und entsprechen damit einer Fläche von 251,92 ha. Anhand dieser Trainingsgebiete mit bekannten Klassenzugehörigkeiten wurde pro Pixel ein Trainingsbeispiel bestehend aus dem entsprechenden Merkmalsvektor und der korrespondierenden Klassenzugehörigkeit definiert. Um eine quantitative Evaluierung der vorgestellten Methodik zu ermöglichen, wurden zudem 46 Kontrollgebiete definiert, welche insgesamt 11514 Pixel und damit eine Fläche von 460,56 ha abdecken. Anhand der Kontrollgebiete mit bekannten Klassenzugehörigkeiten werden wiederum Kontrollbeispiele definiert, die aus dem entsprechenden Merkmalsvektor und der korrespondierenden Klassenzugehörigkeit bestehen. Die entsprechenden Verteilungen über die einzelnen Klassen sind in Tab. 1 und Tab. 2 angegeben.

Tab. 1: Verteilung der Trainingsbeispiele über die definierten Klassen.

Klasse	# Beispiele	Fläche [ha]	Anteil [%]
Siedlung	999	39,96	15,86
Industrie	854	34,16	13,56
Wasser	975	39,00	15,48
Wald	929	37,16	14,75
Wein	898	35,92	14,26
Wiese	604	24,16	9,59
Feld	1039	41,56	16,50
$\Sigma$	6298	251,92	100,0

Tab. 2: Verteilung der Kontrollbeispiele über die definierten Klassen.

Klasse	# Beispiele	Fläche [ha]	Anteil [%]
--------	-------------	-------------	------------

Siedlung	1582	63,28	13,74
Industrie	1475	59,00	12,81
Wasser	1305	52,20	11,33
Wald	1936	77,44	16,81
Wein	2226	89,04	19,33
Wiese	1000	40,00	8,69
Feld	1990	79,60	17,28
$\Sigma$	11514	460,56	100,0

### 4.3 Evaluierungsmetriken

Neben der qualitativen Bewertung der erzielten Klassifizierungsergebnisse in Form von thematischen Karten wurde auch eine quantitative Bewertung durchgeführt, bei welcher die Gesamtgenauigkeit (OA) und der  $\kappa$ -Index sowie die klassenweisen Metriken des Recall-Werts und des Precision-Werts herangezogen wurden. Aus den letzteren beiden wurde für jede Klasse der entsprechende F<sub>1</sub>-Wert abgeleitet und zur Bewertung über alle Klassen der Mittelwert gebildet (mF<sub>1</sub>).

### 4.4 Ergebnisse

Im Rahmen dieser Arbeit wurden Sentinel-2-Daten, welche für das Gebiet des Kaiserstuhls für verschiedene Zeitpunkte vorliegen, sowohl pro Zeitpunkt als auch für die Kombination bezüglich verschiedener Zeitpunkte analysiert. Dabei erfolgte eine Klassifizierung für die gesamte verfügbare Szene sowie eine quantitative Bewertung der erzielten Ergebnisse anhand der definierten Kontrollgebiete.

Die unter Berücksichtigung einzelner Zeitpunkte erzielten Klassifizierungsergebnisse sind in Tab. 3 zusammengefasst und zeigen starke Variationen bezüglich der betrachteten Evaluierungsmetriken. Der generelle Trend zeigt mit einer OA von 76-84% relativ schlechte Klassifizierungsergebnisse für den ersten Zeitpunkt im April jedes betrachteten Jahres sowie für den letzten Zeitpunkt im September jeden betrachteten Jahres, während bei den Zeitpunkten dazwischen pro Jahr zumeist signifikant bessere Klassifizierungsergebnisse mit einer OA von 88-91% vorliegen. Die jeweils besten Ergebnisse pro Jahr sind in Abb. 3 dargestellt.

Die erzielten Ergebnisse für die Kombination von Sentinel-2-Daten verschiedener Zeitpunkte als Grundlage für die Klassifizierung sind in Abb. 4, Abb. 5 und Tab. 4 angegeben. Die Kombination der Sentinel-2-Daten für die jeweils sechs verschiedenen Zeitpunkte pro Jahr liefert mit einer OA von 93,56% und 94,36% eine signifikante Verbesserung der Klassifizierungsergebnisse. Demgegenüber ist mit einer OA von 94,90% lediglich eine leichte Verbesserung gegeben, wenn die Sentinel-2-Daten für alle 12 verschiedenen Zeitpunkte als Grundlage für die Klassifizierung genutzt werden.

Tab. 3: Klassifizierungsergebnisse unter Berücksichtigung der Sentinel-2-Daten von einzelnen Zeitpunkten in den Jahren 2017 und 2018.

Zeitpunkte	OA [%]	$\kappa$ [%]	mF <sub>1</sub> [%]
10.04.2017	82,35	78,65	80,85
10.05.2017	87,68	84,99	85,07
19.06.2017	89,19	86,78	87,81
19.07.2017	80,36	76,01	80,38
23.08.2017	88,00	85,29	84,24
22.09.2017	80,15	75,89	79,05
20.04.2018	76,25	71,51	76,48
05.05.2018	85,39	82,27	83,71
09.07.2018	89,55	87,24	87,49
24.07.2018	87,84	85,12	87,44
03.08.2018	90,97	88,92	89,59
12.09.2018	83,61	80,10	81,86

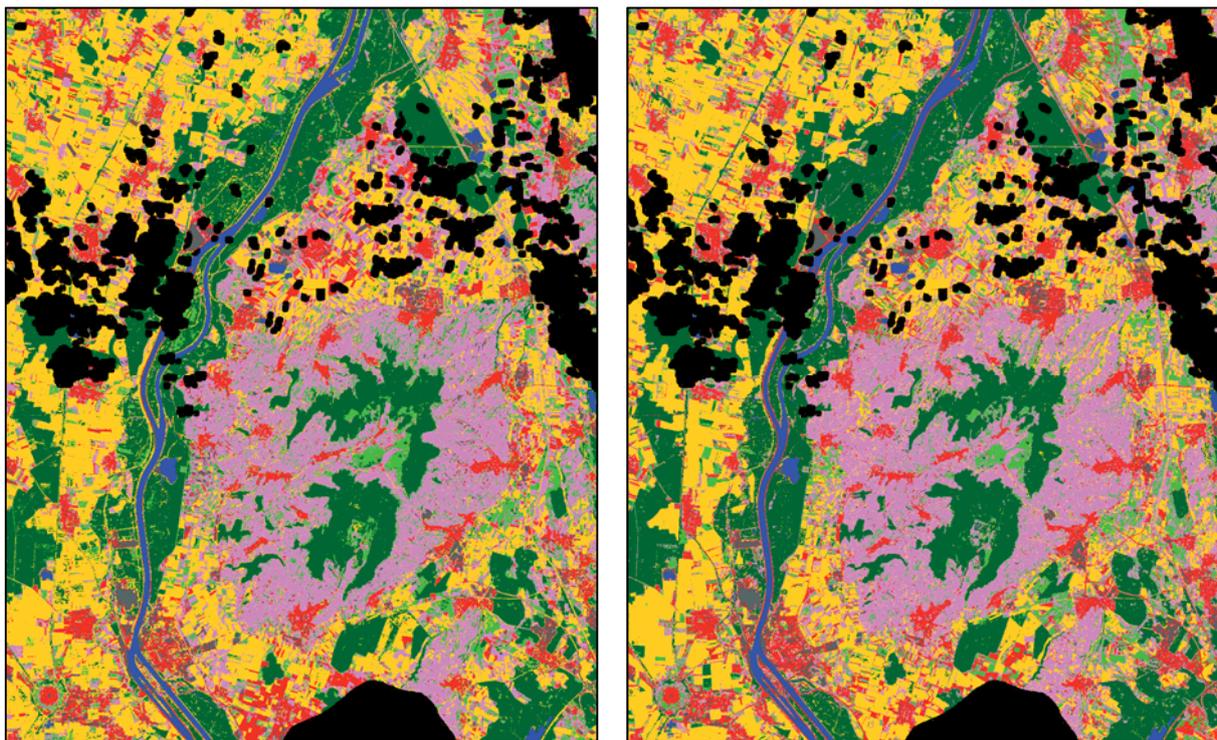


Abb. 3: Erzielte Klassifizierungsergebnisse unter Berücksichtigung der Sentinel-2-Daten für die Zeitpunkte des 19.06.2017 (links) und des 03.08.2018 (rechts): die Farben entsprechen den Klassen *Siedlung* (in Rot), *Industrie* (in Grau), *Wasser* (in Blau), *Wald* (in Dunkelgrün), *Wein* (in Rosa), *Wiese* (in Hellgrün) und *Feld* (in Gelb). Für die schwarzen Bereiche liegen aufgrund der Verwendung einer Wolken- und Schattenmaske keine Ergebnisse vor.

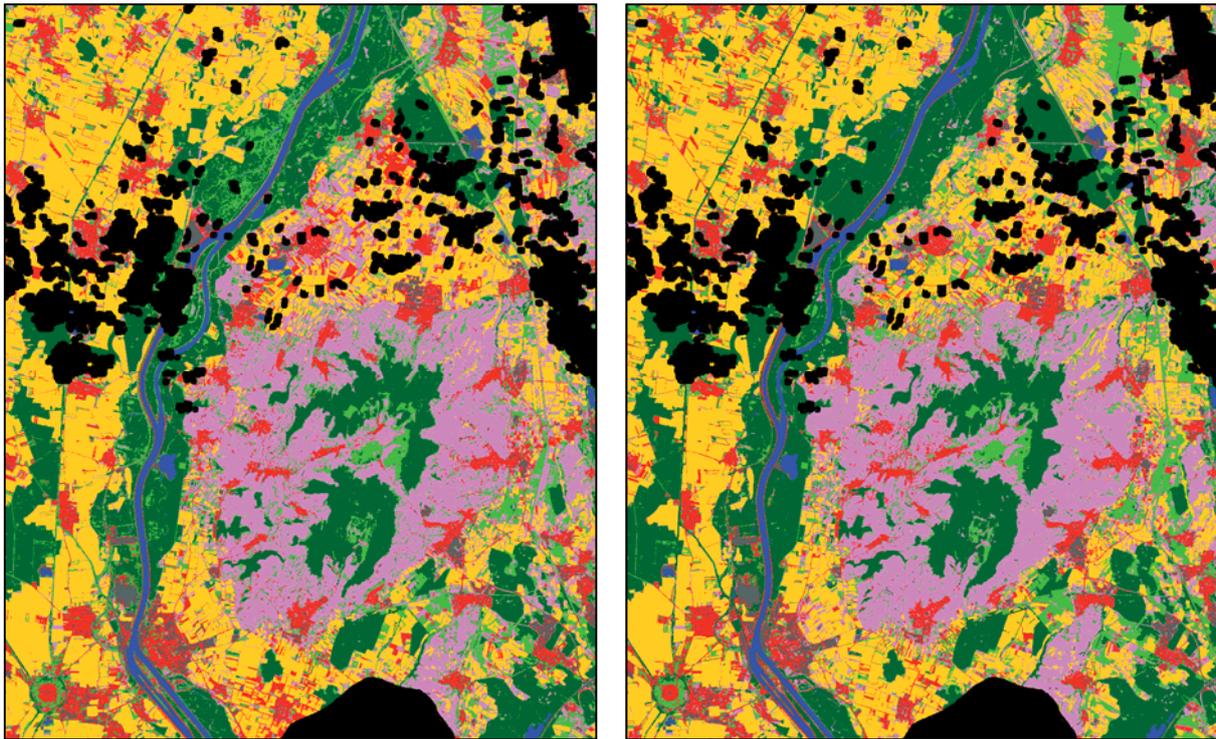


Abb. 4: Klassifizierungsergebnisse für die Jahre 2017 (links) und 2018 (rechts) unter Berücksichtigung der Sentinel-2-Daten für jeweils sechs Zeitpunkte: die Farben entsprechen den Klassen *Siedlung* (in Rot), *Industrie* (in Grau), *Wasser* (in Blau), *Wald* (in Dunkelgrün), *Wein* (in Rosa), *Wiese* (in Hellgrün) und *Feld* (in Gelb). Für die schwarzen Bereiche liegen aufgrund der Verwendung einer Wolken- und Schattenmaske keine Ergebnisse vor

Tab. 4: Klassifizierungsergebnisse unter Berücksichtigung der Sentinel-2-Daten für jeweils sechs verschiedene Zeitpunkte pro Jahr sowie für die Kombination aller Zeitpunkte.

Zeitpunkte	OA [%]	$\kappa$ [%]	mF <sub>1</sub> [%]
alle in 2017	93,56	92,11	91,47
alle in 2018	94,36	93,09	92,16
alle in 2017 und 2018	94,90	93,75	92,66

Die erzielten Ergebnisse für die Kombination der Sentinel-2-Daten für jeweils drei verschiedene Zeitpunkte pro Jahr sind in Tab. 5 angegeben. Dabei wird ersichtlich, dass sowohl für das Jahr 2017 als auch für das Jahr 2018 die Berücksichtigung der ersten drei Zeitpunkte zu deutlich besseren Klassifizierungsergebnissen führt als die Berücksichtigung der letzten drei Zeitpunkte. Der Unterschied beträgt hier bezüglich der OA ca. 2-4 %. Insgesamt liegt jedoch eine geringere Qualität der Klassifizierungsergebnisse vor im Vergleich zu den Ergebnissen für die Kombination der Sentinel-2-Daten für die jeweils sechs verschiedenen Zeitpunkte pro Jahr und für die Kombination der Sentinel-2-Daten für alle 12 verfügbaren Zeitpunkte.

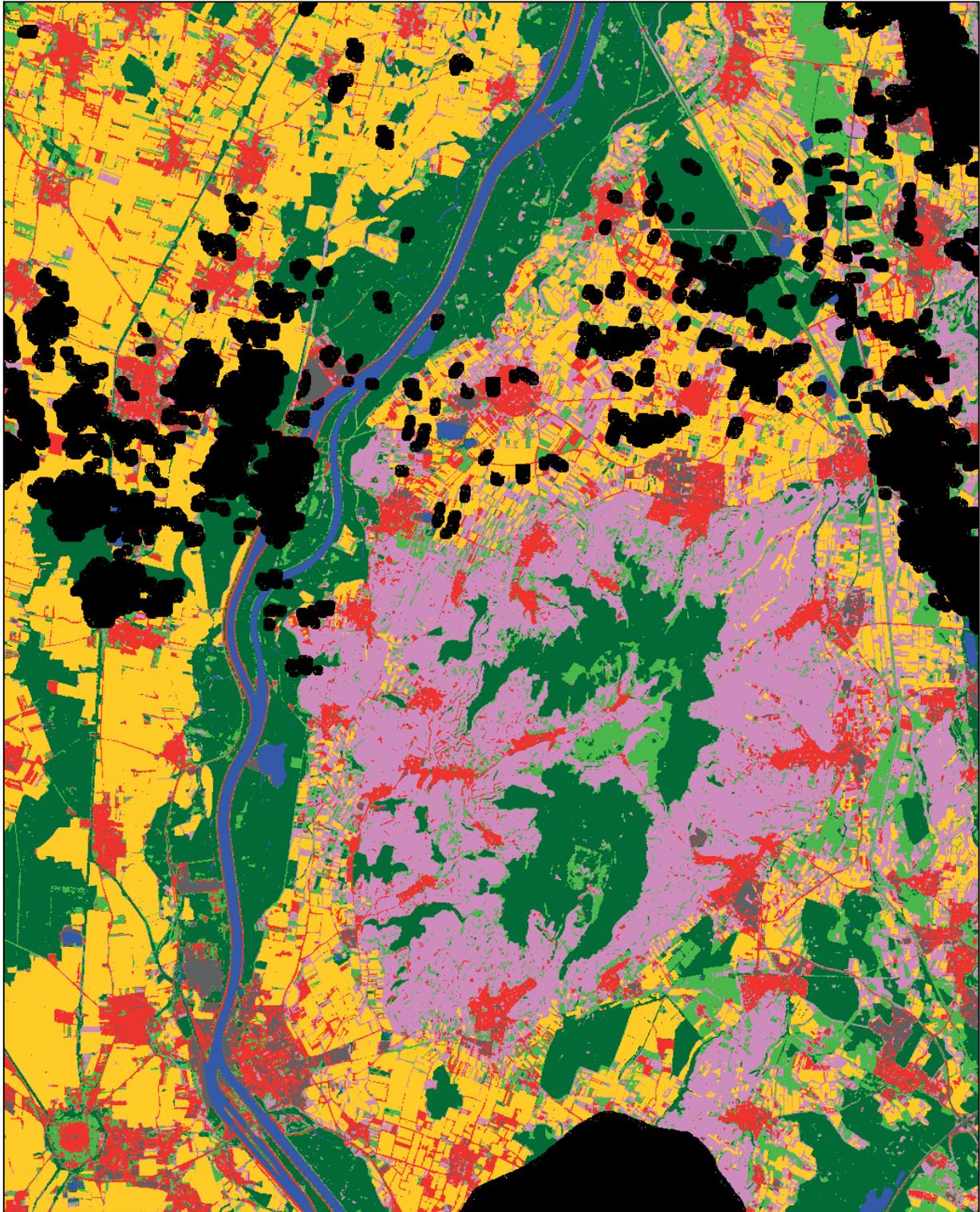


Abb. 5: Klassifizierungsergebnisse unter Berücksichtigung der Sentinel-2-Daten für alle 12 verfügbaren Zeitpunkte: die Farben entsprechen den Klassen *Siedlung* (in Rot), *Industrie* (in Grau), *Wasser* (in Blau), *Wald* (in Dunkelgrün), *Wein* (in Rosa), *Wiese* (in Hellgrün) und *Feld* (in Gelb).

Tab. 5: Klassifizierungsergebnisse unter Berücksichtigung der Sentinel-2-Daten für jeweils drei verschiedene Zeitpunkte pro Jahr.

<b>Zeitpunkte in 2017</b>	<b>OA [%]</b>	<b><math>\kappa</math> [%]</b>	<b>mF<sub>1</sub> [%]</b>
10.04., 10.05., 19.06.	92,43	90,75	90,36
19.07., 23.08., 22.09.	88,48	85,96	85,40

<b>Zeitpunkte in 2018</b>	<b>OA [%]</b>	<b><math>\kappa</math> [%]</b>	<b>mF<sub>1</sub> [%]</b>
20.04., 05.05., 09.07.	92,55	90,89	90,34
24.07., 03.08., 12.09.	90,44	88,31	88,59

Unter Verwendung der abgeleiteten Relevanzmetrik sowie der internen Metrik des Random Forest scheinen insbesondere die Spektralkanäle B2, B3, B4 und B5 als relevant bezüglich der betrachteten Klassifizierungsaufgabe. Werden daher zunächst die Reflektanzwerte der Spektralkanäle B11 und B12 während der Klassifizierung ignoriert und die Reflektanzwerte der Spektralkanäle B2, B3, B4, B5, B6, B7 und B8a herangezogen, zeigen die Klassifizierungsergebnisse eine etwas geringere Qualität bei einem Verlust von etwa 1-2% bezüglich der OA im Vergleich zu den für die Nutzung aller Spektralkanäle erzielten Ergebnissen (siehe Tab. 6). Werden demgegenüber die Reflektanzwerte der Spektralkanäle B6, B7 und B8a während der Klassifizierung ignoriert und die Reflektanzwerte der Spektralkanäle B2, B3, B4, B5, B11 und B12 herangezogen, beträgt der Verlust bezüglich der OA unter 1% im Vergleich zu den für die Nutzung aller Spektralkanäle erzielten Ergebnissen (siehe Tab. 7). Werden schließlich die Reflektanzwerte der Spektralkanäle B6, B7, B8a, B11 und B12 während der Klassifizierung ignoriert und lediglich die Reflektanzwerte der Spektralkanäle B2, B3, B4 und B5 herangezogen, beträgt der Verlust bezüglich der OA wiederum etwa 1-2% im Vergleich zu den für die Nutzung aller Spektralkanäle erzielten Ergebnissen (siehe Tab. 8).

Tab. 6: Klassifizierungsergebnisse unter Berücksichtigung der Sentinel-2-Daten für jeweils sechs verschiedene Zeitpunkte pro Jahr sowie für die Kombination aller Zeitpunkte: die pro Zeitpunkt verwendeten Spektralkanäle sind durch B2, B3, B4, B5, B6, B7 und B8a gegeben.

<b>Zeitpunkte</b>	<b>OA [%]</b>	<b><math>\kappa</math> [%]</b>	<b>mF<sub>1</sub> [%]</b>
alle in 2017	92,26	90,53	89,93
alle in 2018	93,14	91,60	90,91
alle in 2017 und 2018	93,68	92,26	91,35

Tab. 7: Klassifizierungsergebnisse unter Berücksichtigung der Sentinel-2-Daten für jeweils sechs verschiedene Zeitpunkte pro Jahr sowie für die Kombination aller Zeitpunkte: die pro Zeitpunkt verwendeten Spektralkanäle sind durch B2, B3, B4, B5, B11 und B12 gegeben.

<b>Zeitpunkte</b>	<b>OA [%]</b>	<b><math>\kappa</math> [%]</b>	<b>mF<sub>1</sub> [%]</b>
alle in 2017	93,43	91,96	91,64
alle in 2018	93,55	92,09	91,82
alle in 2017 und 2018	93,98	92,62	92,47

Tab. 8: Klassifizierungsergebnisse unter Berücksichtigung der Sentinel-2-Daten für jeweils sechs verschiedene Zeitpunkte pro Jahr sowie für die Kombination aller Zeitpunkte: die pro Zeitpunkt verwendeten Spektralkanäle sind durch B2, B3, B4 und B5 gegeben.

Zeitpunkte	OA [%]	$\kappa$ [%]	mF <sub>1</sub> [%]
alle in 2017	92,95	91,37	91,21
alle in 2018	92,42	90,71	90,39
alle in 2017 und 2018	92,83	91,23	90,88

## 5 Zusammenfassung und Ausblick

In diesem Beitrag wurde die Analyse von multi-temporalen Sentinel-2-Daten hinsichtlich diverser Landbedeckungs- und Landnutzungsklassen behandelt. Dabei wurden für das Untersuchungsgebiet des Kaiserstuhls anhand von verfügbaren Sentinel-2-Daten der Jahre 2017 und 2018 über die Reflektanzwerte ausgewählter Spektralkanäle sowohl pro Zeitpunkt als auch für die Kombination bezüglich verschiedener Zeitpunkte Merkmalsvektoren abgeleitet und jeweils als Eingabe für einen Random Forest verwendet. Letzterer lieferte für die resultierenden Fälle eine Klassifizierung hinsichtlich der definierten Landbedeckungs- und Landnutzungs-klassen. Die vergleichenden Betrachtungen der erzielten Klassifizierungsergebnisse zeigten das große Potential einer multi-temporalen Analyse deutlich auf, wobei für das betrachtete Gebiet durch die Kombination der Reflektanzwerte ausgewählter Spektralkanäle bezüglich verschiedener Zeitpunkte im Vergleich zur Betrachtung der Reflektanzwerte ausgewählter Spektralkanäle pro Zeitpunkt eine signifikante Steigerung der Gesamtgenauigkeit mit bis zu >10% erreicht wurde. Ferner konnte gezeigt werden, dass die Qualität der erzielten Klassifizierungsergebnisse insbesondere durch die Information der Reflektanzwerte in den Spektralkanälen B2, B3, B4 und B5 bedingt ist und durch das Ignorieren der Reflektanzwerte anderer Spektralkanäle kaum beeinträchtigt wird.

Im Rahmen zukünftiger Arbeiten ist eine zusätzliche Betrachtung von klassenweisen Evaluierungsmetriken geplant, da dies eine detailliertere Analyse des Verhaltens der erzielten Klassifizierungsergebnisse hinsichtlich einzelner Landbedeckungs- und Landnutzungsklassen ermöglicht. Zudem wird die Erweiterung der Trainingsdaten angestrebt, so dass im Vergleich zu den mit der vorgestellten Methodik für ein Untersuchungsgebiet erzielten Ergebnissen auch die Leistungsfähigkeit von modernen Deep-Learning-Ansätzen analysiert werden kann. Dabei ist auch der Einfluss von fehlerbehafteten Zuordnungen von Klassenzugehörigkeiten bei den zugrundeliegenden Trainingsdaten („Label Noise“) zu untersuchen, da diese in der Regel durch ein effizienteres manuelles Annotieren großer Szenen bedingt sind und oftmals einen nicht vernachlässigbaren Anteil an den Trainingsdaten ausmachen.

## 6 Literaturverzeichnis

- BREIMAN, L., 2001: Random forests. *Machine Learning* **45**(1), 5-32.
- BRUZZONE, L., BOVOLO, F., PARIS, C., SOLANO-CORREA, Y. T., ZANETTI, M. & FERNÁNDEZ-PRIETO, D., 2017: Analysis of multitemporal Sentinel-2 images in the framework of the ESA Scientific Exploitation of Operational Missions. 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), 1-4.
- CORTES, C. & VAPNIK, V., 1995: Support-vector networks. *Machine Learning* **20**(3), 273-297.
- EHMER, F., 2019: Untersuchung zur Nutzung von Sentinel2-Daten für die Region des Kaiserstuhls. Bachelorarbeit, Institut für Photogrammetrie und Fernerkundung, Karlsruher Institut für Technologie.
- GÓMEZ, C., WHITE, J. C. & WULDER, M. A., 2016: Optical remotely sensed time series data for land cover classification: a review. *ISPRS Journal of Photogrammetry and Remote Sensing* **116**, 55-72.
- GUYON, I. & ELISSEEFF, A., 2003: An introduction to variable and feature selection. *Journal of Machine Learning Research* **3**, 1157-1182.
- HOBERG, T., ROTTENSTEINER, F., FEITOSA, R. Q. & HEIPKE, C., 2015: Conditional random fields for multitemporal and multiscale classification of optical satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing* **53**(2), 659-673.
- HOCHREITER, S. & SCHMIDHUBER, J., 1997: Long short-term memory. *Neural Computation*, **9** (8), 1735-1780.
- KARASIAK, N., SHEEREN, D., FAUVEL, M., WILLM, J., DEJOUX, J. F. & MONTEIL, C., 2017: Mapping tree species of forests in southwest France using Sentinel-2 image time series. 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), 1-4.
- KUKAWSKA, E., LEWINSKI, S., KRUPINSKI, M., MALINOWSKI, R., NOWAKOWSKI, A., RYBICKI, M. & KOTARBA, A., 2017: Multitemporal Sentinel-2 data – Remarks and observations. 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), 1-4.
- LIEBEL, L., 2018: Deep convolutional neural networks for semantic segmentation of multispectral Sentinel-2 satellite imagery: an open data approach to large-scale land use and land cover classification. *Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V.*, Band **27**, T. P. Kersten, E. Gülch, J. Schiewe, T. H. Kolbe & U. Stilla (Hrsg.), 38. Wissenschaftlich-Technische Jahrestagung der DGPF, 7.-9. März 2018 in München, Tagungsband auf CD-ROM, 790-793.
- RUBWURM, M. & KÖRNER, M., 2017: Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 1496-1504.
- SAEYS, Y., INZA, I. & LARRANAGA, P., 2007: A review of feature selection techniques in bioinformatics. *Bioinformatics* **23**(19), 2507-2517.
- SIACHALOU, S., MALLINIS, G. & TSAKIRI-STRATI, M., 2015: A hidden Markov models approach for crop classification: linking crop phenology to time series of multi-sensor remote sensing data. *Remote Sensing* **7**(4), 3633-3650.

- SOLANO-CORREA, Y. T., BOVOLO, F., BRUZZONE, L. & FERNÁNDEZ-PRIETO, D., 2017: Spatio-temporal evolution of crop fields in Sentinel-2 satellite image time series. 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), 1-4.
- SPOTO, F., SY, O., LABERINTI, P., MARTIMORT, P., FERNANDEZ, V., COLIN, O., HOERSCH, B. & MEYGRET, A., 2012: Overview of Sentinel-2. IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 1707-1710.
- VALERO, S., MORIN, D., INGLADA, J., SEPULCRE, G., ARIAS, M., HAGOLLE, O., DEDIEU, G., BONTEMPS, S., DEFOURNY, P. & KOETZ, B., 2016: Production of a dynamic cropland mask by processing remote sensing image series at high temporal and spatial resolutions. Remote Sensing **8**(1), 55:1-55:21.
- WEINMANN, M., MAIER, P. M., FLORATH, J. & WEIDNER, U., 2018: Investigations on the potential of hyperspectral and Sentinel-2 data for land-cover / land-use classification. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences **IV-1**, 155-162.
- WEINMANN, M. & WEIDNER, U., 2018: Erste Erfahrungen mit der Nutzung von Sentinel-2 Daten zur multispektralen und multitemporalen Analyse mit verschiedenen Klassifizierungsverfahren. Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V., Band **27**, T. P. Kersten, E. Gülch, J. Schiewe, T. H. Kolbe & U. Stilla (Hrsg.), 38. Wissenschaftlich-Technische Jahrestagung der DGPF, 7.-9. März 2018 in München, Tagungsband auf CD-ROM, 58-70.
- WEINMANN, M. & WEIDNER, U., 2019: Relevance assessment of spectral bands for land cover and land use classification: A case study involving multispectral Sentinel-2-like and hyperspectral data. Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V., Band **28**, T. P. Kersten (Hrsg.), 39. Wissenschaftlich-Technische Jahrestagung der DGPF, 20.-22. Februar 2019 in Wien, Tagungsband auf CD-ROM, 138-153.
- ZHANG, T., SU, J., LIU, C., CHEN, W. H., LIU, H. & LIU, G., 2017: Band selection in Sentinel-2 satellite for agriculture applications. 23rd International Conference on Automation and Computing (ICAC), 1-6.
- ZHAO, Z., MORSTATTER, F., SHARMA, S., ALELYANI, S., ANAND, A. & LIU, H., 2010: Advancing Feature Selection Research – ASU Feature Selection Repository. Technical Report, School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ, USA, 2010.
- ZHU, X. X., TUIA, D., MOU, L., XIA, G., ZHANG, L., XU, F. & FRAUNDORFER, F., 2017: Deep learning in remote sensing: a comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine **5**(4), 8-36.

# Comparison of Deep-Learning Classification Approaches for Indoor Point Clouds

VLADETA STOJANOVIC<sup>1</sup>, MATTHIAS TRAPP<sup>1</sup>, RICO RICHTER<sup>1</sup> & JÜRGEN DÖLLNER<sup>1</sup>

*Abstract: In this paper we present an approach for classifying indoor point clouds - particularly areas of clutter, noise, and missing segments that are often featured in point clouds captured using commodity mobile devices with photogrammetric or depth-sensing capabilities. We describe and evaluate two deep-learning approaches for classification of point clouds: using a 2D or 3D Convolutional Neural Network (CNN), for multiview and object-based classification tasks. We present a case study for classification of furniture items using two different CNN architectures (Inception V3 and PointNet++). We report on the classification accuracy and the practicality of using both approaches, and evaluate an experimental approach for generation 3D floorplans from bounding boxes of segmented point clusters. The experimental results of the case study show that the use of PointNet++ CNN offers superior classification accuracy in comparison to the multiview classification, though multiview classification offers a better alternative in terms of computational requirements and lightweight software component integration.*

## 1 Introduction

The built environment around us is constantly changing, and reflecting this change digitally is a key challenge in research and development for Smart Cities and Real Estate applications. Point clouds allow us to capture and digitally represent the built environment. While point clouds can capture the physical surface of a real world object and its location, they lack associated semantics that are often required for further analysis and decision making. With recent advancements in machine learning (particularly deep learning), new methods and approaches have become available for generating semantics for point cloud data. Convolutional Neural Networks (CNN) allow for a given classification model to predict the semantics of point clouds. Approaches for training CNN-based classification models can either make use of point clouds, geometric approximations of point regions (e.g., voxels), 2D images of specific 3D objects and/or real-life photographs of their counterparts. The main focus of this research is a case study that examines and compares the practical aspects of training and using 2D and 3D CNNs for classification of indoor point clouds.

### 1.1 Problem Statement

Access to training data for indoor point clouds required for training 3D CNNs is usually difficult to acquire, as there is a paucity of freely available datasets. 3D CNNs that make use of unstructured point data as the primary data source for unsupervised learning have versatile applications for semantic classification of the built environment. However, their use is also often restricted by

---

<sup>1</sup> Hasso Plattner Institute, Faculty of Digital Engineering, University of Potsdam, Prof.-Dr.-Helmert-Straße 2-3, D-14482 Potsdam, Email: [vladeta.stojanovic, matthias.trapp, rico.richter, juergen.doellner]@hpi.de

computational hardware requirements that are usually far greater when compared to those of using 2D CNNs for training and classification. In most cases, the use of GPU-accelerated computation is required, and point cloud data used for training can be very large in terms of memory and storage requirements (often Gigabytes in size). An alternative to 3D CNNs is the use of 2D CNNs trained on image data. In such cases, a multiview classification approach can be used to classify images of spatially-partitioned point clusters, and stream the classification results of the associated point cluster from the model classification result outputs.

## 1.2 Research Contributions

For this research we have compared two approaches for detecting common office furniture objects: The first using multiview classification with a retrained version of the Inception V3 2D CNN (SZEGEDY et al. 2016), retrained using real-life photographs of common office furniture, and the second approach using PointNet++ 3D CNN (QI et al. 2017), trained on the publicly available Stanford Large-Scale 3D Indoor Spaces Dataset (S3DIS) (ARMENI et al. 2016), using a semantic segmentation approach for detecting furniture objects. We also focused on the practicality of using such approaches on commodity computer hardware, in order to assess the feasibility of such an approach without access to high-end workstation or cluster computation resources. Finally, we present and discuss a practical approach for generating bounding-box and triangulated 3D geometry from segmented point clouds representing the floorplan of a given indoor area, which can be combined with the classified furniture point clusters to create a semantically-enriched indoor representation model.

## 2 Related Work

Semantic segmentation enables the assignment of points to surface categories (e.g, wall, chair). Furthermore, semantically enriched point clouds can be used as base-data for Building Information Modeling (BIM) and Digital Twin (DT) representations, particularly within the realm of Real Estate 4.0 (STOJANOVIC et al. 2018a). BABACAN et al. (2017) present an approach of semantically enriching indoor point clouds using a combined approach of classification with a 3D CNN and further segmentation of planar point clusters using the RANSAC algorithm. Other recent approaches for semantic segmentation of indoor scenes make combined use of both 2D and 3D CNNs for clustering and semantic labelling of unstructured indoor point clouds (WANG et al. 2019; CHIANG et al. 2019).

The use of 3D CNNs for classification tasks has proven to be a viable option for semantic enrichment of unstructured point clouds. The PointNet++ CNN architecture is able to learn specific features of indoor point clouds in a supervised manner, and apply this model to semantic segmentation of various point cloud models - including point clouds representing the built environment (MALINVERNI et al. 2019), and LIDAR representations of outdoor and urban scenes (WINIWARTER et al. 2019; WOLF et al. 2019). PointNet++ is able to classify unstructured point clouds regardless of their scale, density or orientation, and is invariant to permutations. It uses local distance between neighbouring points projected in Euclidean space to enable direct feature

learning of point cloud data, and recursively sub-samples point regions for evaluating the local features of a point neighbourhood. The training of the network is based on mapping an input point set to a vector representation that is first fed to a multilayer perceptron, then to a max pooling layer. The global feature vector used for generating the output score is obtained from the subsequent local feature vectors of subsampled regions.

However, the use of CNNs such as PointNet++ impose a considerable amount of overhead in terms of hardware resources and computation times. An alternative to using 3D CNNs for classification of point cloud is using 2D CNNs instead, which classify images depicting multiple views of point clusters and associating the classification results of such views back with corresponding point clusters (SU et al. 2015). Subsequent evaluation of this approach, known as *multiview classification*, has shown that it can in certain cases provide just as accurate results as using a 3D CNN, but without the computational overheads and data bandwidth requirements needed for parsing and processing large point cloud datasets (WANG et al. 2017). Another approach is a combination of both 2D and 3D CNNs for classification of 3D data, and variants of such approaches have been described by IOANNIDOU et al. (2017).

Finally, one of the other main benefits of using point clouds is that planar regions representing walls, floors and ceilings can be segmented and reconstructed as BIM data at a given Level-of-Detail (LOD). Since manual generation of BIM data tends to be a laborious and expensive process (VOLK et al. 2014), the use of various segmentation algorithms can automate the process of “scan to BIM” reconstruction (OCHMANN et al. 2019).

### 3 Approach

We implemented and evaluated two different point cloud classification methods, one using image-based classification with the multiview approach with a 2D CNN (Inception V3), and the other using point cloud-based classification (object-based) using a 3D CNN (PointNet++). Both CNN retraining and classification methods were implemented using Python 3.6, with Tensorflow (for Inception V3) and PyTorch (for PointNet++). Segmentation and generation of bounding boxes of point clusters were implemented as custom command line tools in C++ using the PCL framework (RUSU and COUSINS 2011). Finally, the visualization outputs of the semantically-enriched point clouds were implemented using the WebGL-based Three.js framework in JavaScript and HTML 5, in order to enable accessible and flexible web-based 3D visualization (CABELLO 2010). Both classification approaches were tested using the S3DIS dataset and on a custom point cloud dataset. The training and testing of both CNNs and classification approaches were performed on a commodity laptop with an Intel i5 1.8 GHz CPU, 8 GB RAM, and NVidia GeForce MX150 GPU with 2 GB video memory.

#### 3.1 Custom Dataset Capture and Pre-Processing

The custom dataset used for testing of the semantic segmentation was captured using a Google Tango specification compatible mobile phone, and utilizing Time of Flight (ToF) and depth-

sensing. The capture of the main hallway, conference room, kitchen and three office rooms was completed during day time under natural lighting conditions. Windows and highly-reflective surfaces were not fully captured, and were thus excluded from the scan. The point cloud segments were then manually aligned using the official floorplan as guidance. The initially captured point cloud was further manually edited where noisy and partially scanned clusters without significance were captured. The point cloud originally contained 2 506 858 points, but was sub-sampled to 501 372 points to decrease processing time. Finally, the office rooms, hallway, kitchen and communal area were segmented to form 7 different point cloud scenes for evaluation (Fig. 1).

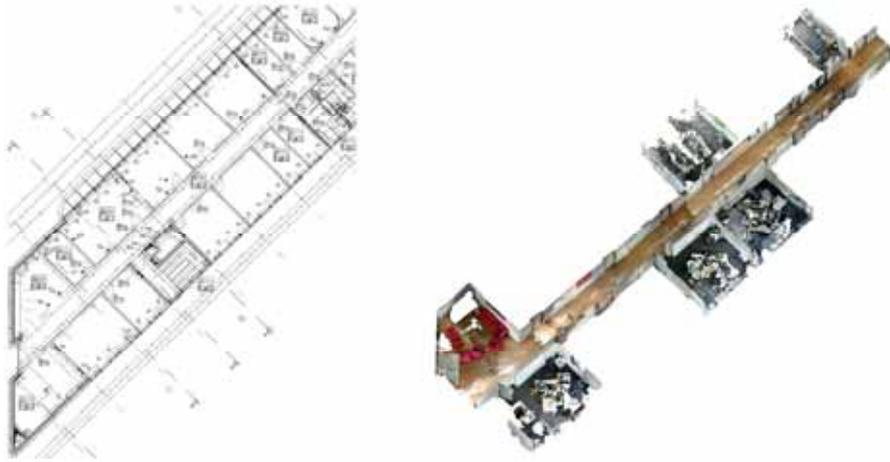


Fig. 1: The floorplan (left) and the captured point cloud of selected areas of the office (right) used to generate the custom dataset for the case study.

### 3.2 Multiview Classification Implementation Details

The multiview classification is based on the approach described in STOJANOVIC et al. (2019b). This approach uses a retrained 2D CNN (Inception V3) to classify images obtained from entropy-selected viewpoints of 3D point clusters. The last bottleneck layer of the CNN was retrained with 4000 epochs, with a learning rate of 0.01 and using 9759 different RGB images of chairs, tables and sofas (with a 70/30 split for training and validation images). Two CNN versions were used, one for detecting only images chairs and tables, and the other for detecting images of sofas as well. The predicted classification accuracy of the CNN version used to classify only chairs and tables (the most common office furniture), was 92.9% while version of the CNN used to classify scenes with both chairs, sofas and tables was 94.5% (with sofas being a less common office furniture item). Prior to classification, the 3D point clouds are clustered, so that point clusters representing cluttered and smaller items can be separated from furniture objects that we want to classify. We used the  $k$ -means clustering approach in order to spatially partition the furniture objects represented by point clusters prior to multiview image generation and classification. We then use randomly sampled 3D point normal vectors as an entropy descriptor to calculate the direction and position of the virtual camera placed around furniture object clusters. The viewpoint entropy method then synthesizes 2D images of a given cluster from multiple views with positions and

directions that have highest visual entropy prior to classification. On average, our multiview approach generates approximately 10 images per point cluster, which are then classified by the retrained CNN and their probability-based results are assigned back to each of the corresponding point clusters.

### 3.3 Object-Based Classification Implementation Details

A PyTorch implementation of the PointNet++ 3D CNN architecture was used in order to assess the feasibility of using 3D CNNs and point clouds for training and classification of indoor office environments. We trained two different versions of the PointNet++ model using the S3DIS dataset (one with and one without additional RGB information of the point data). Both the RGB and non-RGB versions of the PointNet++ CNN model were trained using 112 different scenes from the SI3D dataset, featuring typical indoor offices (with a 81/31 split for training/testing data). The CNN was trained for 200 epochs, with a learning rate of 0.001. We sampled 8192 points for each training cycle. The validation accuracy for the RGB version of the CNN was 64.4%, while for the non-RGB version it was 67.2% (this includes the validation accuracy of the whole scenes, not just furniture objects).

We did not use precomputed point normals as a training feature, since by default the S3DIS set does not contain any normals. A small batch size of 1 was chosen due to computational hardware restraints. The S3DIS already featured labelled data, which was used later on to generate ground truth models to test the accuracy for semantic segmentation of the specific scenes (for the multiview classification approach as well).

### 3.4 As-is BIM Generation Approach

An object-oriented bounding box (OOBB) representation of point clusters can be extracted and used for generating a 3D floorplan model. These planar clusters are the result of the *Region Growing* (RG) segmentation process, which is suited towards segmenting planar region point clusters that most commonly represent walls, floors and ceilings in indoor point clouds (BASSIER et al. 2017). The 3D floorplan generation involves computing the bounding box for each of the segmented planar point clusters, and exporting them as a text file along with the associated semantic label. We implemented this as a custom command-line tool using the PCL framework. The resulting clusters are exported and used for further editing, analysis, representation and decision making tasks, such as generation of Industry Foundation Classes (IFC) data.

## 4 Case Study

We presented experimental results for the multiview and object-based classification approaches for common office furniture (Fig. 2), as well as basic reconstruction of segmented core structural features (e.g., walls, floors, and ceilings). For comparing the classification accuracy of the multiview and object-based classification approaches, we empirically evaluated and compared the

semantically segmented point clouds of 14 offices from Area 1 of the S3DIS dataset. We evaluated both RGB and non-RGB based semantic segmentation using PointNet++. The 14 office areas from the S3DIS set used for testing were chosen due to featuring a large amount of visual clutter. They contain three of the furniture object categories used to train the multiview CNN to classify the data (tables, chairs and sofas).



Fig. 2: Examples of point clouds representing common office furniture and indoor locations

Furthermore, we evaluated the classification capabilities of the multiview-based approach for semantic segmentation on our custom point cloud dataset. The custom point cloud dataset features extracted office furniture items from the main office, kitchen and common areas. The accuracy of the two classification approaches for semantic segmentation was obtained by detecting and finding the average of intersection points between the ground truth and predicted point sets - where an intersecting point is defined as having the same location and color as the corresponding point in the ground truth set. Finally, we present experimental results for 3D mesh generation of the floorplan of our custom dataset using OOB-based reconstruction of point clusters segmented using the RG algorithm.

#### 4.1 Classification Results

We present classification results using both the multiview and object-based classification methods, tested using the SI3D dataset and our custom dataset. The results from the multiview-based semantic segmentation show that the multiview-based approach has an accuracy of 52.22% for the custom dataset (Tab. 1), and an average accuracy of 39.56% for the SI3D dataset (Tab. 2). The obtained results from the object-based semantic segmentation of the SI3D dataset show that it has an average classification accuracy of 83.51% for the non-RGB version of the CNN, and an average classification accuracy of 81.14% for the RGB version of the CNN (Tab. 3).

#### 4.2 Experimental 3D Floorplan Generation Results

In addition, we also investigated the use of segmented point clusters of walls, floors and ceilings for the generation of as-is BIM and geometry data. The point clusters were segmented using RG

segmentation, and reconstructed using the OOBBs of the point clusters. Since the OOBBs are generally not correctly aligned, further manual editing was needed in order to create the final 3D floorplan representation (Fig.3). An alternative method for floorplan generation is also possible using horizontally segmented regions of the complete indoor point clouds. Using such an approach, concave shapes obtained from the boundary evaluation of the point cloud in the 2D projected plane is used to generate vectorized contours that can be extruded into 3D polygonal shapes (Stojanovic et al. 2019c).

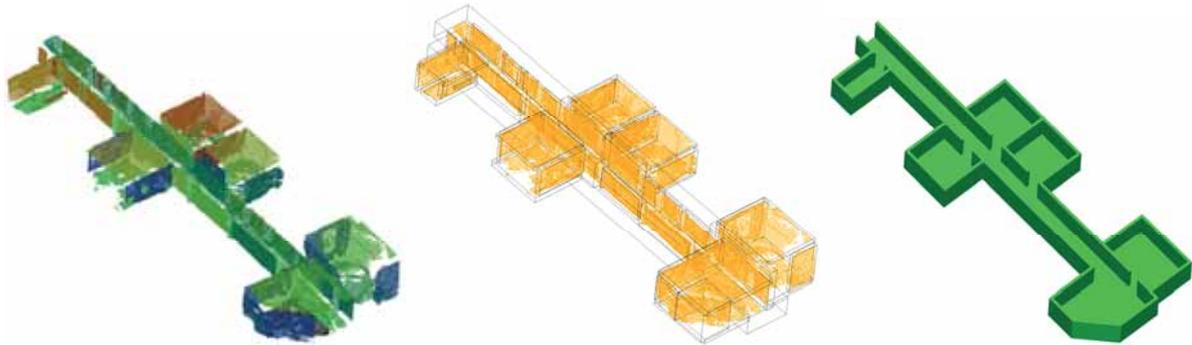
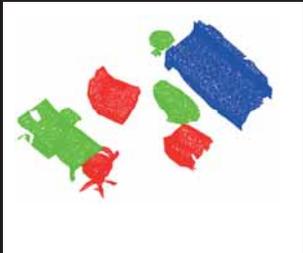
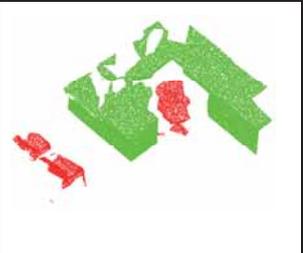
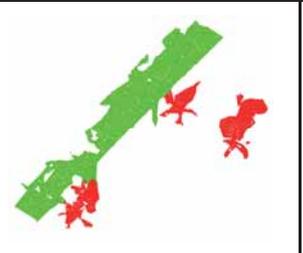
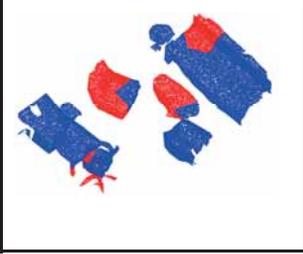
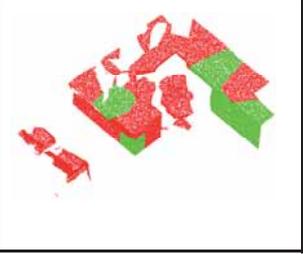
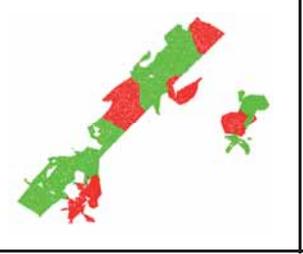


Fig. 3: Example RG-based segmentation of the main hallway and office areas of the custom dataset (left). The resulting OOBBs (middle) can then be exported and used to generate an as-is BIM representation at LOD300 level (right), using various specialist BIM software or software libraries.

Tab. 1: The ground truth (top) and predicted (bottom) results for the custom dataset classified using the multiview-based approach. Red point clusters represent chair objects, blue represent sofas and green represent tables.

Common Area	Kitchen	Office 1	Office 2

Tab. 2: The ground truth (top) and predicted (bottom) results for the SI3D classified the using the multiview-based approach

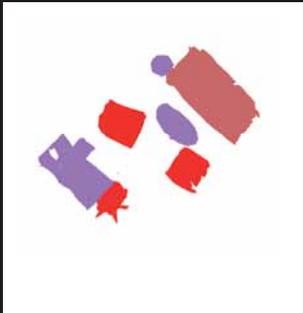
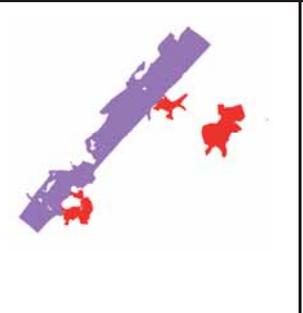
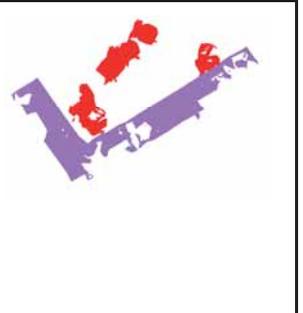
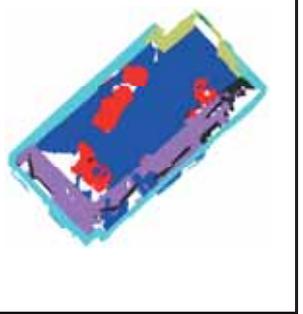
Area 1 Office 2	Area 1 Office 6	Area 1 Office 18	Area 1 Office 24
			
			

## 5 Discussion

For applications where classification accuracy is more critical, the use of PointNet++ is a better option if high-end computing resources are available. The retraining of the Inception V3 CNN with images of real-life furniture, used for the multiview-based approach, takes only a fraction of the time compared to training the PointNet++ 3D CNN using the S3DIS dataset. The retraining time for Inception V3 is approximately 30 minutes, while the training time for the PointNet++, which had to be trained from scratch, was approximately 15 hours. The semantic segmentation using PointNet++ on each of the offices from the SI3D takes on average approximately 90 minutes, while the multiview-based approach on segmented office clusters from the same dataset takes on average approximately 6 minutes. The main challenge when using PointNet++ for semantic segmentation is that it relies on point cloud data for training, and this is impractical for integration with commodity computing hardware compared to the multiview-based classification approach (as most implementations of PointNet++ are designed to leverage the computing power of parallel multi-GPU architectures). We have also observed that the PointNet++ CNN versions making use of RGB point color features had slightly worse evaluated classification accuracy. While the multiview-based approach offers worse classification accuracy in comparison to the object-based approach, it is generally more adaptable for use on commodity hardware, and retraining the CNN (using only the last bottleneck layer), is more convenient with access to potentially millions of images on the internet depicting common office furniture in different categories (in comparison to having to find and use point clouds of such objects, which is often difficult). One drawback of the multiview-based approach is that the viewpoint entropy method used for generating the multiview images relies on correctly calculated point normals that conform to the surface curvature (in terms of viewing the object from a given distance from its center with 3D perspective). Finally, the

reconstruction of 3D floor plans from RG-segmented point clusters provides favourable experimental results for further use for generation of as-is BIMs, and can be expanded to include reconstruction of higher levels of detail and information.

Tab. 3: The ground truth (top) and predicted non-RGB CNN version (middle), and RGB CNN version (bottom) results for the SI3D classified the using the multiview-based approach. Red point clusters represent chair objects, copper represents sofas and purple represents tables. For these results the semantically segmented walls, floors and ceilings are also included, but were not evaluated as part of the classification accuracy result

Area 1 Office 2	Area 1 Office 6	Area 1 Office 18	Area 1 Office 24
			
			
			

## 6 Conclusions and Outlook

Our research investigated the practical use of CNNs for semantic enrichment of indoor point clouds. We compared multiview versus object-based classification for indoor point clouds, with a particular focus on semantic segmentation. We also experimented with a basic method for reconstructing 3D floor plans using bounding boxes of point clusters representing walls, floors and ceilings. We have found that the use of the use of object-based classification with PointNet++

offers superior classification accuracy in comparison to the multiview-based classification approach. However, there is a trade-off between practicality and performance versus classification accuracy when deciding which approach to use for deep-learning-based semantic segmentation. Future work will investigate the possible combined use of a 2D CNN with a 3D CNN for classification of indoor point clouds. We predict that a balance in terms of computational performance and classification accuracy can be achieved using a combined 2D and 3D CNN approach, along with the use of semantic reasoning.

## 7 Acknowledgments

This work has been partially funded by the Research School on Service-Oriented Systems Engineering of the Hasso Plattner Institute, Faculty of Digital Engineering, University of Potsdam, Germany.

## 8 References

- ARMENI, I., SENER, O., ZAMIR, A. R., JIANG, H., BRILAKIS, I., FISCHER, M. & SAVARESE, S., 2016: 3D Semantic Parsing of Large-Scale Indoor Spaces. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1534-1543.
- BABACAN, K., CHEN, L. & SOHN, G., 2017: Semantic Segmentation of Indoor Point Clouds Using Convolutional Neural Network. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, 101-108.
- BASSIER, M., BONDUÉL, M. VAN GENECHTEN, B., & VERGAUWEN, M., 2017: Segmentation of Large Unstructured Point Clouds Using Octree-Based Region Growing and Conditional Random Fields. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **42**(2W8), 25-30.
- CABELLO, R., 2010: Three.js. URL: <https://github.com/mrdoob/three.js>.
- CHIANG, H. Y., LIN, Y. L., LIU, Y. C. & HSU, W. H., 2019: A Unified Point-based Framework for 3D Segmentation. IEEE International Conference on 3D Vision (3DV), 155-163.
- IOANNIDOU, A., CHATZILARI, E., NIKOLOPOULOS, S. & KOMPATSIARIS, I., 2017: Deep Learning Advances in Computer Vision with 3D Data: A Survey. ACM Computing Surveys (CSUR) **50**(2), 20.
- MALINVERNI, E. S., PIERDICCA, R., PAOLANTI, M., MARTINI, M., MORBIDONI, C., MATRONE, F. & LINGUA, A., 2019: Deep Learning for Semantic Segmentation of 3D Point Cloud. ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **4215**, 735-742.
- OCHMANN, S., VOCK, R. & KLEIN, R., 2019: Automatic Reconstruction of Fully Volumetric 3D Building Models from Oriented Point Clouds. ISPRS journal of photogrammetry and remote sensing, **151**, 251-262.
- QI, C. R., YI, L., SU, H. & GUIBAS, L. J., 2017: Pointnet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. In Advances in neural information processing systems, 5099-5108.

- RUSU, R. B. & COUSINS, S., 2011: 3D is here: Point cloud library (pcl). IEEE international conference on robotics and automation, 1-4.
- STOJANOVIC, V., TRAPP, M., RICHTER, R., HAGEDORN, B. & DÖLLNER, J. 2018a: Towards The Generation of Digital Twins for Facility Management Based on 3D point Clouds. In Proceedings of the 34th Annual ARCOM Conference.
- STOJANOVIC, V., TRAPP, M., DÖLLNER, J. & RICHTER, R. 2019b: Classification of Indoor Point Clouds Using Multiviews. In The 24th International Conference on 3D Web Technology, 1-9.
- STOJANOVIC, V., TRAPP, M., RICHTER, R. & DÖLLNER, J. 2019c: Generation of Approximate 2D and 3D Floor Plans from 3D Point Clouds. In Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications **1**, 177-184.
- SU, H., MAJI, S., KALOGERAKIS, E. & LEARNED-MILLER, E., 2015: Multi-View Convolutional Neural Networks for 3D Shape Recognition. In Proceedings of the IEEE international conference on computer vision, 945-953.
- SZEGEDY, C., VANHOUCHE, V., IOFFE, S., SHLENS, J. & WOJNA, Z., 2016: Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2818-2826.
- VOLK, R., STENDEL, J. & SCHULTMANN, F., 2014: Building Information Modeling (BIM) for Existing Buildings—Literature Review and Future Needs. *Automation in construction* **38**, 109-127.
- WANG, C., PELILLO, M. & SIDDIQI, K., 2019: Dominant Set Clustering and Pooling for Multi-View 3D Object Recognition. arXiv preprint arXiv:1906.01592.
- WANG, W., YU, R., HUANG, Q. & NEUMANN, U., 2018: SGPN: Similarity Group Proposal Network for 3D Point Cloud Instance Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2569-2578.
- WINIWARTER, L. & MANDLBURGER, G., 2019: Classification of 3D Point Clouds using Deep Neural Networks. *Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V.*, Band **28**, 663-674.
- WOLF, J., RICHTER, R. & DÖLLNER, J., 2019: Techniques for Automated Classification and Segregation of Mobile Mapping 3D Point Clouds. In Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications **1**, 201-208.

# Monitoring einer Schleuse mittels satellitengestützter DInSAR-Techniken

PHILIPP J. SCHNEIDER<sup>1</sup> & UWE SÖRCEL<sup>1</sup>

*Zusammenfassung: Die Auswertung hochaufgelöster Synthetic Aperture Radar (SAR) Daten mittels Persistent Scatterer Interferometry (PSI) kann kleinste Oberflächenbewegungen detektieren und bietet so eine Möglichkeit, bodengebundenen Überwachungsmessungen durch eine Satellitentechnik zu ergänzen. Im Rahmen eines Forschungsprojektes soll ein Schleusenbauwerk überwacht werden. Es kann gezeigt werden, dass sich die PSI Technik dazu eignet, kleinräumige Deformationen im Umfeld der Schleuse aufzudecken. Die Auswertung zweier verschiedener Beobachtungsgeometrien erlaubt eine Dekomposition der Line-of-Sight-Messungen in vertikale und horizontale Bewegungen. Eine statistische Analyse der Zeitserien mittels EOF legt nahe, dass sich die Bewegung des Bauwerks hauptsächlich auf einen Jahresgang zurückführen lassen.*

## 1 Motivation

Satellitengestützte differentielle SAR-Interferometrie-Methoden (DInSAR) wie die Persistent Scatterer Interferometry (PSI) haben sich mittlerweile als Verfahren zum Monitoring flächenhafter Deformationsprozesse der Erdoberfläche etabliert (CROSETTO 2016), etwa zur Überwachung von Vulkanen oder langsam fortschreitenden Hangrutschungen. Klassische Ansätze zur Bauwerküberwachung wie Nivellement und kontinuierliche Totalstationsmessungen liefern zwar hochgenaue Koordinatenwerte, die jedoch auf eine vergleichsweise geringe Menge von Messpunkten beschränkt bleiben müssen.

In einem Forschungsprojekt der Bundesanstalt für Gewässerkunde (BfG) und des Instituts für Photogrammetrie der Universität Stuttgart wird die Schleuse Hessigheim am Neckar überwacht. Das Umfeld der Schleuse sinkt durch instabilen Boden und untergrabenden Wasserfluss um bis zu 2 cm pro Jahr ab (KAUTHER & WOLF 2018).

In diesem Projekt untersuchen wir unter Anderem das Potential von PSI, die genannten traditionellen Methoden zur Überwachung von Wasserbauwerken zu ergänzen, insbesondere was die Ausweitung der Überwachung in die Fläche angeht.

## 2 Methoden

Im Folgenden werden wir zunächst auf die verwendeten Synthetic Aperture Radar (SAR) Daten eingehen, insbesondere was deren zeitliche und räumliche Auflösung angeht. Des Weiteren wird die PSI-Technik mit ihren speziellen Stärken und Einschränkungen erklärt. Durch Fusion zweier Messreihen, aus unterschiedlichen Beobachtungsrichtung, kann ein umfassenderes Bild des tatsächlichen Deformationsgeschehens gewonnen werden. Schlussendlich wenden wir Methoden der Zeitserienanalyse an, um räumliche und zeitliche Muster aus den PSI Ergebnissen ableiten.

---

<sup>1</sup> Universität Stuttgart, Institut für Photogrammetrie, Geschwister-Scholl-Str. 24D, D-70174 Stuttgart, E-Mail: [philipp.schneider, uwe.soergel@ifp.uni-stuttgart.de

## 2.1 Datengrundlage

Für die hier beschriebene Analyse wurden hochaufgelöste SAR Bilder des Satelliten TerraSAR-X (TSX) über einem Zeitraum von 4.5 Jahren verwendet (Siehe Abb. 1). Die Auflösung von TSX im „High Resolution Spotlight Mode“ beträgt etwa 1m x 1m. Der Satellit arbeitet im X-Band (9.65 GHz), was einer Wellenlänge von etwa 3 cm entspricht. Für den Zeitraum 2018 bis 2019 liegen zwei Bildstapel vor (aufsteigender und absteigender Orbit), sodass die LOS-Messungen in horizontale und vertikale Komponenten zerlegt werden können. Für den Zeitraum 2015 bis Ende 2017 ist hingegen nur der absteigende Orbit verfügbar. Das kommerzielle TSX-System ist theoretisch dazu in der Lage, alle 11 Tage ein Bild von derselben Position im Orbit aufzunehmen, jedoch können konkurrierende Aufnahmen höherer Priorität zum Aufnahmeverzicht und somit zu Datenlücken führen. Es wurde in einer vorherigen Studie auch untersucht, ob sich in ähnlicher Weise die frei verfügbaren Sentinel-1-Daten für eine PSI-Analyse des Studiengebiets nutzen lassen. Durch den räumlichen Auflösungsunterschied der beiden Systeme ist hier jedoch mit einer zu geringen PSI-Punktdichte zu rechnen, sodass die kleinräumigen lokalen Deformationsgeschehen nicht zufriedenstellend abgebildet werden können.

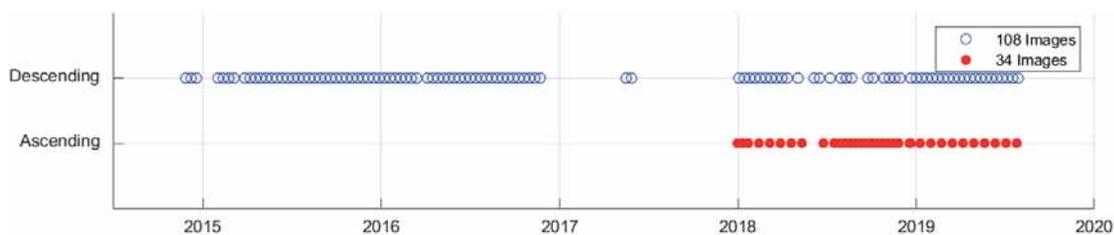


Abb. 1: Aufnahme Zeitpunkte der TSX Bilder. Blau: Absteigender Orbit. Rot: Aufsteigender Orbit.

## 2.2 Persistent Scatterer Interferometrie

Die SAR-Auswertetechnik Persistent Scatterer Interferometrie (PSI) zählt zu einer speziellen Klasse der DInSAR-Methoden. Die prinzipielle Idee ist hierbei, in einer Zeitreihe von SAR-Satellitenbildern solche Pixel zu detektieren, welche über die Zeit ein „kohärentes“ also stabiles Signal aufweisen. SAR-Bilder liegen komplexwertig vor, was eine Analyse des zeitlichen Verlaufs der Phasenlage dieser Pixel ermöglicht, welche wiederum Rückschlüsse auf den zugrundeliegenden Deformationsverlauf erlaubt (FERRETTI et al. 2000).

PSI ist ein opportunistisches Verfahren, das bedeutet, dass die genaue Position und Verteilung der PS-Punkte vor der Auswertung nicht bekannt ist. Es kann jedoch davon ausgegangen werden, dass bei höherer räumlicher Auflösung der ausgewerteten Bilder mehr PS Punkte pro Fläche erhalten werden. Die Rückstreuungseigenschaften von Metall und rechtwinkligen Strukturen führen dazu, dass auf menschengemachten Objekten mit einer hohen PS-Punktdichte zu rechnen ist.

PSI hat sich in den letzten zwei Jahrzehnten zu einem gutverstandenen Standardverfahren entwickelt, für das auch kommerzielle Auswertungsprogramme angeboten werden (CROSETTO 2016). Für eine aussagekräftige Analyse wird in der Regel eine Zeitreihe von mindestens 20 Bildern vom selben Orbit benötigt.

In der PSI-Auswertung kann neben der 3D-Lage auch eine Deformationszeitserie für jeden der Reflexionspunkte bestimmt werden. Unter bestimmten Voraussetzungen ist das Verfahren dazu in der Lage, Geländedeformationen in der Größenordnung von Millimetern aufzudecken. Wird von

einem linearen Deformationsgeschehen ausgegangen, kann die Geschwindigkeit durch lineare Regression hierbei mit besser als 1mm/Jahr bestimmt werden (FERRETTI et al. 2001).

### 2.3 Line-of-Sight Dekomposition

Von einem beliebigen 3D-Deformationsgeschehen kann bei interferometrischen SAR-Auswertungen nur jene Komponente erfasst werden, welche dessen Projektion auf den sogenannten „Line-of-Sight“-Vektor (LOS), also der Radialrichtung bezogen auf den Satelliten, entspricht (Abb. 2). Im Rahmen des Projekts werden deshalb zwei parallel erfasste Zeitreihen von TerraSAR-X-Daten betrachtet. Es werden hochauflösende TSX-Bilder von auf- und absteigenden Satellitenbahnen verarbeitet und zunächst unabhängig voneinander mit der PSI-Methode analysiert, was beutet, dass die Szene einmal von Westen und einmal von Osten her beobachtet wurde. Die Zweiorbitkonfiguration ermöglicht es, die LOS-Messung in horizontale und vertikale Komponenten zu zerlegen und mögliche Verformungen zu validieren. Um die beiden unabhängigen Messungen zu verknüpfen, muss als erster Schritt eine Interpolation der punkthaften Messungen auf ein gleichförmiges Gitterraster durchgeführt werden. Anschließend kann so für jede Gitterzelle eine Zerlegung in horizontale und vertikale Bewegungen durchgeführt werden.

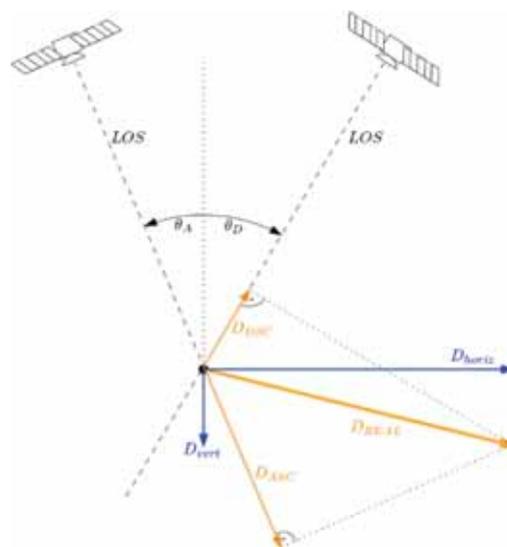


Abb. 2: Die tatsächliche Bewegung  $D_{REAL}$  wird auf den Line-Of-Sight Vektor des Satelliten projiziert. Aus zwei unabhängigen Messungen kann auf die vertikale und horizontale Komponente geschlossen werden.

Die horizontale und vertikale Komponente  $D_{horiz}$  und  $D_{vert}$  ergeben sich aus der Lösung folgenden Gleichungssystems,

$$\begin{bmatrix} D_{ASC} \\ D_{DSC} \end{bmatrix} = \begin{bmatrix} \cos \theta_A & \sin \theta_A / \cos \Delta\alpha \\ \cos \theta_D & \sin \theta_D \end{bmatrix} \cdot \begin{bmatrix} D_{vert} \\ D_{horiz} \end{bmatrix}$$

wobei  $\theta$  der Einfallswinkel und  $\Delta\alpha$  der Unterschied im Azimut der Satelliten sind (vgl. Abb. 2) (SAMIEI ESFAHANY et al. 2010).

## 2.4 Zeitserienanalyse mittels Empirischer Orthogonaler Funktionen

Die PSI-Punkte enthalten neben der recht groben Lage- und Höheinformation (Genauigkeiten im Meter-bis Dezimeterbereich, unter anderem abhängig von der Anzahl der SAR-Bilder) auch die LOS-Bewegung des Punktes über die Zeit. Da Aufnahmen über mehrere Jahre vorliegen, bietet die Auswertung einer Zeitserie von 2015 bis 2019 die Möglichkeit, mittels statistischer Verfahren wie „Empirischer Orthogonaler Funktionen“ (EOF) unterschiedliche Bewegungsmuster zu identifizieren und so lineare Trends von saisonal bedingten Jahregängen zu trennen (BJORNSSON & VENEGAS 1997).

Sortiert man die  $N$  Deformationszeitserien in einem Gebiet, welche jeweils aus  $M$  Messungen bestehen in eine Matrix  $F$

$$F = \begin{bmatrix} d_1(t_1) & d_2(t_1) & \dots & d_{N-1}(t_1) & d_N(t_1) \\ d_1(t_2) & d_2(t_2) & \dots & d_{N-1}(t_2) & d_N(t_2) \\ \vdots & \vdots & \dots & \vdots & \vdots \end{bmatrix}$$

und führt eine Singulärwertzerlegung (ökonomische Variante) dieser durch,

$$U \cdot S \cdot V^T = svd(F)$$

so erhält man mit den Spalten von  $U$  orthogonale Funktionen  $M$ , sogenannte „Moden“.

Durch Analyse der Spur von  $S$  kann diesen Moden ein entsprechender Anteil der Gesamtvarianz der Zeitserien zugeordnet werden. Typischerweise gehen lineare Trends oder periodische Signale mit signifikanten Moden einher. Enthalten die Zeitserien allerdings weitere Auffälligkeiten, welche aus einer fehlerhaften Prozessierung oder dem Einfluss der Atmosphäre herrühren können, so bilden diese auch einen Modus aus.

Möchte man nun feststellen, welche der Zeitserien hauptsächlich aus einer der Moden aufgebaut ist, können die Spalten von  $V$  analysiert werden, deren Einträge einer Gewichtung der Moden zur Rekonstruktion der einzelnen Zeitserien entsprechen.

## 3 Ergebnisse

Aufgrund der hohen PSI-Punktdichte der TerraSAR-X Ergebnisse können flächenhafte Deformationsprozesse beobachtet und lokal eingegrenzt werden. So zeigt Abb. 3 Bereiche mit kritischer Absenkung deutlich. Diese decken sich mit Erwartungsbereichen, die sich durch geophysikalische Modellierung der Beschaffenheit des Untergrundes und der darauf einwirkenden Kräfte ergeben, welche die Bundesanstalt für Wasserbau durchgeführt hat (KAUTHER & SCHULZE 2015).

Die gemittelten Deformationszeitserien in den markierten Bereichen „Osten“ in Abb. 3 sind in Abb. 4 jeweils für die aufsteigende und absteigende Messung dargestellt, es zeigt sich hierbei signifikante, lineare Absenkungen von ca. 15mm/Jahr in LOS Richtung. Die Unterschiede in der beobachteten Geschwindigkeit können hierbei auf die leicht unterschiedlichen Einfallswinkel der LOS zurückgeführt werden. Die Dekomposition der beiden Messungen in ihre vertikale und horizontale Komponente in Abb. 3 zeigt, dass das Deformationsgeschehen vornehmlich von vertikale Natur ist. Es bestätigt die Absenkungen am südöstlichen Ufer, sowie kleinere Setzungen im Norden und Nordwesten.

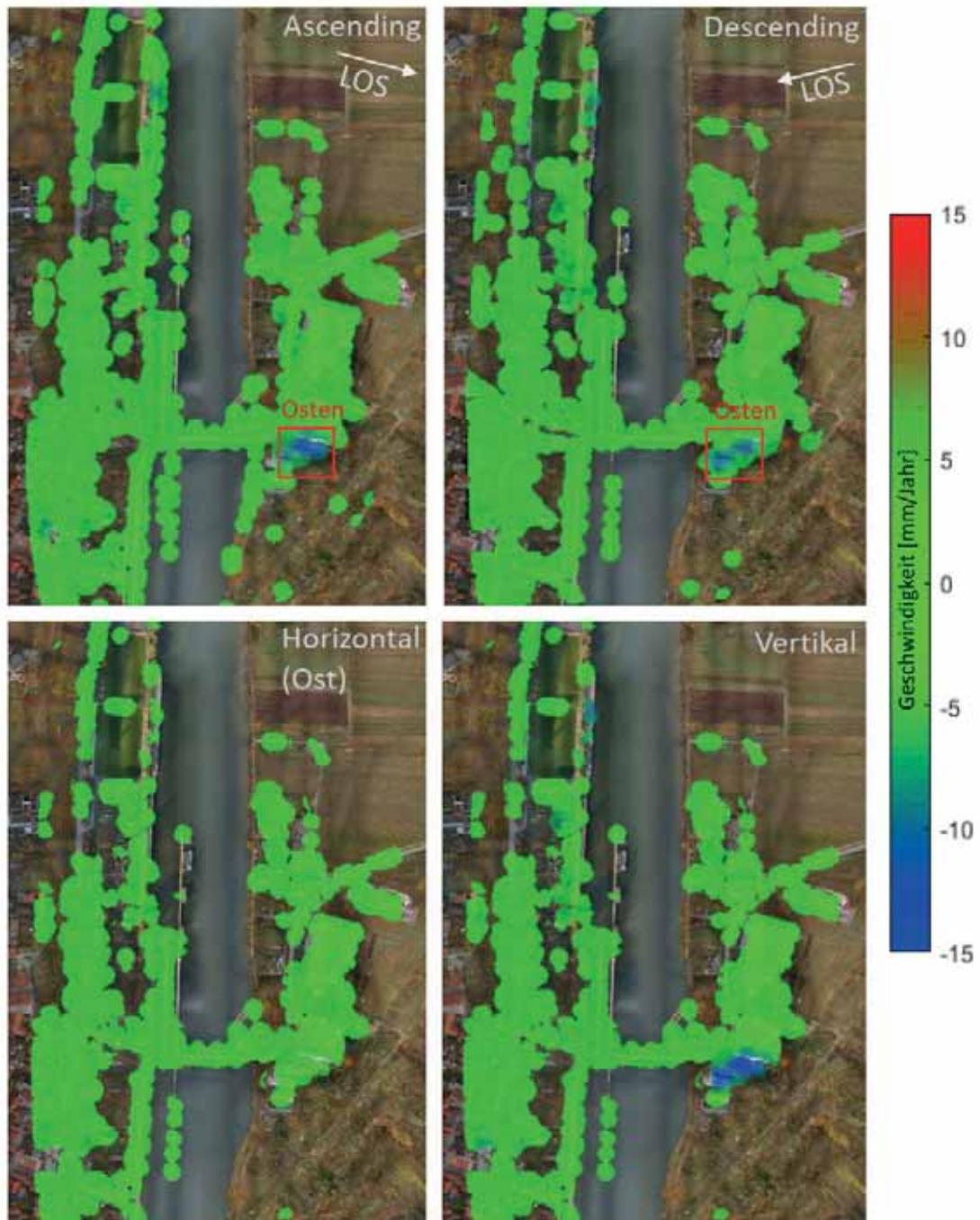


Abb. 3: Oben: LOS-Deformationen in mm/Jahr aus TSX PSI (auf Raster) mit ausgewähltem Bereich „Osten“. (Links: ascending, rechts descending Orbit). Darstellung der Zeitserien der markierten Bereiche in Abb. 4. Unten: Horizontale und vertikale Bewegung nach Dekomposition.

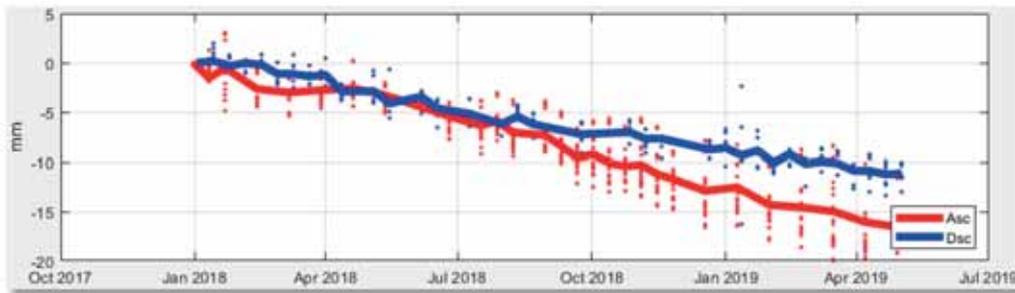


Abb. 4: Zeitreihe der LOS-Deformationen im markierten Gebiet „Osten“ in Abb. 3. Rot: Aufsteigender Orbit, blau: Absteigender Orbit.

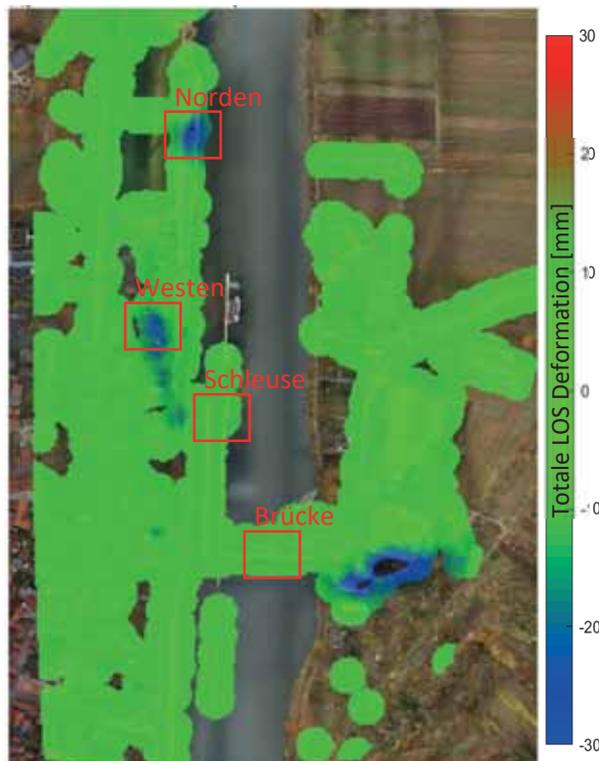


Abb. 5: LOS-Deformationen für gesamten verfügbaren Bilder des absteigenden Orbits (2015-2019). Die Zeitserien für die markierten Bereiche sind in Abb. 6 dargestellt.

Durch die Auswertung der mehrjährigen Zeitserie können die Bodenbewegungen am südöstlichen Ufer bestätigt werden. Des Weiteren heben sich nun die Deformationen in den Bereichen „Norden“ und „Westen“ in Abb. 5 deutlich vom Messrauschen ab. Eine Analyse der Punkte auf dem Schleusenbauwerk („Schleuse“ und „Brücke“ in Abb.5) zeigt lediglich eine periodische Bewegung, welche als Jahresgang identifiziert werden kann (Abb.6).

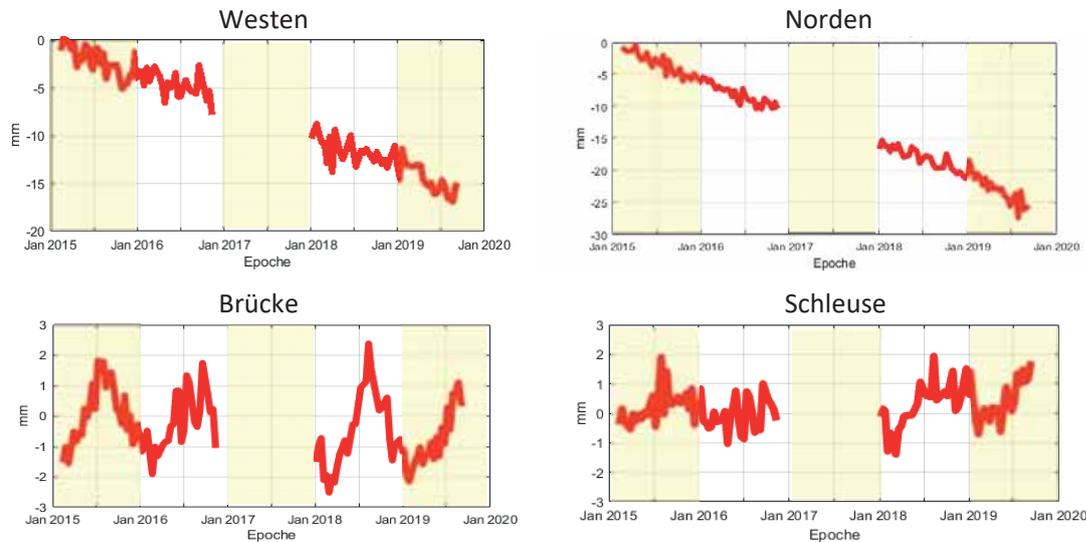


Abb. 6: Die LOS Deformationen für die Markierten Bereiche in Abb. 5. Die Bereiche „Brücke“ und „Schleuse“ weisen einen deutlichen Jahrgang auf, wohingegen „Westen“ und „Norden“ einen klaren linearen Trend aufweisen.

Die Dekomposition der Zeitserien des gesamten absteigenden Bildstapels von 2015 bis 2019 in Empirische Orthogonale Funktionen (EOF) zeigt, dass die beiden signifikantesten Bewegungsmuster (Moden) ein linearer Trend und ein Jahrgang sind.

Eine Analyse der Gewichtungen dieser beiden Moden für jeden der einzelnen Punkte ergibt, dass die zuvor detektierten Deformationsbereiche im Südosten und Westen der mehrjährigen linearen Bewegung unterliegen. Für den Schleusenbereich liegen keine generellen Trends vor. Die dort beobachteten Bewegungen können plausibel durch saisonale Effekte erklärt werden, welche periodischen auftreten. In Abb. 7 werden alle PS-Punkte auf der Schleuse und dem Uferbereich dargestellt. Alle Punkte, deren Deformationshistorie sich hauptsächlich durch eine Mode beschreiben lässt sind farbig dargestellt. Die Einfärbung gibt an wie groß der Betrag der Gewichtung der jeweils darunter dargestellten Mode ist.

## 4 Diskussion

Wir konnten zeigen, dass sich die PSI-Analyse von hochaufgelösten TerraSAR-X Daten über dem Zeitraum von einem Jahr dazu eignet, kleinräumige, lineare Deformationen in der Größenordnung von kleiner als 1cm/Jahr zuverlässig zu erkennen. Durch die geometrische Dekomposition zweier Line-of-Sight-Messungen ergab sich, dass die Verformungen hauptsächlich vertikaler Natur sind. Die Analyse einer mehrjährigen Zeitserie bestätigt die zuvor ausgemachten Bereiche einer Geländebewegung und kann weitere, langsame Deformationen aufdecken. Durch die EOF-Analyse konnten die Bewegungsmuster in lineare und periodische Bewegungen getrennt werden. Daraus geht die Erkenntnis hervor, dass das Schleusenbauwerk an sich, neben dem unkritischen Jahrgang, keiner signifikanten Bewegung mehr unterliegt.

Ausgehend von dieser Erkenntnis schlagen wir vor, die Überwachungsmessungen der Schleuse auch auf die umliegenden Bereiche auszudehnen. Um eine Einschätzung zu bekommen, in wiefern

Auslaugungsprozesse die Ursache für die Senkungen sind, könnten ggf. weitere geologische Untersuchungen und eventuell Sondierungsbohrungen in den von uns detektierten Bereichen durchgeführt werden.

Die PSI-Technik stellt sich zusammenfassend als nützliches Werkzeug in einem umfassenden Monitoringkonzept dar. Eine Ausweitung und gegebenenfalls routinemäßige Überprüfungen aller Binnenschleußen könnte helfen, Deformationen im Umland frühzeitig zu erkennen und existierende Messprogramme in der Fläche zu ergänzen.

Die durch die zukünftig weiter steigende Verfügbarkeit von hochaufgelösten SAR-Daten kann davon ausgegangen werden, dass sich satellitengestützte InSAR-Verfahren zu einer weiteren Standardmethode im Repertoire des Bauwerksmonitoring entwickeln.

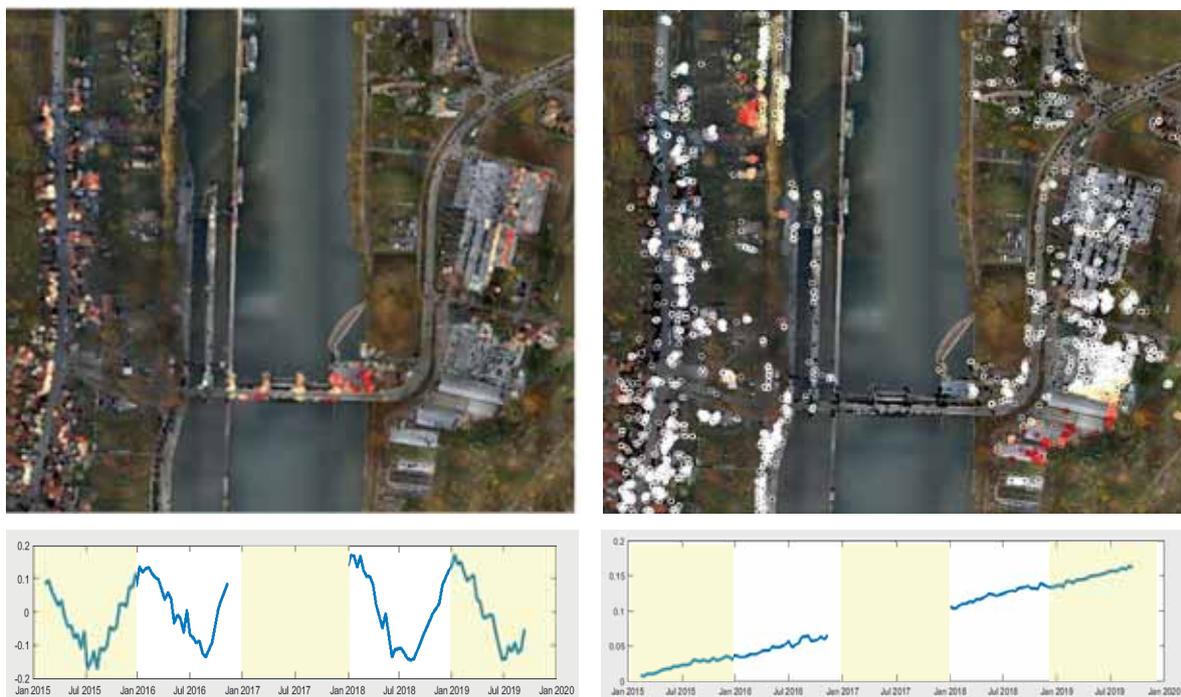


Abb. 7: Links: Punkte die sich hauptsächlich durch die „Jahresgang“ Mode (unten links) beschreiben lassen. Rechts: Punkte die sich hauptsächlich durch die „Lineare Mode“ (unten rechts) beschreiben lassen. Es wurden die Spalten von  $V$ , wie in Kap. 2.4 beschrieben analysiert und die Punkte mit dem Betrag der Gewichtung farbcodiert eingefärbt, sodass auch eine hohe negative Gewichtung der Mode als hoher Anteil dargestellt wird.

## 5 Literaturverzeichnis

- BJORNSSON, H., & VENEGAS, S. A., 1997: A Manual for EOF and SVD Analyses of Climatic Data. Centre for Climate and Global Change Research Report No. 97-1, 52.
- FERRETTI, A., PRATI, C., ROCCA, F., 2001: Permanent scatterers in SAR interferometry. IEEE Transactions on Geoscience and Remote Sensing, 8-20.
- FERRETTI, A., PRATI, C., ROCCA, F.: 2000: Nonlinear subsidence rate estimation using permanent scatterers in differential SAR Interferometry. Geoscience and Remote Sensing. IEEE Transactions on, 2202-2212.
- KAUTHER, R., & SCHULZE, R., 2015: Detection of subsidence affecting civil engineering structures by using satellite InSAR. Proceedings of the Ninth Symposium on Field Measurements in Geomechanics, Australian Centre for Geomechanics, 207-218.
- KAUTHER, R., & WOLF, C., 2018: Detection of Cavities in Karst Subsoil Affecting Civil Engineering. IFMGM 2018, 89-93.
- MICHELE CROSETTO, O. M.-G., 2016: Persistent Scatterer Interferometry: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 78-89.
- SAMIEI ESFAHANY, S., HANSEN, R., THIENEN-VISSER, K., & MUNTENDAM-BOS, A. G., 2010: On the effect of horizontal deformation on InSAR subsidence estimates. Fringe 2009 Proceedings.

## Analysis of Airborne SAR and InSAR Data for Coastal Monitoring

TOBIAS BOLZ<sup>1</sup>, SYLVIA SCHMITZ<sup>2,3</sup>, ANTJE THIELE<sup>2,3</sup>, UWE SOERGEL<sup>1</sup> & STEFAN HINZ<sup>2</sup>

*Abstract: The German Wadden Sea is a very fragile ecosystem listed as UNESCO World Heritage Site and therefore closely monitored. This monitoring includes but is not limited to the waterways, the erosion of the islands causes by tides and storms, the condition of coastal protection buildings and the land cover and use. The goal of the project called GeoWAM is to investigate whether or not it is possible to replace airborne Laser scanning, which is the state of the art for this task, by airborne Synthetic Aperture Radar. As part of the project consortium, the Karlsruhe Institute for Technology and the University of Stuttgart are investigating new methods for the automatic evaluation of radar data. In this paper we describe the efforts on the subject of classification, georeferencing and the extraction of 3D structures.*

### 1. Motivation

In order to be able to respond to urgent issues of water science, water management and ecology with regard to coastal protection, high-resolution and accurate geodata of coastal areas and water bodies are required. Today and in the future, we will be facing an increasing demand for geo basic products, especially in the form of digital terrain and surface models. These can be used to ensure the traffic safety of seafaring roads and support various federal and state authorities and NGOs in the implementation of their work on traffic safety and services of general interest in the coastal region. Nowadays, usually airborne laser scanning (ALS) is applied to derive demanded geodata products. With regard to the modelling of coastal areas, data collection on dry fallen mudflats and foreshore area is of particular interest. Hence, only extremely short time windows are available for the acquisition of the corresponding data, due to changing water levels and weather conditions. This is bottleneck for ALS, which suffers from clouds and fog and is usually bound to stripe width of some 500 m only. For this reason, remote sensing methods are desirable that allow a higher degree of flexibility. Current developments in airborne Interferometric Synthetic Aperture Radar (InSAR) technology lead us to expect new technical and economic advantages over operational technologies, such as ALS. These advantages are based in particular on the up to five times higher area coverage performance and the weather independence of the InSAR technology. However, for corresponding operational applications and application-related product provision, intensive research activities are still required with regard to data acquisition, data processing, user-oriented data analysis and product development. Therefore, we are investigating how to improve the entire development chain starting from more efficient capturing of airborne SAR and InSAR

---

<sup>1</sup> University of Stuttgart, Institute for Photogrammetry, Geschwister-Scholl-Straße 24D, 70174 Stuttgart, Germany, Email: [Tobias.Bolz, Uwe.Soergel]@ifp.uni-stuttgart.de

<sup>2</sup> Karlsruhe Institute of Technology, Institute of Photogrammetry and Remote Sensing, Englerstraße 7, 76133 Karlsruhe, Germany, Email: [Sylvia.Schmitz, Antje.Thiele, Stefan.Hinz]@kit.edu

<sup>3</sup> Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Gutleuthausstraße 1, 76275 Ettlingen, Germany, Email: [Sylvia.Schmitz, Antje.Thiele]@iosb.fraunhofer.de

data up to the deployment of ready-to-use coastal-relevant geodata products. The individual steps of this development chain are presented in Fig. 1.

In this paper, we focus in particular on the step of SAR analysis and interpretation. This includes the investigation and evaluation of the data acquired via airborne SAR with regard to user requirements for the purpose of water management in tidal affected areas. This includes questions such as the separation of soil and non-soil points and a further subdivision of the land cover into classes relevant for coastal areas, the optimization of georeferencing to geocode the results at the best possible rate as well as the extraction of structural edges within the dry fallen mudflats.



Fig. 1: Development chain, including flight planning, data acquisition, processing and analysing as well as generating and providing geodata products.

## 2. Data acquisition

As part of our project, we planned three flights for the data acquisition, whereof two have already taken place so far. We have chosen the coastal region near Spiekeroog and the Elbe estuary near Brunsbüttel as representative test areas. For the timing of the flight, a low tide level is of great importance, because the analysis of dry fallen mudflats is of particular interest. In addition, the measurement campaigns are arranged in such a way that two different vegetation conditions can be observed.

The airborne SAR system F-SAR of the German Aerospace Centre (Deutsches Zentrum der Luft- und Raumfahrt; DLR) (HORN et al. 2017) is used for data acquisition. It is capable to record simultaneously data at different wavelengths and polarizations in very high resolution. For our measurements, we used the X- and S-band antennas. To enable interferometry on the one hand a repeat-pass mode was used, additionally the configuration of the X- and S-band antennas allows single-pass interferometry. Amplitudes images, single look complex data, and interferometric intermediate products such as phase- and coherence images are generated by DLR based on the recorded raw data obtained during the flights. These data provide the basis for further analysis and interpretation and the generation of high-resolution geodata products.

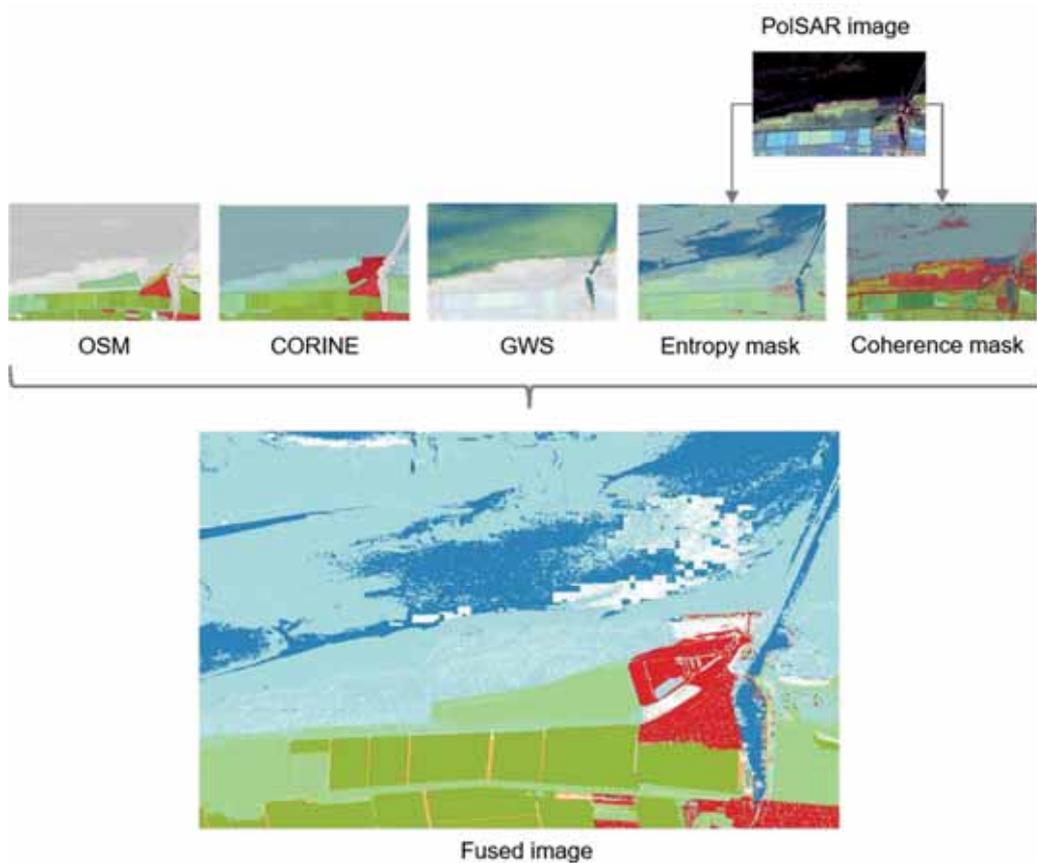


Fig. 2: Fused reference map, used for data labelling

### 3. SAR data analysis

#### 3.1 Land use land cover classification

In order to interpret the processed SAR images, suitable algorithms can be applied to derive statements about soil properties and thus about land use and land cover. Another aim is to distinguish between water areas, wetland areas, meadows, shrubs, forests and human made infrastructure. This shall be achieved by novel classification algorithms based on supervised learning strategies to be developed in the framework of this project. For this purpose, a part of the acquired SAR- data needs to be annotated. In order to accomplish this task in an efficient manner, we fused information given by available land use land cover products and information derived from a previous unsupervised classification of our collected data (Fig. 2). We use the automatically labelled data to train a Random Forest, which assign the desired classes by evaluating various SAR-features. The chosen features are extracted using Eigenvalue decomposition (CLOUDE & POTTIER 1996) as well as the model based Freeman-Durden decomposition (FREEMAN & DURDEN 1993). The resulting pixel-wise classification needs to be post-processed and generalized in order to obtain the desired user-oriented products, such as coastal lines or hydraulic constructions.

### 3.2 Optimization of georeferencing

The usability of such classification products is also depending on the quality of its geocoding. Hence, the optimization of the georeferencing step is one of the main project challenges, since this step specifies the geoposition accuracy of all other products. To realize the optimization of geocoding, we plan to conceive, analyse and evaluate different strategies in terms of resource effort and improvement level. The following options are feasible:

- Usage of corner reflectors, transponders or InSAR DSM control areas,
- Usage of image-to-image matching by integration of UAV data or SAR satellite data, or
- Generation of control points by using radargrammetric image pairs.

First, we will focus on the integration of SAR satellite data to make use of their high precision orbit information. For this purpose, TerraSAR-X data were ordered in parallel to the airborne data acquisition time. Since the satellite passes the area of interest only twice the day at fixed times, the tide height cannot be taken into account as opposed to the airborne data acquisition. Nevertheless, the good match between airborne and spaceborne SAR data shown in Fig. 3 gives reason to hope.



Fig. 3 TerraSAR-X image overlaid by F-SAR data strips and marked reflector positions

### 3.3 Structural edge detection

Besides the classification of land cover, the extraction of 3D structures becomes important for coastal protection. With every ebb and flood, the silt is exposed to high currents and thus moved. The position of tidal creeks with its breaking edges is monitored to observe wide area changes in the mud flat over time. To identify the tidal creeks its breaking edges have to be detected automatically in the SAR data.

In the project, we will develop different approaches to extract those breaking edges. First, we transfer state of the art algorithms (e.g. MANDLBURGER et al. 2016) developed to extract breaking edges from ALS DEMs to InSAR DEMs. Since the generation of InSAR DEMs is a

computationally intensive process, we further examine methods based solely on single look complex data. For this task, SAR specific edge detectors like CFAR need to be integrated. Another attempt is based on applying a SAR specific Canny-like algorithm that is used to extract edges on a single SAR image. A result of this approach is shown in Fig. 4. As additional option, the use of polarimetric features to detect structure lines is planned.

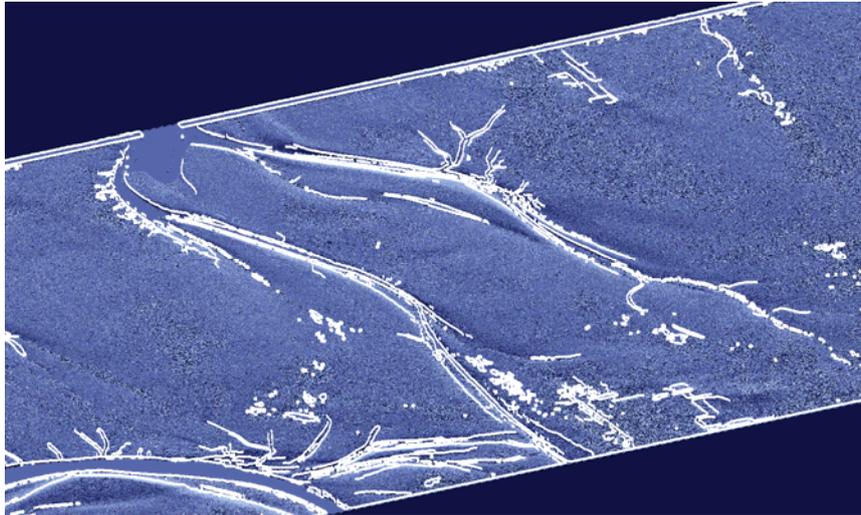


Fig. 4: InSAR DEM overlaid with result of the SAR adopted Canny algorithm

## 4. Acknowledgement

This study is part of the GeoWAM project that is funded by the German Federal Ministry of Transport and Digital Infrastructure (BMVI) within the framework of the Modernity Fund ("mFUND").

## 5. Literature

- CLOUDE, S. R & POTTIER, E., 1996: A review of target decomposition theorems in radar polarimetry. *IEEE Transactions on Geoscience and Remote Sensing* **34**(2), 498-518.
- HORN R., JAEGER M., KELLER M., LIMBACH M., NOTTENSTEINER A., PARDINI M., REIGBER A. & SCHEIBER R., 2017: F-SAR - recent upgrades and campaign activities. 18<sup>th</sup> International Radar Symposium (IRS), 1-10.
- FREEMAN, A. & DURDEN, S. L., 1993: A three component scattering model to describe polarimetric SAR data. *Radar Polarimetry* **1748**, 213-224.
- MANDLBURGER, G., OTEPKA J., BRIESE C., MÜCKE W., SUMMER G., PEIFER N., BALTRUSCH S., DORN C. & BROCKMANN H., 2016: Automatische Ableitung von Strukturlinien aus 3D-PUNKTWOLKEN. *Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e. V* **25**, 131-142.

# Methodische Untersuchungen zum Einsatz von Copernicus-Daten für das Monitoring im Rahmen der EU-Agrarförderung

JULIAN HELGE BRAATZ-NEBEL<sup>1</sup>, HEIKE GERIGHAUSEN<sup>1</sup> & HOLGER LILIENTHAL<sup>1</sup>

*Zusammenfassung: Die Europäische Kommission hat 2018 den Einsatz von Daten der Copernicus-Sentinel-Satelliten im integrierten Verwaltungs- und Kontrollsystem der Gemeinsamen Agrarpolitik beschlossen. Vor diesem Hintergrund wurde in der vorliegenden Arbeit exemplarisch für drei Untersuchungsgebiete in Niedersachsen das Potenzial von Sentinel-1 und Sentinel-2 Daten zur satellitengestützten Kontrolle der gemeldeten Kulturarten untersucht. Im Vergleich zwei verschiedener Ansätze zeigte sich das Random Forest Verfahren der Maximum-Likelihood-Methode leicht überlegen. Dabei erzielte die kombinierte Nutzung von Sentinel-1 und -2 Daten die besten Ergebnisse. Die Kombination beider Daten wirkte sich deutlich positiv auf die Klassifikationsgüte sehr kleiner Flächen (<8 ha) aus.*

## 1 Einleitung

Die Gemeinsame Agrarpolitik (GAP) ist eine der ältesten und finanziell bedeutendsten Aufgabengebiete europäischer Politik, die seit ihrer Verankerung in den Römischen Verträgen von 1957 vielfach reformiert wurde. Ein wichtiger Baustein zu Verwaltung und Kontrolle der gezahlten EU-Beihilfen ist das integrierte Verwaltungs- und Kontrollsystem (InVeKoS) (VO (EU) 1306/2013). Neue bildgebende Technologien sollen nun zur Verringerung der zeit- und kostenintensiven Vor-Ort-Kontrollen (VOK) beitragen und ein flächendeckendes Monitoring aller landwirtschaftlichen Flächen in den Mitgliedstaaten ermöglichen (DVO (EU) 809/2014). Mit den Sentinel-1 (S-1) und Sentinel-2 (S-2) Satelliten des europäischen Erdbeobachtungsprogramms Copernicus stehen für Deutschland alle 1-2 (S-1) bzw. 2-5 (S-2) Tage räumlich hochaufgelöste Bilddaten zweier unterschiedlicher Satellitensysteme frei zur Verfügung (VO (EU) 1159/2013). Sie sollen beispielsweise der Überprüfung der gemeldeten Kulturarten, der Kontrolle der landwirtschaftlichen Tätigkeit oder der Prüfung gezahlter Greeningprämien dienen.

Zahlreiche Autoren haben sich mit der Erkennung von Fruchtarten aus Satellitendaten unter Nutzung verschiedener methodischer Ansätze beschäftigt. Hierzu zählen herkömmliche statistische Schätzverfahren wie die Maximum-Likelihood-Methode (z.B. CHUST et al. 2004; OK et al. 2012), Verfahren des maschinellen Lernens, wie Support Vector Machine, Random Forest, (z.B. DURO et al. 2012; OK et al. 2012; FORKUOR et al. 2014), Entscheidungsbäume (z.B. LÖW et al. 2015) oder künstliche neuronale Netze (z.B. MCNAIRN et al. 2009; KENDUIYWO et al. 2018). Einige Autoren nutzen hierarchische Klassifikationsansätze (z.B. VILLA et al. 2015) oder integrieren zusätzliches Expertenwissen (z.B. WALDHOF et al. 2017).

Ziel der vorliegenden Studie war es, Aufschluss über den Einfluss der Datenzusammenstellung, der Schlaggröße und -form sowie der Anwendung eines räumlichen Puffers auf die Genauigkeit

---

<sup>1</sup> Julius Kühn-Institut, Institut für Pflanzenbau und Bodenkunde, Bundesallee 69, 38116 Braunschweig, E-Mail: [Heike.Gerighausen, Holger.Lilienthal]@julius-kuehn.de

einer Fruchtartenerkennung zu erhalten. Zu diesem Zweck wurden zwei unterschiedliche methodische Ansätze, die klassische Maximum-Likelihood-Methode und das Random Forest Verfahren, gegenübergestellt.

## 2 Material und Methoden

### 2.1 Untersuchungsgebiet

Zur Durchführung der Vor-Ort-Kontrollen werden derzeit in jedem Bundesland mehrere sogenannte *Fernerkundungszonen* festgelegt. Drei dieser Zonen befanden sich 2018 im westlichen Niedersachsen nahe der Gemeinde Kluse (KLUS), im Zentrum Niedersachsens nahe Bad Fallingb. (BAFA) und im südlichen Niedersachsen nahe Adenbüttel (ADBU) mit einer räumlichen Ausdehnung von ca. 30 km × 40 km. Sie dienen als Untersuchungsgebiet in der vorliegenden Studie (Abb. 1).

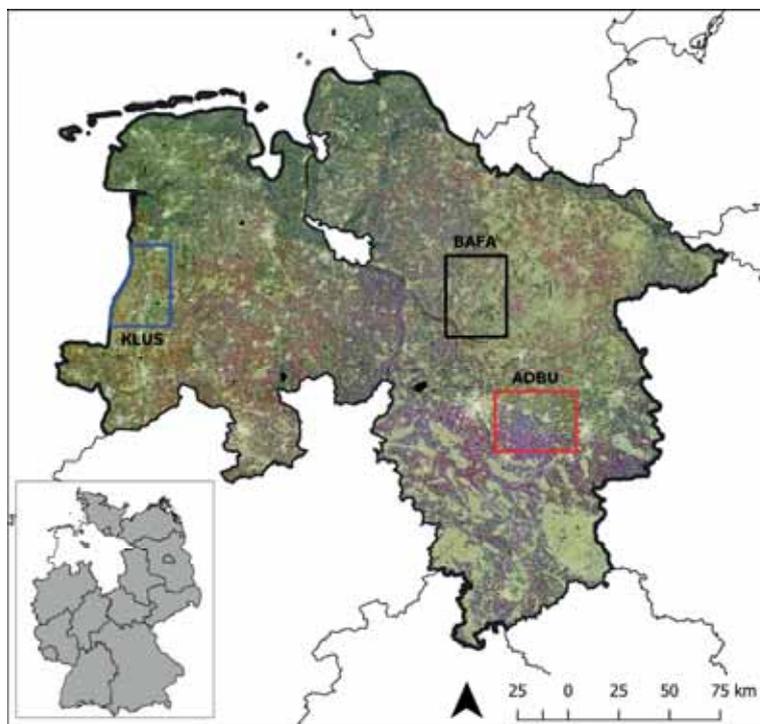


Abb. 1: Sentinel-1 Monatskomposit in RGB-Farbdarstellung und Untersuchungsgebiete in Niedersachsen. Die RGB-Farbkanäle entsprechen den Monaten Juli, Mai, März des Jahres 2018.

Die Nutzung der Landwirtschaftsfläche in diesen Zonen variiert boden- und witterungsbedingt teilweise stark. Während in der Zone ADBU Winterweizen (34 %) gefolgt von Zuckerrüben (20 %) und Mais (12 %) dominieren, bedecken Mais (41 %), Kartoffeln (23 %) und Triticale (9 %) in der Zone KLUS den größten Teil der Flächen. In der Zone BAFA werden vorwiegend Mais (34 %), Winterroggen (13 %) und Sommergerste (9 %) kultiviert. Die mittlere Schlaggröße beträgt in allen drei Zonen weniger als 3 ha (Median), das 0,75-Quantil beträgt 5 ha. Das entspricht ca. 41

% der Ackerfläche in den Zonen ADBU und BAFA sowie 53 % in der Zone KLUS. Alle Angaben beziehen sich auf das Erntejahr 2018.

## 2.2 Satellitendaten

Die Sentinel-1 Daten wurden als Level-1 Ground Range Detected (GRD) Daten vom *Copernicus Open Access Hub* der Europäischen Weltraumorganisation (ESA) heruntergeladen und mit der Software SNAP (Version 6.0) hinsichtlich der Radiometrie und Geometrie und der Topographie korrigiert. Im Ergebnis standen Gamma0 Backscatterdaten für die Polarisationen VH und VV für die Ascending (ASC) und Descending (DSC) Orbits und einer Pixelgröße von 10m x 10m zur Verfügung. Alle Einzeltracks eines Kalendermonats wurden durch Mittelwertbildung zu Monatskompositen zusammengefasst und anschließend zu deutschlandweiten Mosaiken zusammengesetzt. Für die Analysen standen somit jeweils 12 Monatskomposite für das Jahr 2018 als 1) ASC-Daten in VV-Polarisation, 2) ASC-Daten in VH-Polarisation, 3) DSC-Daten in VV-Polarisation und 4) DSC-Daten mit VH-Polarisation zur Verfügung, d.h. insgesamt 48 Datenlayer.

Die Sentinel-2 Daten wurden ebenso vom *Copernicus Open Access Hub* als L1C-Daten bezogen und mit ATCOR-3 (GEOSYSTEMS, 2014) unter Verwendung des DGM10 (BKG, 2015) atmosphären- und terrainkorrigiert. Alle Bänder wurden auf eine Pixelgröße von 10m x 10m skaliert und die Hauptkomponenten sowie der Normalized Difference Vegetation Index (NDVI) berechnet. Es wurden ausschließlich wolkenfreie Aufnahmen aus den Monaten April bis August 2018 verwendet (Tab. 1). Alle Datensätze wurden auf die räumliche Ausdehnung der Untersuchungsgebiete zugeschnitten. Sofern erforderlich erfolgte zuvor eine Mosaikierung einzelner Sentinel-2 Kacheln.

Tab. 1: Übersicht der verwendeten Sentinel-2 Daten aus dem Jahr 2018

Fernerkundungszone	Kachel	Aufnahmetermine
ADBU	UNC/ UND	10.4., 20.4., 20.5., 29.6., 19.7., 24.7., 3.8.2018
BAFA	UND	10.4., 20.4., 20.5., 29.6., 19.7., 24.7., 3.8.2018
KLUS	ULD	18.4., 8.5., 7.6., 2.7., 27.7., 6.8.2018

## 2.3 Klassifikation

Die Kulturartenklassifikation erfolgte mit der Maximum-Likelihood-Methode (ML) (LILLESAND & KIEFER 1994) und dem Random Forest-Verfahren (RF) (BREIMANN 2001). Während der ML-Methode ein parametrisches Schätzverfahren zugrunde liegt, zählt das RF-Verfahren zu den Methoden des maschinellen Lernens und basiert auf zahlreichen unkorrelierten Entscheidungsbäumen, die nach einer bestimmter Art der Randomisierung während des Lernprozesses generiert werden. Für die ML-Klassifikation wurde das semi-automatic classification Plugin (SCP) (CONGEDO 2018), für das RF-Verfahren das dzetsaka Plugin (KARASIAK 2018) in QGIS 3.4 genutzt.

Vor der Klassifikation wurden alle Satellitendaten mit den Schlagvektoren der Antragsdaten aus der Agrarförderung Niedersachsen des Servicezentrum Landentwicklung und Agrarförderung maskiert. Für die verbleibenden Ackerflächen wurden für jeden Sentinel-1 und Sentinel-2 Datensatz (vgl. Kap. 2.2) der Mittelwert pro Schlag bestimmt. Jede Rasterzelle innerhalb eines Schlages wurde anschließend mit dem entsprechenden Mittelwert belegt. Um den Einfluss von Randeffekten und Mischpixeln zu prüfen, wurde dieser Vorgang unter Verwendung eines Puffers von 10m

für alle Datensätze wiederholt. Abschließend wurden unterschiedliche Datenzusammenstellungen generiert (vgl. Tab. 2). Informationen zur zeitlichen Abfolge der Satellitenaufnahmen wurden nicht explizit integriert.

Auf Basis der Antragsdaten wurden die Kulturartenklassen definiert. Kulturarten, die weniger als < 2 % der Fläche bedecken, wurden in einer Klasse „Sonstiges“ zusammengefasst. In QGIS wurde aus diesem Datensatz über eine randomisierte Auswahl ein Trainings- und ein Validationsdatensatz im Verhältnis 30/70 erstellt. Die Klassifikationsgenauigkeiten wurden mithilfe einer Konfusionsmatrix ermittelt.

Um den Einfluss der Schlaggröße auf die Gesamtgüte der Klassifikation zu analysieren, wurden 10.000 Schläge (ca. 69 % aller Validationsflächen) zufällig ausgewählt. Diese wurden aufsteigend nach der Größe sortiert und in 10 Klassen zu je 1.000 Schlägen gruppiert. Für jede dieser Klassen wurden Konfusionsmatrizen zur Ermittlung der größenabhängigen Klassifikationsgüte erstellt. Zusätzlich wurde für jede Klasse die mittlere Schlaggröße und die Standardabweichung ermittelt. Als Maß für die Flächenform wurde für alle Flächen zusätzlich die Kompaktheit nach POLSBY & POPPER (1991) berechnet.

### 3 Ergebnisse und Diskussion

Für die Fernerkundungszone ADBU wurde mehrere Klassifikationen mit unterschiedlichen Datenzusammenstellungen (vgl. Tab. 2) mit bis zu 69 Bändern durchgeführt (Tab. 3). Unabhängig vom Eingangsdatensatz und der Vorverarbeitung erzielte die RF-Methode in allen Fällen eine etwas höhere Gesamtgüte als das ML-Verfahren. Das RF-Verfahren zeigte sich aber vor allem in Bezug auf die Prozessierungsdauer deutlich überlegen. Eine Klassifikation mit 69 Bändern war mit der ML-Methode nicht möglich, da der Prozess nach 38 Minuten durch das Programm abgebrochen wurde. Alle Zeitangaben in Tab. 3 beziehen sich dabei auf einen Arbeitsplatzrechner mit einem Intel Core TM i3-6100CPU, 3.7GHz mit 16Gb RAM und einen Windows 10/ 64bit Betriebssystem.

Einen wesentlichen Einfluss auf die Höhe der Klassifikationsgüte (Gesamtgüte) hatte die Art und die Anzahl der Eingangsdaten. Die besten Ergebnisse wurden bei Verwendung von Sentinel-1 und -2 Daten erzielt. Die separate Verwendung von Sentinel-1 oder Sentinel-2 resultierte in etwas besseren Ergebnissen bei Nutzung der Sentinel-1 Daten. Dabei erwies sich insbesondere eine Kombination von ASC-VH und DSC-VV als zielführend. Die Verwendung eines räumlichen Puffers führte nur zu einer geringfügigen Verbesserung der Gesamtgüte. Sie verbesserte vor allem die Trefferquote der Klasse *Sonstiges*. Flächen dieser Klasse wiesen eine mittlere Flächengröße von weniger als 0,7 ha auf.

Die Nutzer- und Herstellergenauigkeiten (nicht dargestellt) variierten je nach Fruchtart. Für die Klassifikation mit der höchsten Gesamtgüte (RF-69p) wurden sehr gute Ergebnisse bei der Erkennung von Winterweizen, Winterraps und Hackfrüchten einschließlich Mais erzielt. Schwierigkeiten traten bei der Identifikation von Sommerweizen auf, der in vielen Fällen der Klasse Winterweizen zugeordnet wurde. Außerdem kam es zu Fehlzugeisungen zwischen Winterroggen, Wintergerste und Winterweizen.

In den Fernerkundungszonen KLUS und BAFA konnten mit dem RF-Verfahren und kombinierten Sentinel-1 und -2 Daten Gesamtgenauigkeiten von 94,8% und 89,5% erreicht werden, wobei die

niedrigere Klassifikationsgüte für die Zone BAFA durch einen hohen Anteil der Klasse Sonstiges (12%) zur erklären ist.

Tab. 2: Übersicht über die Datenzusammenstellungen (DZ) für die Fernerkundungszone ADBU

DZ	Anzahl der Bänder	Eingangsdaten
I	69	S-1 (48 Bänder): ASC-VV, ASC-VH, DSC-VV, DSC-VH S-2 (14 Bänder): 1. + 2. Hauptkomponente, NDVI
II	24	S-1 (24 Bänder): ASC-VH, DSC-VV
II	14	S-2 (14 Bänder): 1.+2. Hauptkomponente

Tab. 3: Ergebnisse verschiedener Klassifikationsvarianten für die Fernerkundungszone ADBU; RF ... Random Forest, ML... Maximum-Likelihood, S1 ... Sentinel-1, S2... Sentinel-2.

Name	DZ	Methode	Anzahl der Bänder	Daten	Puffer (10m)	Gesamtgüte [%]	Dauer [min] †)
RF-69p	I	RF	69	S1+S2	ja	95,7	8
ML-69p	I	ML	69	S1+S2	ja	- <sup>†)</sup>	- <sup>†)</sup>
RF-69	I	RF	69	S1+S2	nein	95,4	8
RF-24p	II	RF	24	S1	ja	92,8	6
ML-24p	II	ML	24	S1	ja	92,1	31
RF-14p	III	RF	14	S2	ja	91,1	2
ML-14p	III	ML	14	S2	ja	89,6	16

†) Abbruch durch das Programm nach 38 min; ‡) Die Zeiten beziehen sich auf einen Arbeitsplatzrechner mit einem Intel Core TM i3-6100CPU, 3.7GHz mit 16Gb RAM, Windows 10 64bit

Abb. 2 zeigt die Klassifikationsgüte (nur RF) in Abhängigkeit von der Schlaggröße und der Datenzusammenstellung. In allen Fällen steigt die Klassifikationsgüte mit Zunahme der Schlaggröße logarithmisch an. Flächen kleiner 8 ha werden deutlich besser bei Verwendung von Sentinel-1 und Sentinel-2 Daten erfasst. Die Klassifikationsgüte für die kleinste Größenklasse (<0,4 ha) beträgt mit Sentinel-1 und -2 Daten 87 %, bei Nutzung von Sentinel-2 nur noch 83 % und auf Basis der Sentinel-1 Daten 79 %. Für Flächen >8 ha wurden indes bessere Ergebnisse mit Sentinel-1 als mit Sentinel-2 erreicht. Bei Flächen größer 10 ha spielt die Art der Datenzusammenstellung eine zunehmend geringere Rolle. Bislang liegen keine Erkenntnisse vor, ob und inwiefern sich die Heterogenität landwirtschaftlicher Bestände in ausgeräumten Agrarlandschaften wie beispielsweise in Mecklenburg-Vorpommern mit sehr großen Nutzungseinheiten auf den verwendeten Klassifikationsansatz auswirkt. Die ist in weiterführenden Studien zu prüfen. Bei kleinen Flächen lassen erste Analysen einen signifikanten Zusammenhang zwischen Klassifikationsgüte und Kompaktheit der Schläge erkennen.

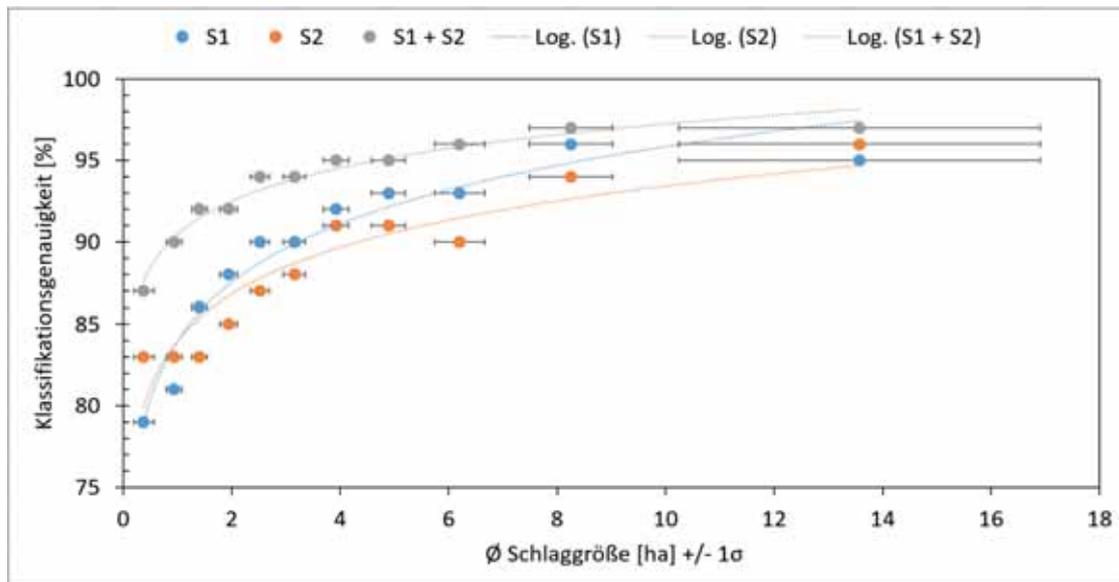


Abb. 2: Klassifikationsgenauigkeit in Abhängigkeit von der Schlaggröße und verwendeten Eingangsdaten: blau: Sentinel-1 (24 Bänder, RF-24p), orange: Sentinel-2 (14 Bänder, RF-14p), grau: Sentinel-1 und Sentinel-2 (69 Bänder, RF-69p).

## 4 Fazit & Ausblick

Die Ergebnisse dieser Studie zeigen, dass für ein flächendeckendes Monitoring der Kulturarten im Rahmen des integrierten Verwaltungs- und Kontrollsystem der EU die kombinierte Verwendung von Sentinel-1 und Sentinel-2 Daten anzustreben ist. Dies gilt nicht nur im Hinblick auf Einschränkungen in der Datenverfügbarkeit durch Wolken (S-1) und Starkniederschläge (S-2), sondern vor allem in Bezug auf die Überprüfung gezahlter Beihilfen in stark kleinstrukturierten Agrarräumen wie beispielsweise in Hessen, Baden-Württemberg oder Bayern. In diesem Zusammenhang sind weitere Untersuchungen zum Einfluss der Schlagform erforderlich.

Darüber hinaus ist die Monitoringfähigkeit einer Vielzahl anderer Prüfkriterien wie die Mindesttätigkeit auf Brachen oder die landwirtschaftliche Tätigkeit auf Grünland zu testen, um das Potenzial der Sentinel-Satelliten voll auszuschöpfen.

## 5 Literaturverzeichnis

BKG, 2015: Digitales Geländemodell Gitterweite 10m, DGM10. Bundesamt für Kartographie und Geodäsie, URL: [https://sg.geodatenzentrum.de/web\\_public/gdz/dokumentation/deu/dgm10.pdf](https://sg.geodatenzentrum.de/web_public/gdz/dokumentation/deu/dgm10.pdf)

BREIMANN, L., 2001: Random Forests, Machine Learning **45**, 5-32.

CHUST, G., DUCROT, D. & PRETUS, J.L.L., 2004: Land cover discrimination potential of radar multitemporal series and optical multispectral images in a Mediterranean cultural landscape. International Journal of Remote Sensing **25**, 3513-3528.

CONGEDO, L., 2018: Semi-Automatic Classification Plugin Documentation - Release 5.3.6.1.

- DURO, D.C., FRANKLIN, S.E. & DUBÉ, M.G., 2012: A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment*, 118, 259-272.
- DVO (EU) 809/2014: Durchführungsverordnung (EU) 2018/746 der Kommission vom 18. Mai 2018, Amtsblatt der Europäischen Union, 22.5.2018.
- FORKUOR, G., CONRAD, C., THIEL, M., ULLMANN, T. & ZOUNGRANA, E., 2014: Integration of optical and synthetic aperture radar imagery for improving crop mapping in northwestern Benin, West Africa. *Remote Sensing* 6, 6472-6499.
- GEOSYSTEMS, 2014: ATCOR for IMAGINE 2016. User manual ATCOR 2 und ATCOR 3. Manual Version 29/07/2016, Geosystems GmbH.
- KARASIAK, N. & PERBET, P., 2018: Remote Sensing of Distinctive Vegetation in Guiana Amazonian Park. In: *QGIS and Applications in Agriculture and Forest*, 215-245.
- KENDUIYWO, B., K., BARGIEL, D. & SOERGEL, U. 2018: Crop-type mapping from a sequence of Sentinel 1 images. *International Journal of Remote Sensing* 39(19), 6383-6404.
- LILLESAND, T.M. & KIEFER, R.W., 1994: *Remote Sensing and Image Interpretation*. 3rd. Edition, John Wiley & Sons, Inc.
- LÖW, F., CONRAD, C. & MICHEL, U., 2015: Decision fusion and non-parametric classifiers for land use mapping using multi-temporal RapidEye data. *ISPRS J. Photogramm. Remote Sensing* 108, 191-204.
- McNAIRN, H., CHAMPAGNE, C., SHANG, J., HOLMSTROM, D. & REICHERT, G., 2009: Integration of optical and Synthetic Aperture Radar (SAR) imagery for delivering operational annual crop inventories. *Journal of Photogrammetry and Remote Sensing* 64, 434-449.
- OK, A.O., AKAR, O. & GUNGOR, O., 2012: Evaluation of random forest method for agricultural crop classification. *European Journal of Remote Sensing* 45, 421-432.
- VILLA, P., STROPPIANA, D., FONTANELLI, G., AZAR, R. & BRIVIO, P.A., 2015: In-season mapping of crop type with optical and X-band SAR data: a classification tree approach using synoptic seasonal features. *Remote Sensing* 7, 12859-12886.
- VO (EU) 1306/2013: Verordnung (EU) Nr. 1306/2013 des Europäischen Parlaments und des Rates vom 17. Dezember 2013, Amtsblatt der Europäischen Union, 20.12.2013.
- VO (EU) 1159/2013: Delegierte Verordnung (EU) Nr. 1159/2013 der Kommission, Amtsblatt der Europäischen Union, 19.11.2013.
- WALDHOFF, G., LUSSEM, U. & BARETH, G., 2017: Multi-Data Approach for remote sensing-based regional crop rotation mapping: A case study for the Rur catchment, Germany. *International Journal of Applied Earth Observation and Geoinformation* 61, 55-69.

# Impacts of Climatic Oscillations on Precipitation in an Extended Mediterranean Area

LAURA CROCETTI<sup>1</sup> & WOUTER DORIGO<sup>1</sup>

*Abstract: Oceanic-atmospheric oscillation patterns, described by so-called climate modes, control a large portion of the variability in the water cycle. However, the relation between ocean-atmospheric oscillations and hydrology is still poorly understood. Better knowledge about these connections is needed to accurately predict the impact of climate change on hydrology. In this study, a supervised machine learning approach based on least absolute shrinkage and selection operator (LASSO) models is used to identify and disentangle the impact of 17 major climate modes on precipitation anomalies over an extended Mediterranean area. Analysing the results of the winter model and using precipitation from the Climate Research Unit (CRU) as a response variable, shows that up to 70% of the variability can be explained by ocean-atmospheric oscillations. These results help to improve the general understanding of how individual climate modes affect different parts of the Mediterranean area.*

## 1 Introduction

In recent decades, global warming has led to significant changes in the processes of our climate system, including the cycles of energy, water, and biogeochemicals. A thorough understanding of these cycles is needed to assess the potential implications of global warming on our future climate. A large portion of the natural variability of climate is controlled by the coupled state of ocean and atmosphere, the so-called ocean-atmospheric oscillations. These climate modes characterized by a repeating time-space pattern are important drivers of the hydrology and can even influence regions that are far apart. Therefore, they are known as teleconnections (WALLACE & GUTZLER 1981; WANG & SCHIMMEL 2003). Each climate mode is expressed through its corresponding Climate Oscillation Index (COI) that describes the state of the atmospheric-oceanic oscillations in numbers. One of the most relevant teleconnections affecting the Mediterranean hydrology is the well-known North Atlantic Oscillation (NAO) (BARNSTON & LIVEZEY 1987). Its index is defined through the difference between the normalized sea-surface pressure over Lisbon, Portugal and Reykjavik, Iceland. However, not only the NAO influences the Mediterranean climate but also other large-scale phenomena, like the East Atlantic Oscillation (EA), the East Atlantic West Russia Pattern (EAWR) or the Scandinavian Pattern (SCAND).

A lot of studies investigate the relationship between individual climate modes and their impact on the Mediterranean climate (MIKHAILOVA & YUROVSKY 2016; BLADÈ et al. 2012; CHRONIS et al. 2011). It is shown that teleconnection patterns exert a significant impact on climatic drivers (KRICHAK et al. 2014; MARTENS et al. 2018). Although several climate modes show co-varying behaviour, i.e. individual climate modes are highly correlated, most of the studies do not address

---

<sup>1</sup> TU Wien, Department of Geodesy and Geoinformation, Wiedner Hauptstraße 8, A-1040 Wien  
E-Mail: [laura.crocetti, wouter.dorigo]@geo.tuwien.ac.at

the coupled interactions between individual climate modes. However, these interactions are important and their consideration improves the accuracy of the results compared to correlations. Therefore, further research is necessary to improve the general understanding of the connection between ocean-atmospheric oscillations and precipitation anomalies.

The objective of this study is to reveal the dominant climate modes controlling precipitation anomalies in an extended Mediterranean area. Therefore, a supervised machine learning approach, called least absolute shrinkage and selection operator (LASSO) regression is used. It is a data-driven multivariate method that analyses the joint behaviour of multiple variables concurrently. This means the co-varying behaviour of the individual climate modes is considered and the impact of many climate modes on precipitation anomalies is determined simultaneously.

## 2 Data

### 2.1 Climate modes

Climate modes describe the state of the ocean-atmospheric circulations and follow distinct repeating spatio-temporal patterns. Each climate mode has a corresponding Climate Oscillation Index (COI) that monitors the strength and the temporal pattern of the individual climate mode. In this study, 17 major climate modes, expressed through their associated COI, listed in Table 1 are used as predictive features in a LASSO regression framework. The COIs have a monthly temporal resolution. The raw time series of the COIs get pre-processed. First, the multi-year climatology is eliminated to remove the seasonal variations of the signal, then the COIs are standardized by months and the whole time series gets normalized by dividing it by the l2-norm.

Tab. 1: Climate modes used in this study with their corresponding COI

Climate mode	Climate Oscillation Index (COI)
Atlantic Meridional Mode (AMM)	AMM sea-surface temperature (AMMSST) index
Atlantic Multidecadal Oscillation (AMO)	AMO index
Indian Ocean Dipole (IOD)	Dipole Mode Index (DMI)
East Atlantic Pattern (EA)	EA index
East Atlantic West Russia Pattern (EAWR)	EAWR index
East Pacific North Pacific Pattern (EPNP)	EPNP index
Northern Annular Mode (NAM)	NAM index
North Atlantic Oscillation (NAO)	NAO index
Pacific Decadal Oscillation (PDO)	PDO index
Polar Eurasia Pattern (PEA)	PEA index
Pacific-North American Pattern (PNA)	PNA index
Southern Annular Mode (SAM)	SAM index
Scandinavian Pattern	SCAND index
El Niño-Southern Oscillation (ENSO)	Southern Oscillation Index (SOI)
Tropical Northern Atlantic Dipole (TNA)	TNA index
Tropical Southern Atlantic Dipole (TSA)	TSA index
West Pacific Pattern (WP)	WP index

### 2.2 Precipitation data

The target variable in the LASSO framework are precipitation anomalies. In this study, precipitation provided by the Climatic Research Unit (CRU) TS v.3.23 (HARRIS et al. 2013) is used. It is a gridded high-resolution climate dataset from monthly observations at meteorological

stations across the world’s land areas. It is provided on a regular latitude-longitude grid with a common spatial resolution of 0.5°. The LASSO regression is trained by the anomalies of the precipitation data instead of the raw time series to remove seasonal signals. First, the linear trend is removed, then the anomalies are calculated based on the multi-year climatology.

### 3 Method – LASSO regression

For this study, a supervised machine learning algorithm, called least absolute shrinkage and selection operator (LASSO) regression (TIBSHIRANI 1996) is used to determine and disentangle the impact of climate modes on precipitation anomalies. In this framework, 17 COIs (see Table 1) are used as predictive features, and anomalies of precipitation of CRU are used as the target variable. One benefit of the LASSO regression is, that it takes the cross-correlations of the features into account. This is important since some climate modes are highly correlated, as shown in Figure 1. Furthermore, the method uses feature selection and regularization to quantify the impact of each feature on the target variable. The regularization helps to prevent overfitting by setting the coefficients of weak features to zero. Thus, only a subset of features remain, making the final predictive model simpler and improving the prediction accuracy of the model. Compared to other machine-learning algorithms like random forest, results from the LASSO regression are easier to understand and evaluate.

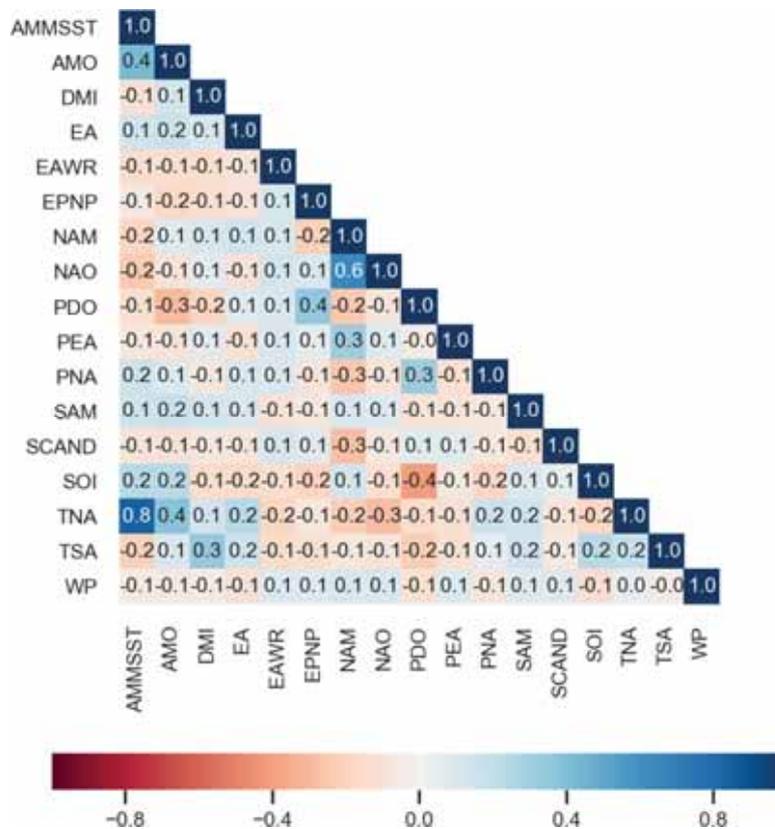


Fig. 1: Cross-correlation of COI anomalies

The LASSO regression determines the estimated regression coefficients  $\hat{\beta}$  by minimizing a penalized residual sum of squares (TIBSHIRANI 1996).

$$\hat{\beta} = \operatorname{argmin} \left\{ \sum_{i=0}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \alpha \sum_{j=1}^p |\beta_j| \right\} \quad (1)$$

In equation (1),  $\hat{\beta}$  is the  $p$ -dimensional vector with the estimated regression coefficients,  $n$  the number of training samples in the dataset,  $y_i$  the value of the target variable in sample  $i$ ,  $p$  the number of features,  $x_{ij}$  the value of features  $j$  in sample  $i$ , and  $\alpha$  a hyper-parameter controlling the amount of shrinkage (MARTENS et al. 2018).

The estimated regression coefficients  $\hat{\beta}$  are used to predict the target variable  $y_{pred}$ ,

$$y_{pred} = \sum_{feature=1}^{17} \left\{ \sum_{lag=0}^5 (x_{feature\ lag} \cdot \hat{\beta}_{feature\ lag}) + bias \right\} \quad (2)$$

where  $x$  is the feature vector for a given time lag and  $\hat{\beta}$  is the vector with the corresponding estimated regression coefficients.

To validate the predictive performance of the model, the coefficient of determination  $R^2$  is calculated. It is defined as the squared correlation between the target variable  $y$  and the predicted target variable  $y_{pred}$ .

$$R^2 = \operatorname{corr}(y, y_{pred})^2 \quad (3)$$

## 4 Experimental set up

In this study, analysis over an extended Mediterranean area spanning from 28.5°N to 56.5°N and from 10°W to 46°E are carried out. A LASSO regression model is fitted individually for every grid point using 17 major COIs as features and precipitation anomalies as the target variable.

Figure 2 shows the concept of the LASSO regression. The left block is the target variable vector  $y$  and the right block represents the feature matrix  $X$ . Six time lags ranging between zero and five months are introduced for every feature to take a potential delayed response of precipitation anomalies to ocean-atmospheric oscillations into account. This results in a total of 102 features (17 COIs times 6 time lags). The overlapping period for the COIs and the anomalies of the CRU precipitation dataset is from 1957 to 2014. This period is used for the analysis.



Fig. 2: Basic framework of the LASSO regression

Besides developing models based on the entire available time series, additional models for each season are calculated. The winter model includes values from December, January and February, the spring model includes March, April and May, the summer model includes June, July and August and the autumn model includes September, October, and November.

To optimize and validate the predictive LASSO regression models, two five-fold cross-validations are applied. One cross-validation is used to determine the regularization parameter  $\alpha$  and the second determines the coefficient of determination  $R^2$ . To ensure that the results are robust and trustful, a significance test using the Benjamini-Hochberg procedure (BENJAMINI & HOCHBERG 1995) at a significance level of 95% is applied and areas which do not fulfil this test are masked. Furthermore, to increase the predictive performance of the model and make the results more robust, the  $3 \times 3$  neighbourhood is included. The framework is extended by the data of the eight neighbouring grid points. It is expected that by including the neighbourhood information and, therefore, extending the time series by a factor of nine, more information can be extracted from the time series. Thus, if one individual grid point does not provide enough information to create a good model, at least the neighbouring grid points can help to improve the prediction.

## 5 Results and Discussion

Figure 3 depicts the explained variance  $R^2$  of the precipitation anomalies obtained by the LASSO regression for the winter model from 1957 to 2014. In the left plot, the LASSO regression is trained for each pixel individually, while in the right plot the  $3 \times 3$  neighbourhood is taken into account. By comparing the two maps it can be seen that the explained variance  $R^2$  gets significantly higher when the  $3 \times 3$  neighbourhood is included and variabilities up to 70% can be estimated. The spatial pattern detected look very similar, however, the explained variances  $R^2$  are significantly higher for the model including the  $3 \times 3$  neighbourhood. The reason for the higher coefficient of determination  $R^2$  is, that the impact of the noise in the data is reduced due to the nine times longer time series. Thus, the time series gets more robust.

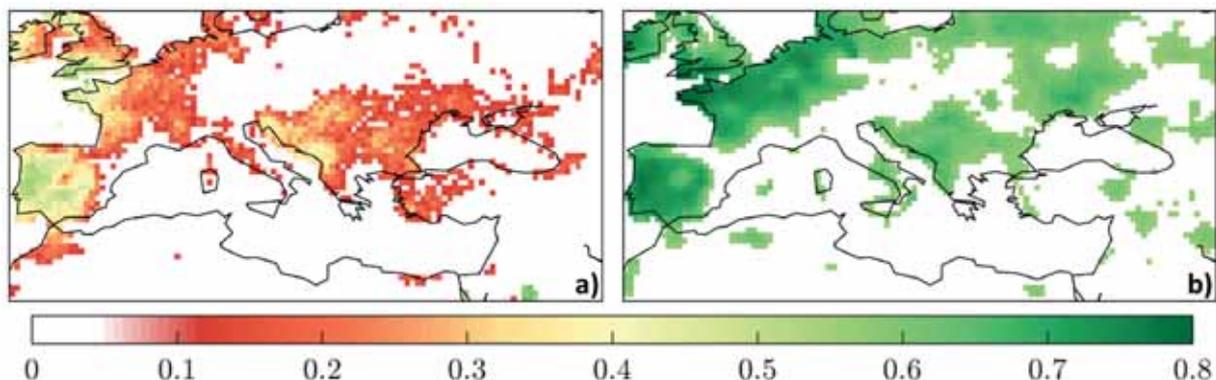


Fig. 3: Explained variance  $R^2$  of the LASSO regression for the winter model. Areas that are classified as non-significant are marked white. a) the model is trained for each pixel individually, b) additional  $3 \times 3$  neighbourhood is taken into account

Precipitation anomalies in the winter model show a high sensitivity to the climate modes. In particular, the Iberian Peninsula, Ireland, the United Kingdom, France, Belgium, the Netherlands, as well as some parts of Italy, and the eastern coast of the Adriatic Sea are influenced by the modes of climate variability. The white areas in the maps in Figure 3 have been declared as non-significant by the permutation test using the Benjamini-Hochberg procedure at a significance level of 95%.

The impact of each COI is investigated by calculating the explained variance for each COI  $R_{COI}^2$ . This value is calculated by taking the six regression coefficients  $\hat{\beta}$  of one individual COI from the LASSO regression and based on these coefficients the predicted target variable  $y_{pred_{COI}}$  and finally the explained variance  $R_{COI}^2$  is calculated.

Figure 4 visualizes the explained variance for the winter model for the five COIs  $R_{COI}^2$  detected having a significant impact on precipitation anomalies in this area, i.e. the East Atlantic Pattern (EA), the East Atlantic West Russia Pattern (EAWR), the Northern Annular Mode (NAM), the North Atlantic Oscillation (NAO), and the Scandinavian Pattern (SCAND).

The explained variance of each COI  $R_{COI}^2$  is considerably lower than the explained variance  $R^2$  of the complete model. This can be explained by the now missing interaction of the individual COIs.

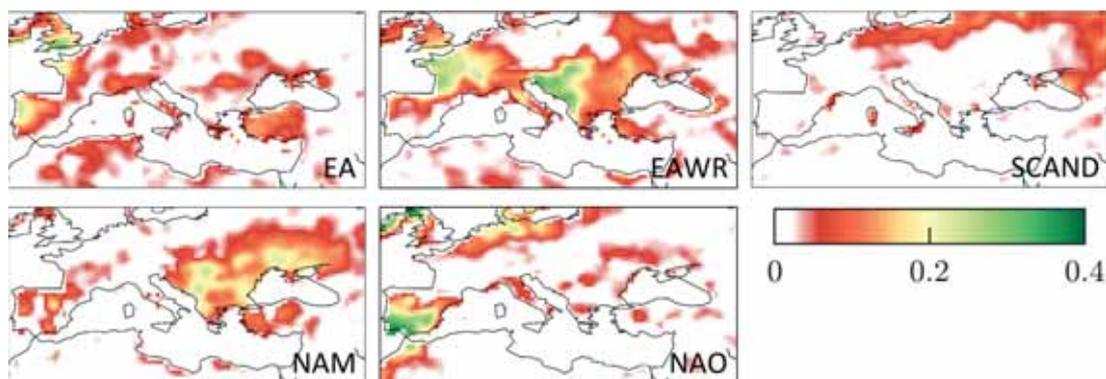


Fig. 4: Impact of the five climate modes identified having the highest impact on precipitation anomalies  $R_{COI}^2$  for the winter model with additional neighbourhood information. The respective abbreviations of each climate mode are written in the bottom right corner.

To better understand how the LASSO model works and what the coefficient of determination  $R^2$  shows, the results of one randomly chosen grid point are discussed in more detail. The grid point is located at a latitude of  $38.75^\circ\text{N}$  and  $7.25^\circ\text{W}$ . It lies in the south-western region of the Iberian Peninsula.

Figure 5 shows the results of a step-wise addition of COIs to the LASSO regression model for that grid point. The LASSO model is trained by using all 17 COIs. In case of the no neighbourhood model, for this specific grid point only three features, namely EA, NAO, and NAM are selected by the automatic feature selection of the LASSO regression to build the final model.

The first plot in Figure 5 shows the result, if only the regression coefficients  $\hat{\beta}$  from the full LASSO regression model of one feature, namely EA, would be used to predict the precipitation anomalies. The corresponding coefficient of determination would result in  $R^2=0.12$ . By adding the regression coefficients  $\hat{\beta}$  of the second feature, i.e. NAO, the coefficient of determination would increase

considerably to  $R^2=0.54$ . Adding the third feature, i.e. NAM, shows the final LASSO result with a  $R^2$  of 0.55. The pattern of the predictions  $y_{pred}$  of the final LASSO model follow the temporal patterns of the target variable  $y$  well but the amplitude is lower.

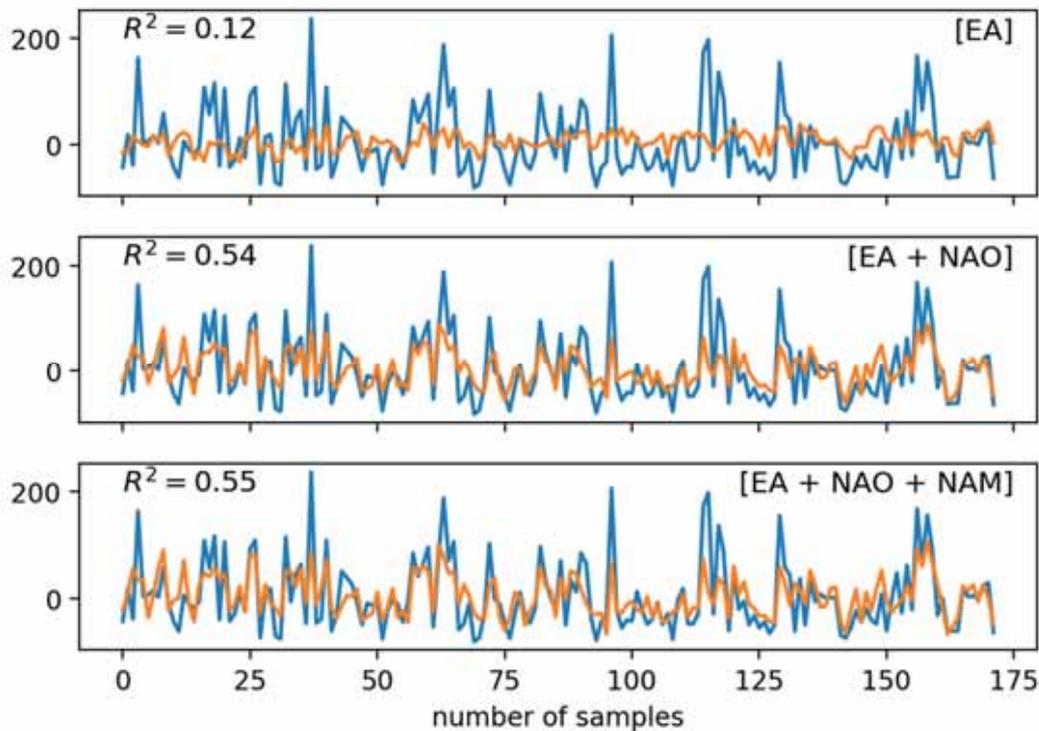


Fig. 5: Step-wise addition of COIs building the LASSO winter model for precipitation anomalies at  $38.75^{\circ}\text{N}$  and  $7.25^{\circ}\text{W}$  without including the neighbourhood. Blue: target variable  $y$ , orange: predicted target variable  $y_{pred}$ . Only three COIs are selected by the LASSO regression model.

Figure 6 shows the same step-wise addition of individual features as Figure 5 but includes the  $3 \times 3$  neighbourhood. By adding the neighbourhood information the LASSO regression uses a lot more features to build its prediction model. Therefore, three to four features are added simultaneously in Figure 6 to reduce the number of necessary plots. The final LASSO regression model is the last plot in Figure 6. It is clearly visible, that the final result including the neighbourhood corresponds better to the target variable  $y$  as it is the case in Figure 5. This time the temporal patterns and amplitudes are represented better. This can also be verified by the higher explained variance  $R^2$  of 0.77. However, the contribution per COI is lower.

Although this is only the result of one grid point, the same can be seen at nearly every location. As shown in Figure 3 the explained variance  $R^2$  is almost always higher when adding the information of the neighbourhood.

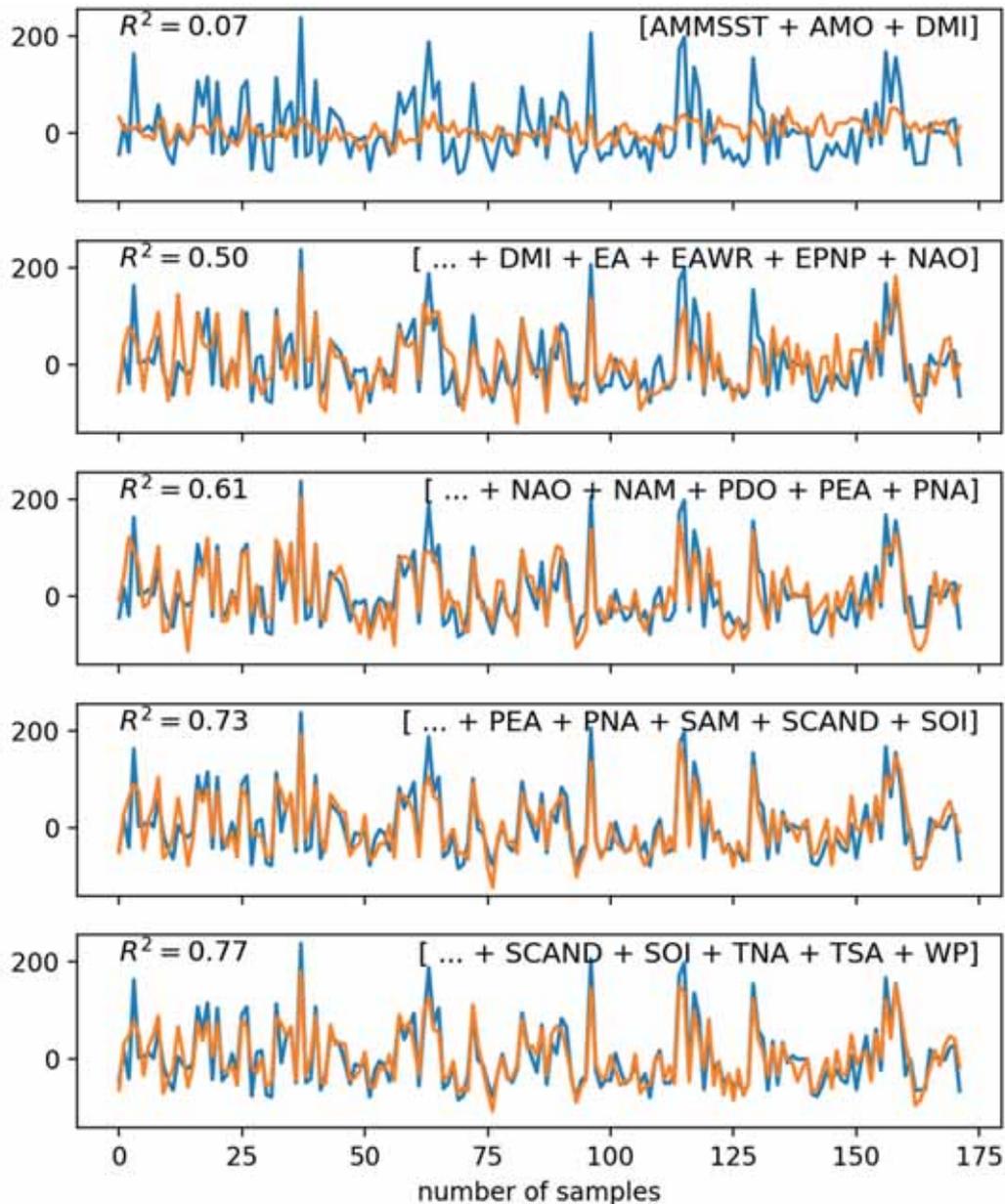


Fig. 6: Step-wise addition of COIs building the LASSO winter model for precipitation anomalies at  $38.75^{\circ}\text{N}$  and  $7.25^{\circ}\text{W}$  including the neighbourhood. Blue: target variable  $y$ , orange: predicted target variable  $y_{pred}$

## 6 Conclusion

In this study, the impact of 17 climate modes on precipitation anomalies over an extended Mediterranean area is analysed. The results show that the variability in precipitation anomalies are strongly affected by ocean-atmospheric oscillations.

A supervised learning approach, called LASSO regression is used to determine the impact of individual climate modes. The method uses feature selection and regularization and can help

revealing the dominant modes of climate variability by disentangling their complex impacts and interactions. The analysis is performed in two ways, once by analysing the time series of each grid point individually and, furthermore, by adding the information of the 3x3 neighbourhood. If the LASSO framework is extended by the 3x3 neighbourhood, the model improves significantly, resulting in a high explained variance  $R^2$ . Additionally, seasonal models are calculated which lead to a significantly higher explained variance  $R^2$  compared to the full-year model that uses all months of a year. Especially, the winter model shows a high coefficient of determination  $R^2$ . In some regions up to 70% of the precipitation anomalies can be explained by the combination of specific climate modes. The LASSO regression detected that EA, EAWR, NAM, NAO, and SCAND have the biggest impact, meaning that these climate modes are mostly responsible for inter-annual precipitation anomalies during winter.

Based on these results, further research can be carried out to improve the understanding of the relation between climatic oscillations and hydrology.

## 7 References

- BARNSTON, A. G. & LIVEZEY, R. E., 1987: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Monthly Weather Review*, **115**(6), 1083-1126.
- BENJAMINI, Y. & HOCHBERG, Y., 1995: Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, **57**(1), 289-300.
- BLADÈ, I., LIEBMANN, B., FORTUNY, D. & OLDENBORGH, G. J., 2012: Observed and simulated impacts of the summer NAO in Europe: Implications for projected drying in the Mediterranean region. *Climate Dynamics*, **39**(3), 709-727.
- CHRONIS, T., RAITOS, D. E., KASSIS, D. & SARANTOPOULOS, A., 2011: The summer North Atlantic Oscillation influences on the eastern Mediterranean. *Journal of Climate*, **24**(21), 5584-5596.
- HARRIS, I., JONES, P., OSBORN, T. & LISTER, D., 2013: Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 dataset. *International Journal of Climatology*, **34**(3), 623-642.
- KRICHAK, S. O., BREITGAND, J. S., GUALDI, S. & FELDSTEIN, S. B., 2014: Teleconnection-extreme precipitation relationships over the Mediterranean region. *Theoretical and Applied Climatology*, **117**(3), 679-692.
- MARTENS, B., WAEGEMAN, W., DORIGO, W.A., VERHOEST, N. E. C. & MIRALLES, D. G., 2018: Terrestrial evaporation response to modes of climate variability. *NPJ Climate and Atmospheric Science*, **1**(1), 1-7.
- MIKHAILOVA, N. & YUROVSKY, A., 2016: The East Atlantic Oscillation: Mechanism and impact on the European climate in winter. *Physical Oceanography*, (4), 25-33.
- TIBSHIRANI, R., 1996: Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B, (Methodological)*, **58**(1), 267-288.
- WALLACE, J. M. & GUTZLER, D. S., 1981: Teleconnections in the geopotential height field during the northern hemisphere winter. *Monthly Weather Review*, **109**(4), 784-812.
- WANG, G. & SCHIMEL, D., 2003: Climate Change, Climate Modes, and Climate Impacts. *Annual Review of Environment and Resources*, **28**(1), 1-28.

# Entwicklung und Evaluierung eines Web-Tools zur bezahlten crowd-basierten Erfassung von Bäumen aus 3D-Punktwolken

YIFEI YIN<sup>1</sup>, MICHAEL KÖLLE<sup>2</sup> & VOLKER WALTER<sup>2</sup>

*Zusammenfassung: Mit dem rapiden Anstieg der Leistungsfähigkeit der Methoden des maschinellen Lernens wächst in den letzten Jahren auch zunehmend das Interesse an großen Mengen qualitativ hochwertiger Trainingsdatensätze. Um solche Datensätze möglichst gänzlich ohne den Einsatz eines Experten abzuleiten, kann bezahltes Crowdsourcing genutzt werden. Der Fokus der Arbeit liegt auf der bezahlten crowd-basierten Erfassung von Bäumen aus 3D-Punktwolken. Diese werden mittels eines implementierten Web-Tools durch minimal umschließende Zylinder annotiert. Die Qualität der Erfassungen wird durch einen Vergleich mit Referenzdaten bewertet. Um die Übereinstimmung mit diesen zu steigern, wird weiterhin das Potential einer Mehrfacherfassung aufgezeigt, um im Sinne der „Wisdom of the Crowd“ die am best-angepassten Zylinder zu rekonstruieren.*

## 1 Einleitung

Bei der Entwicklung überwachter Klassifikationsansätze kommen manuell annotierte Datensätze als Trainingsdaten zum Einsatz, um die zu trainierenden unbekannt Parameter möglichst optimal zu erlernen bzw. zu bestimmen. Dabei spielen die Menge und insbesondere die Qualität der Trainingsdatensätze eine wichtige Rolle. Zur Gewinnung solcher Datensätze bietet bezahltes Crowdsourcing die Möglichkeit, einfache Aufgaben wie Datenannotation ohne den Einsatz eines Experten an eine undefinierte Menge von Arbeitskräften auszulagern (SU et al. 2012). Die finanzielle Entlohnung pro Aufgabe liegt in der Regel im Bereich von einigen Cents, weshalb die Crowdworker auch oft als Microworker bezeichnet werden (HOSSFELD et al. 2012).

Viele bestehende Crowdsourcing-Aufgaben im Bereich der Geoinformation konzentrieren sich auf die Informationsextraktion aus 2D-Geodaten wie Satellitenbildern (ALBUQUERQUE et al. 2016), jedoch werden 3D-Geodaten wie Airborne Laser Scanning (ALS)-Punktwolken bisher kaum als Eingabedaten für Crowdsourcing-Aufgaben eingesetzt. Im Vergleich zu Bildern, die relativ einfach für ungeschulte Nutzer zu interpretieren sind, ist die Datenerfassung aus 3D-Punktwolken erheblich anspruchsvoller, da das Arbeiten mit 3D-Daten auf 2D-Bildschirmen ein stärkeres räumliches Vorstellungsvermögen erfordert. Zudem können sich aufgrund der schwankenden Punktdichte Interpretationsschwierigkeiten ergeben (HACKEL et al. 2017).

In HERFORT et al. (2018) wird ein Crowdsourcingprojekt zur Informationsextraktion aus 3D-Punktwolken von einzelnen grob vorsegmentierten Bäumen präsentiert, welches eine direkte Interaktion mit den 3D-Punktwolken realisiert und Aufgaben wie binäre Klassifikation umfasst. Die Crowdaufgaben werden dabei zum einen freiwillig von Nutzern bearbeitet und erfordern zum anderen keine Interpretation von Objekten in einem räumlichen komplexen Kontext, da die

---

<sup>1</sup> Universität Stuttgart, E-Mail: st146807@stud.uni-stuttgart.de

<sup>2</sup> Universität Stuttgart, Institut für Photogrammetrie, Geschwister-Scholl-Str. 24D, D-70174 Stuttgart, E-Mail: [michael.koelle, volker.walter]@ifp.uni-stuttgart.de

verwendeten 3D-Punktwolken von Bäumen in einem ersten Vorverarbeitungsschritt segmentiert und somit von deren Umgebung bereits getrennt wurden. Nach unserem Wissensstand gibt es bisher noch keine Arbeit, welche die Erfassung von Objekten direkt in komplexen 3D-Szenen mittels bezahlter Crowdsourcing ermöglicht.

Nach WALTER et al. (2018) können mit bezahltem Crowdsourcing zwar hochqualitative Daten geliefert werden, jedoch sind die Ergebnisse hinsichtlich der Qualität typischerweise sehr inhomogen (WALTER et al. 2018). Die Arbeit stellt eine mögliche Lösung dieses Problems dar, indem durch Crowdworker mehrfach gesammelte Daten zu einem Gesamtergebnis integriert werden. Die integrierten Ergebnisse können eine höhere Qualität aufweisen als die einzelnen Ergebnisse.

Um große Mengen von annotierten Daten mit hoher Qualität für das Training von überwachten Ansätzen zur Klassifizierung von 3D-Punktwolken bereitstellen zu können, ist es von großem Interesse, das Potential einer bezahlten crowd-basierten Erfassung von komplexen Objekten wie Bäumen in 3D-Szenen zu nutzen und die Qualität der mittels bezahltem Crowdsourcing erhaltenen Daten im Sinne der „*Wisdom of the Crowd*“ durch Mehrfacherfassung zu steigern.

Im Rahmen dieser Arbeit werden drei Ziele verfolgt, (i) ein Web-Tool zur bezahlten crowd-basierten Erfassung von Bäumen aus 3D-Punktwolken mittels minimal umschließender Zylinder zu erstellen, (ii) die von bezahlten Crowdworkern mit dem entwickelten Web-Tool erfassten Daten mit den Referenzdaten zu vergleichen und hinsichtlich der Qualität zu analysieren und (iii) die durch einzelne Crowdworker mehrfach erfassten Bäume im Sinne einer Mehrheitsentscheidung zu integrieren und die am bestangepassten Zylinder zu rekonstruieren.

## 2 Methodik

### 2.1 Datengewinnung

Für die Datenerfassung wird zunächst ein Web-Tool entwickelt, welches es ermöglicht, Bäume aus 3D-Punktwolken mittels minimal umschließender Zylinder zu erfassen (siehe Abb. 1). Die Crowdworker können per Maus die 3D-Punktwolken drehen, verschieben und hineinzoomen, um einen dreidimensionalen Eindruck der Eingabedaten zu gewinnen. Zur Erfassung von Bäumen können die Crowdworker Zylinder per Maus positionieren, sowie den Radius und die Höhe der Zylinder skalieren, um einen jeden einzelnen Baum optimal zu approximieren.

Als Eingabedaten werden ALS Punktwolken verwendet. In Abb.2. werden die zwei verwendeten Datensätzen dargestellt. Die Grundlage des V3D-Datensatzes ist der *ISPRS Vaihingen 3D Semantic Labeling Benchmark* Datensatz (NIEMEYER et al. 2014). Der sogenannte M3D-Datensatz wurde im Rahmen der turnusmäßigen Landesbefliegung im März 2016 erfasst und umfasst die Umgebung des Mercedes-Benz-Museums in Stuttgart. Fünf Ausschnitte aus dem V3D-Datensatz und 14 Ausschnitte aus dem M3D-Datensatz werden ausgewählt, die jeweils einen kreisförmigen Bereich mit einem Radius von 50 m darstellen. Der Erfassungsbereich beschränkt sich auf einen Kreis mit einem Radius von 30 m, um die Komplexität der Aufgabe zu reduzieren, trotzdem aber den räumlichen Kontext zu erhalten.

Das entwickelte Web-Tool wird dann über die Crowdsourcing-Plattform *microWorkers* ([www.microWorkers.com](http://www.microWorkers.com)) den Crowdworkern zugänglich gemacht. Den Crowdworkern wird jeweils ein Ausschnitt der zu annotierenden Punktwolke zur Verfügung gestellt, wobei die Daten

serverseitig vorgehalten werden. Die von Crowdworkern erfassten Daten, in diesem Fall die Positionen, Radien und Höhen der Zylinder, werden am Ende wieder an den Server gesendet und gespeichert. Im Sinne einer redundanten Erfassung, wird jeder Punktwolkenausschnitt von je 10 Crowdworkern erfasst.



Abb. 1: Implementiertes Web-Tool zur Erfassung von Bäumen aus 3D-Punktwolken



Abb. 2: Kolorierte ALS Punktwolke und davon ausgewählte kreisförmige Erfassungsbereiche mit einem Radius von 30 m (*links*: V3D-Datensatz, *rechts*: M3D-Datensatz)

## 2.2 Qualitätsanalyse

Die Qualitätsanalyse der Objektextraktion aus 3D-Punktwolken kann entweder auf Punktlevel oder Objektlevel basieren (RUTZINGER et al. 2009). Da die Erfassung von Bäumen aus 3D-Punktwolken mittels minimal umschließender Zylinder in erster Linie auf Objektlevel bezogen ist, konzentriert sich die Evaluation in dieser Arbeit daher auf objekt-basierte Qualitätsmaße.

Eine besondere Schwierigkeit der objekt-basierten Auswertung liegt darin, dass die Anzahl der Instanzen in den Referenzdaten und den zu evaluierenden Daten unterschiedlich sein kann. Anders formuliert, wird sich nicht immer eine 1:1-Beziehung zwischen Referenzdaten und den zu evaluierenden Daten ergeben. Wie in Abb. 3 zu sehen ist, können zum einen mehrere Referenzzylinder durch einen großen Zylinder von Crowdworkern erfasst werden ( $n:1$ -Beziehung) und zum anderen kann ein Referenzzylinder zwei oder mehr von Crowdworkern erfasste Zylinder beinhalten ( $1:n$ -Beziehung). Außerdem können  $n:m$ -Beziehungen ebenfalls vorkommen, wobei mehrere Instanzen aus beiden Datensätzen miteinander korrespondieren.

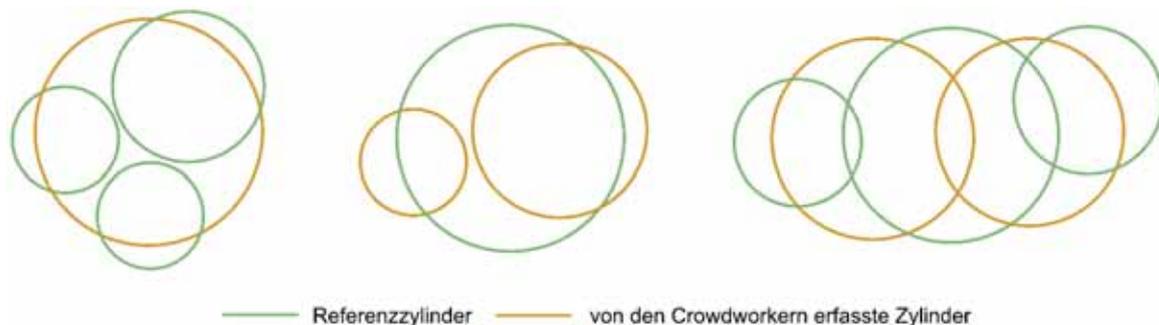


Abb. 3: Unterschiedliche Beziehungen zwischen den Referenzzylindern und den von Crowdworkern erfassten Zylindern (Draufsicht auf die Zylinder, *links*:  $n:1$ -Beziehung, *Mitte*:  $1:n$ -Beziehung, *rechts*:  $n:m$ -Beziehung)

Hierbei wird eine objekt-basierte Auswertungsmethode mittels Evaluation der gegenseitigen Überlappung in (ROTTENSTEINER et al. 2005) angewendet. Zur Erweiterung dieser Methode im dreidimensionalen Fall für Zylinder wird die Berechnung der Überlappung auf Schnittvolumen von Zylindern adaptiert. Die Referenzzylinder werden mit dem implementierten Web-Tool manuell sorgfältig erfasst. Die von den Crowdworkern erstellten Zylinder werden als *True Positive* Erfassungen bezeichnet, wenn diese hinsichtlich des Volumens mindestens 50% von den korrespondierenden Referenzzylindern überdeckt werden.

Zur Evaluation der crowd-basierten Erfassungen von Bäumen werden drei Qualitätsmaße, die Vollständigkeit *Comp* (auch *Recall*, *Producer's Accuracy*), Korrektheit *Corr* (auch *Precision*, *User's Accuracy*) und Qualität  $Q$  (siehe Gl. 1) bestimmt. Die Vollständigkeit *Comp* zeigt den Anteil von Bäumen in den Referenzdaten, die ebenfalls als *Baum* von Crowdworkern erfasst werden. Die Korrektheit *Corr* stellt dar, wie viel Prozent der von Crowdworkern als *Baum* annotierten Instanzen in den Referenzdaten tatsächlich zur Klasse *Baum* gehören. Die Qualität  $Q$  liefert ein Maß zur Bewertung der Gesamtleistung der Erfassung, wobei sowohl die Vollständigkeit als auch die Korrektheit berücksichtigt werden.

$$Q = \frac{Comp \cdot Corr}{Comp + Corr - Comp \cdot Corr} \quad (1)$$

## 2.3 Datenintegration

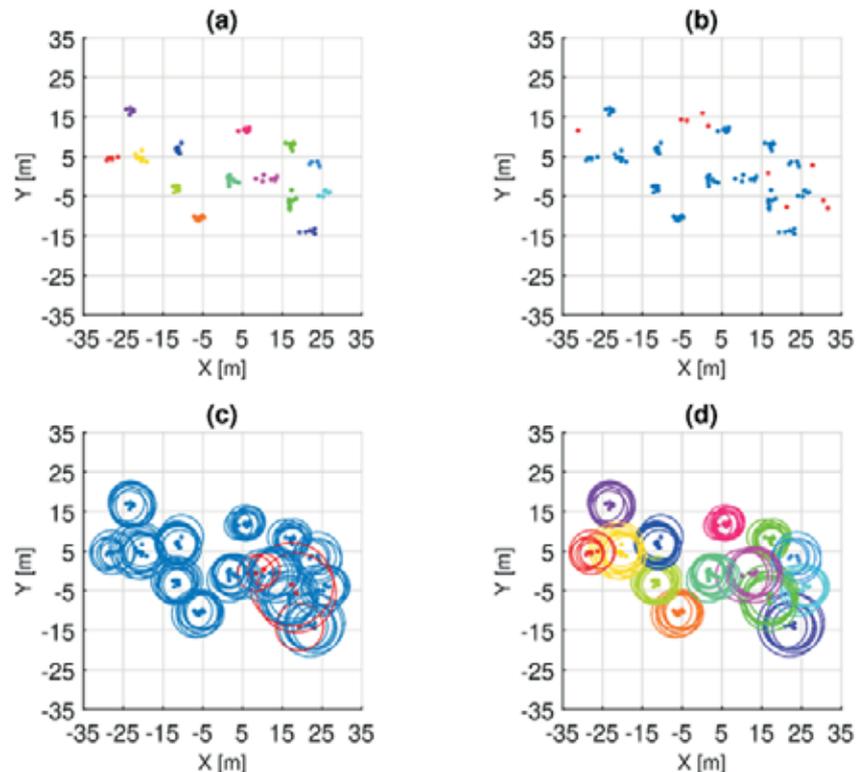


Abb. 4: Ablauf der Datenintegration: (a) Einteilung der Zylinder anhand der Positionen in Gruppen mittels DBSCAN (b) Filterung von Rauschpunkten mittels DBSCAN (*blau*: korrekte Erfassungen, *rot*: Ausreißer) (c) Ausreißerdetektion anhand gegenseitiger Überlappung (*blau*: korrekte Erfassungen, *rot*: Ausreißer) (d) Endergebnisse (farbkodiert je Baumcluster)

Abb. 4 stellt den Ablauf der Datenintegration dar. Die durch einzelne Crowdworker mehrfach erfassten Zylinder werden integriert, indem in erster Linie nur die Positionen der Zylinder im Rahmen eines Clusterings betrachtet werden und nach dem Verfahren DBSCAN (ESTER et al. 1996) ohne Verwendung der Referenzdaten automatisch in Gruppen eingeordnet werden (siehe Abb. 4a). Jede Gruppe beinhaltet die Zylinder, die von Crowdworkern zur Annotation eines einzelnen Baums mehrfach erfasst werden. Außerdem können falsch erfasste Zylinder wegen der stark unterschiedlichen Positionen mit DBSCAN als Ausreißer detektiert und eliminiert werden (siehe Abb. 4b).

Für jede Gruppe wird ein gemittelter Zylinder durch Mittelwertbildung der Positionen, Radien und Höhen der zugehörigen Zylinder rekonstruiert. Zur Eliminierung der Ausreißer mit stark unterschiedlichen Zylindergrößen wird das Schnittvolumen  $V_s$  zwischen dem gemittelten Zylinder  $l_m$  mit einem Volumen  $V_m$  und einem jeden dem Cluster zugehörigen Zylinder  $l_n$  mit einem Volumen  $V_n$  berechnet. Basierend auf den Schnittvolumen werden zwei Quoten  $q_m = V_s/V_m$  bzw.  $q_n = V_s/V_n$  der gegenseitigen Überlappung abgeleitet. Ist  $q_m$  und  $q_n$  kleiner als 50%, wird der entsprechende Zylinder  $l_n$  wegen geringer gegenseitiger Überlappung mit dem gemittelten Zylinder  $l_m$  als Ausreißer angesehen und eliminiert (siehe Abb. 4c). Nach dem Clustering und der Ausreißerdetektion werden die gemittelten Zylinder für jedes Baumcluster rekonstruiert (siehe Abb. 4d).

### 3 Ergebnisse & Diskussion

#### 3.1 Evaluation der Erfassung durch einzelne Crowdworker

Tab. 1: Durchschnittliche Vollständigkeit (*Comp*), Korrektheit (*Corr*) und Qualität (*Q*) der Erfassung von Bäumen aus beiden Datensätzen durch einzelne Crowdworker

Datensatz	<i>Comp</i> (%)	<i>Corr</i> (%)	<i>Q</i> (%)
V3D	85.19	92.64	79.37
M3D	92.42	96.82	89.58

Aus Tab. 1 geht hervor, dass die durchschnittliche Korrektheit *Corr* für beide Datensätze eine Genauigkeit von über 90% erreicht. Dagegen liegt die durchschnittliche Vollständigkeit *Comp* beim V3D-Datensatz bei einem Wert von 85%. Die bessere Performance der Korrektheit gegenüber der Vollständigkeit weist darauf hin, dass die von Crowdworkern als *Baum* annotierten Objekte verglichen mit den Referenzdaten in den meisten Fällen tatsächlich Bäume sind, jedoch nicht alle Bäume bei der Erfassung identifiziert werden.

Die durchschnittliche Qualität *Q* der Erfassungen beim M3D-Datensatz ergibt sich zu 89.58%, welche um 10 Prozentpunkte höher als die entsprechende Qualität des V3D-Datensatzes ausfällt. Durch den Vergleich der Qualitätsmaße von beiden Datensätzen kann festgestellt werden, dass der M3D-Datensatz durch die Crowd genauer erfasst werden kann als der V3D-Datensatz. Ein Grund dafür liegt darin, dass im V3D-Datensatz komplexere Szenen enthalten sind, wobei bei der Erfassung der Bäume eine Abgrenzung zu anderen Vegetationsklassen, wie vor allem Büsche, erforderlich ist. Demgegenüber liegen beim M3D-Datensatz hauptsächlich eindeutig als Baum identifizierbare Instanzen vor. Darüber hinaus kann der Unterschied der Punktdichte zwischen beiden Datensätzen ein weiterer Grund sein. Die Punktdichte des V3D-Datensatzes liegt zwischen 4 und 7 Punkten/m<sup>2</sup>, während die Punktdichte des M3D-Datensatzes zwischen 4 und 32 Punkten/m<sup>2</sup> variiert. Eine höhere Punktdichte kann Crowdworkern bei der Erfassung visuell helfen den räumlichen Kontext von Punktwolken besser zu interpretieren.

#### 3.2 Evaluation der integrierten Ergebnisse aus Mehrfacherfassung

Tab. 2: Durchschnittliche Vollständigkeit (*Comp*), Korrektheit (*Corr*) und Qualität (*Q*) der gemittelten Zylinder nach der Datenintegration aus beiden Datensätzen

Datensatz	<i>Comp</i> (%)	<i>Corr</i> (%)	<i>Q</i> (%)
V3D	97.28	100.00	97.28
M3D	99.68	100.00	99.68

Die durchschnittlichen Qualitätsmaße der mittels Datenintegration rekonstruierten gemittelten Zylinder aus beiden Datensätzen werden in Tab. 2 aufgelistet. Auffällig ist, dass alle Qualitätsmaße nach der Datenintegration eine sehr hohe Genauigkeit von mindestens 97% erreichen. Vor allem erzielen die integrierten Zylinder von beiden Datensätzen eine Korrektheit von 100%. Das bedeutet, dass alle mittels Datenintegration rekonstruierten Zylinder tatsächlich zur Klasse *Baum* gehören. Für den M3D-Datensatz werden die integrierten Zylinder von allen Punktwolkenausschnitten zusammengefasst und in Abb. 5 dargestellt.



Abb. 5: Darstellung der integrierten Ergebnisse des gesamten M3D-Datensatzes (*links*: kolorierte Punktwolke als Eingabedaten, *rechts*: durch Datenintegration rekonstruierte gemittelte Zylinder)

Abb. 6 verdeutlicht die erhebliche Qualitätssteigerung der erfassten Daten durch die vorgestellte Datenintegration. Durch diese lässt sich die durchschnittliche Qualität der integrierten Erfassungen beim V3D-Datensatz um mehr als 15 Prozentpunkte steigern. Auch die Genauigkeit der einzelnen Erfassungen beim M3D-Datensatz kann durch die Datenintegration um ca. 10 Prozentpunkte gesteigert werden. Dieser Anstieg verifiziert das Dogma der „*Wisdom of the Crowd*“. Damit ist gemeint, dass die Integration der Entscheidungen mehrerer Personen oft bessere Ergebnisse liefert als diejenigen einzelner Individuen. Solange ein Teil der Crowdworker einen Baum in den Punktwolken erfolgreich erkennt, werden die Erfassungen von diesem Baum bei der Datenintegration berücksichtigt und schließlich wird ein entsprechend gemittelter Zylinder rekonstruiert. Infolgedessen können die von einigen Crowdworkern übersehenen Bäume dennoch mittels Datenintegration redetektiert werden. Zusätzlich erlaubt diese die Rekonstruktion von am best-angepassten Zylindern.

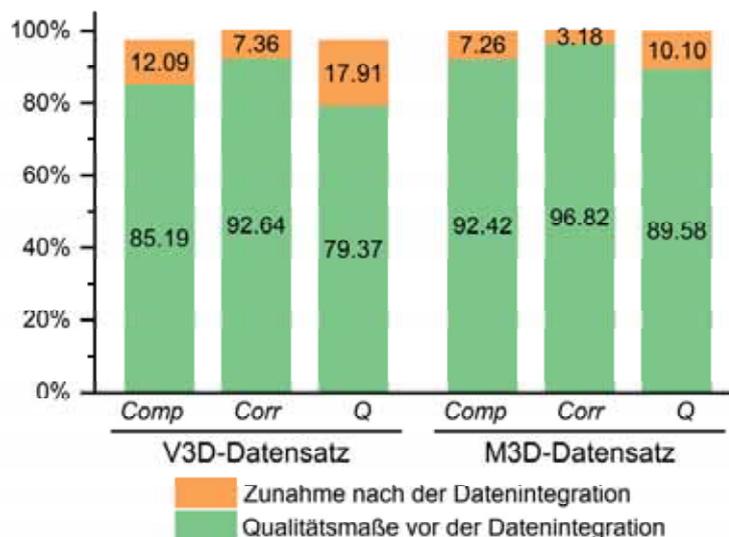


Abb. 6: Steigerung der durchschnittlichen Qualitätsmaße mittels Datenintegration

### 3.3 Evaluation der einzelnen Zylinderparameter

Für die gemittelten Zylinder werden die Differenzen der einzelnen Parameter (Position, Radius und Höhe) unter Verwendung der Referenzdaten gebildet. Abb. 7 stellt die relative Frequenz der Differenzen zwischen den gemittelten Zylindern und Referenzzylindern dar. Nach der Datenintegration liegt die durchschnittliche Abweichung der Position bei 0,73 m. Die Positionen der gemittelten Zylinder weisen in den meisten Fällen geringe Abweichungen von unter 1 m zu den Referenzzylindern auf. Die durchschnittliche Abweichung des Radius beträgt lediglich 0,35 m und der Höhe 0,34 m. Darüber hinaus liegt die maximale relative Frequenz der Differenzen bezüglich des Radius und der Höhe nahe Null. Dies zeigt, dass die Radien und Höhen der gemittelten Zylinder weitgehend an die Referenzzylinder angepasst werden können.

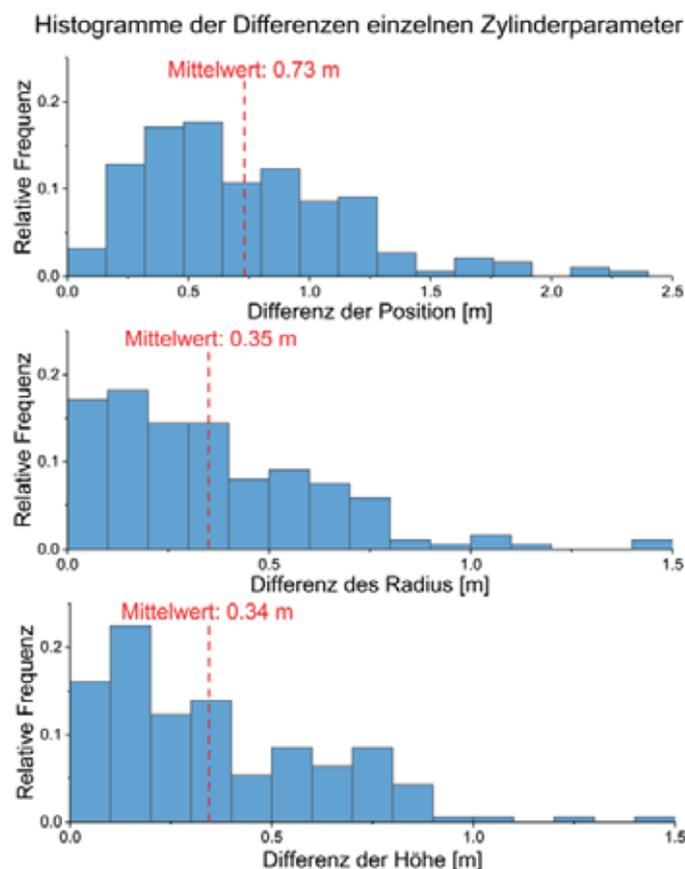


Abb. 7: Verteilungen der Differenzen der einzelnen Parameter zwischen den Referenzzylindern und den gemittelten Zylindern

Ein Teil der Erfassungen vor der Datenintegration beinhaltet fälschlicherweise annotierte Objekte, die tatsächlich keine Bäume repräsentieren. Darüber hinaus werden mehrere räumlich nahliegende Bäume von einigen Crowdworkern in einem großen Zylinder zusammengefasst. Daher ist es schwer, für die von den Crowdworkern erfassten Zylinder eine 1:1-Zuordnung zu den korrespondierenden Referenzzylindern festzulegen und die Abweichungen der einzelnen Parameter zu bestimmen. Zur visuellen Evaluation der Zylinderpositionen vor und nach der

Datenintegration, werden in Abb. 8 die Positionen aller erfassten Zylinder vor der Datenintegration, den gemittelten Zylindern nach der Datenintegration sowie den Referenzzylindern für einen repräsentativen Punktwolkenausschnitt gegenübergestellt. Es zeigt sich, dass mittels Datenintegration Erfassungen mit stark abweichenden Positionen als Ausreißer detektiert und eliminiert werden können. Außerdem nähern sich die gemittelten Zylinder in den meisten Fällen den Referenzzylindern an und weisen eine höhere Übereinstimmung mit den Referenzdaten als die der einzelnen Erfassungen auf.

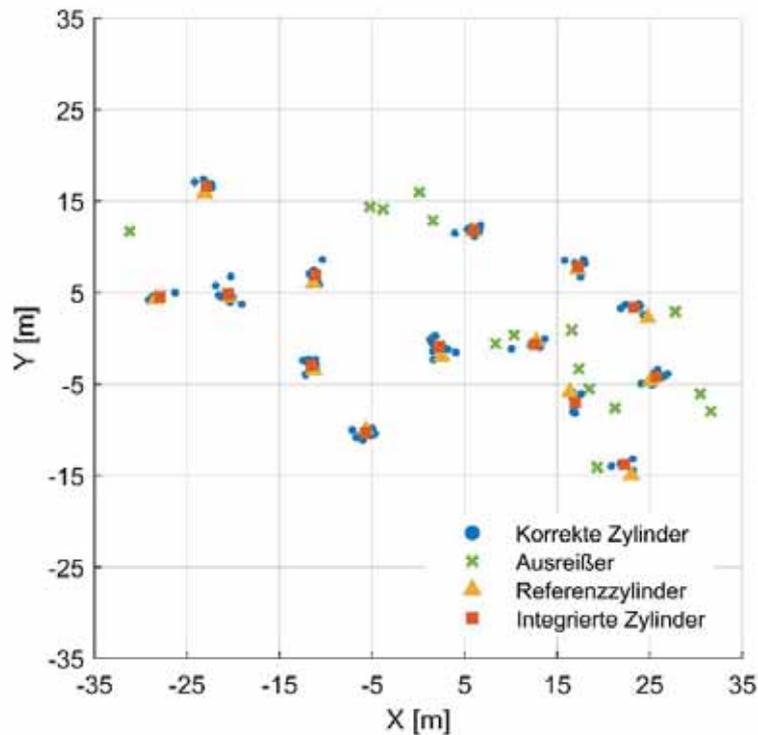


Abb. 8: Positionen der Zylinder vor und nach der Datenintegration für einen Punktwolkenausschnitt

## 4 Fazit & Ausblick

In dieser Arbeit wurde ein Web-Tool zum Zweck der bezahlten crowd-basierten Erfassung von Bäumen aus 3D-Punktwolken mittels minimal umschließender Zylinder implementiert. Die Qualität der mittels des erstellten Web-Tools erfassten Daten hängt von der Komplexität der 3D-Szene ab. Für leicht interpretierbare Datensätze, die hauptsächlich klar strukturierte Bäume mit einem eindeutig sichtbaren Stamm und einer ausgeprägten Krone enthalten, können die von den Crowdworkern erfassten Daten eine Qualität von über 90% erreichen. Hingegen erzielen die Erfassungen bei komplexeren Datensätzen nur eine Qualität von unter 80%. Aufgrund der inhomogenen Qualität der durch bezahltes Crowdsourcing erhaltenen Daten werden die von Crowdworkern mehrfach erfassten Zylinder im Sinne einer Mehrheitsentscheidung integriert. Durch Datenintegration kann die Übereinstimmung mit den Referenzdaten gesteigert werden, sodass eine hohe Qualität von über 95% erreicht wird. Außerdem werden die Positionen und die Ausmaße der durch Datenintegration rekonstruierten Zylinder weitgehend an die Referenzzylinder

angepasst. Die Ergebnisse zeigen, dass Crowdsourcing ein erhebliches Potential hinsichtlich der Annotation von 3D-Punktwolken aufweist und die integrierten Ergebnisse aus Mehrfacherfassung hohe Qualität erreichen können.

Allerdings können die zur Erfassung von Bäumen erstellten Zylinder auch andere Klassen wie *Boden* und *Niedrige Vegetation* beinhalten. Es ist sinnvoll weiter zu untersuchen, wie die mittels Zylinder erfassten Daten in einem weiteren Schritt nachbearbeitet werden können, um irrelevante Objekte zu filtern.

## 5 Danksagung

Die Autoren danken dem Landesamt für Geoinformation und Landentwicklung (LGL) Baden-Württemberg für die Bereitstellung der ALS-Punktwolken der Stadt Stuttgart.

## 6 Literaturverzeichnis

- ALBUQUERQUE, J., HERFORT, B. & ECKLE, M., 2016: The tasks of the crowd: A typology of tasks in geographic information crowdsourcing and a case study in humanitarian mapping. *Remote Sensing* **8**(10), 859.
- ESTER, M., KRIEGEL, H.P., SANDER, J. & XU, X., 1996: A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of Second International Conference on Knowledge Discovery and Data Mining* **96**(34), 226-231.
- HACKEL, T., SAVINOV, N., LADICKY, L., WEGNER, J. D., SCHINDLER, K. & POLLEFEYS, M., 2017: Semantic3d. net: A new large-scale point cloud classification benchmark. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* **IV-1/W1**, 91-98.
- HERFORT, B., HOEFLE, B., & KLONNER, C., 2018: 3D micro-mapping: Towards assessing the quality of crowdsourcing to support 3D point cloud analysis. *ISPRS Journal of Photogrammetry and Remote Sensing* **137**, 73-83.
- HOSSFELD, T., HIRTH, M. & TRAN-GIA, P., 2012: Aktuelles Schlagwort: Crowdsourcing. *Informatik Spektrum* **35**(3), 204-208.
- NIEMEYER, J., ROTTENSTEINER, F. & SOERGEL, U., 2014: Contextual classification of lidar data and building object detection in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, **87**, 152-165.
- ROTTENSTEINER, F., TRINDER, F., CLODE, S. & KUBIK, K., 2005: Using the Dempster-Shafer method for the fusion of LIDAR data and multi-spectral images for building detection. *Information fusion* **6**(4), 283-300.
- RUTZINGER, M., ROTTENSTEINER, F. & PFEIFER, N., 2009: A comparison of evaluation techniques for building extraction from airborne laser scanning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **2**(1), 11-20.
- SU, H., DENG, J. & FEI-FEI, L., 2012: Crowdsourcing annotations for visual object detection. *AAAI Workshops*.
- WALTER, V. & SOERGEL, U., 2018: Implementation, Results, and Problems of Paid Crowd-Based Geospatial Data Collection. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science* **86**(3-4), 187-197.

# Dynamic Aggregation of Geo-Objects for the Interactive Exploration of Research Data

ANNIKA BONERATH<sup>1</sup>, BENJAMIN NIEDERMANN<sup>1</sup> & JAN-HENRIK HAUNERT<sup>1</sup>

*Abstract: In the context of research data management, the interactive exploration of data is a useful tool to make data findable and reusable. For exploring spatio- and temporal data we developed in a previously published work a data structure that provides the user with simple visualizations of the data for time window queries. We considered the data to be points in space each associated with a time stamp and we visualized it using  $\alpha$ -shapes, which generalize convex hulls. In general, time windowed data structures support time window queries for geometric shapes or more general problems from computational geometry such as counting and intersection problems. With this paper, we contribute a review of our previously published method in the context of time windowed data structures, which is a relatively new concept of computational geometry. In particular, we highlight the relevance of our work for a growing domain of research.*

## 1 Introduction

In times of big and heterogeneous research data, the data management and the exploration of available data becomes more and more important. Especially in the scientific context, projects often deal with these large amounts of spatio- and temporal data that need to be findable and reusable. For example, consider a project that deals with a database of meteorological data; see Figure 1. Each object in the data set is a storm event that is represented as a point in space and time. For scientific tasks often the researcher is not interested in all the data but only in a subset that is limited to a time window. In order to find the right time window, the user might want to interactively explore the data, by retrieving simplified visualizations of the data in time windows; see Figure 2.

In this work, we visualize the queried data by sketching its outline. This outline provides the user with the possibility of roughly assessing the spatial distribution of the data. Since for most data sets, simple representations as the convex hull are not adequate, a wide range of more sophisticated polygonal representations exists; some of these are based on Delaunay-triangulations (DE BERG et al. 2011; DUCKHAM 2008; EDELSBRUNNER et al. 1983) while others use spatial grids to define the representation (ATTALI 1997; JONES et al. 2008; PURVES et al. 2005).

In this paper, we use  $\alpha$ -shapes (EDELSBRUNNER et al. 1983; EDELSBRUNNER 2010) for representing point sets, which are a generalization of convex hulls and strongly related to Delaunay triangulations. Among others, this technique finds its application in pattern recognition (VAUHKONEN et al. 2010) and micro-biology (LIANG et al. 1998).

---

<sup>1</sup> Rheinische Friedrich-Wilhelms-Universität Bonn, Institut für Geodäsie und Geoinformation, Meckenheimer Allee 172, D-53115 Bonn, E-Mail: [bonerath, niedermann, haunert]@igg.uni-bonn.de

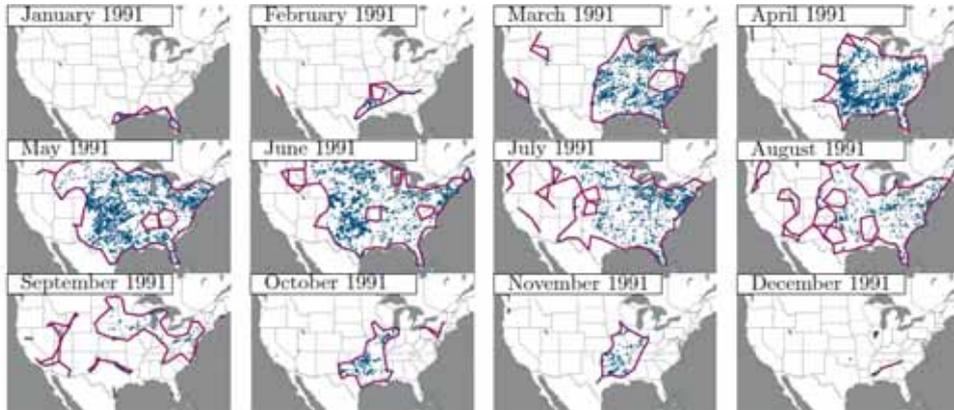


Fig. 1: Storm events for the months of 1991 represented by  $\alpha$ -shapes (lilac). The actual point set (blue) is drawn for illustration. Data retrieved from Data.gov. Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL.

In the following, we give the definition of  $\alpha$ -shapes. Let  $P \subseteq \mathbb{R}^2$  be a set of  $n$  points in the plane and let  $\alpha > 0$ . Further, let  $pq \in P \times P$  be a directed edge with  $|q - p| \leq \alpha$ . We define the *edge domain* of  $pq$  as the open disk  $D_{pq}$  with radius  $\alpha/2$  whose center lies to the right of  $pq$  and whose boundary contains the points  $p$  and  $q$ . Let  $S_\alpha(P) \subseteq P \times P$  be the set of all edges that are shorter than  $\alpha$  and do not contain any point of  $P$  in their edge domain. We call  $S_\alpha(P)$  the  $\alpha$ -shape of  $P$ ; see Figure 3(a). Based on the Delaunay triangulation, the  $\alpha$ -shape of  $n$  points can be computed in  $O(n \log n)$  time (EDELBRUNNER et al. 1983).

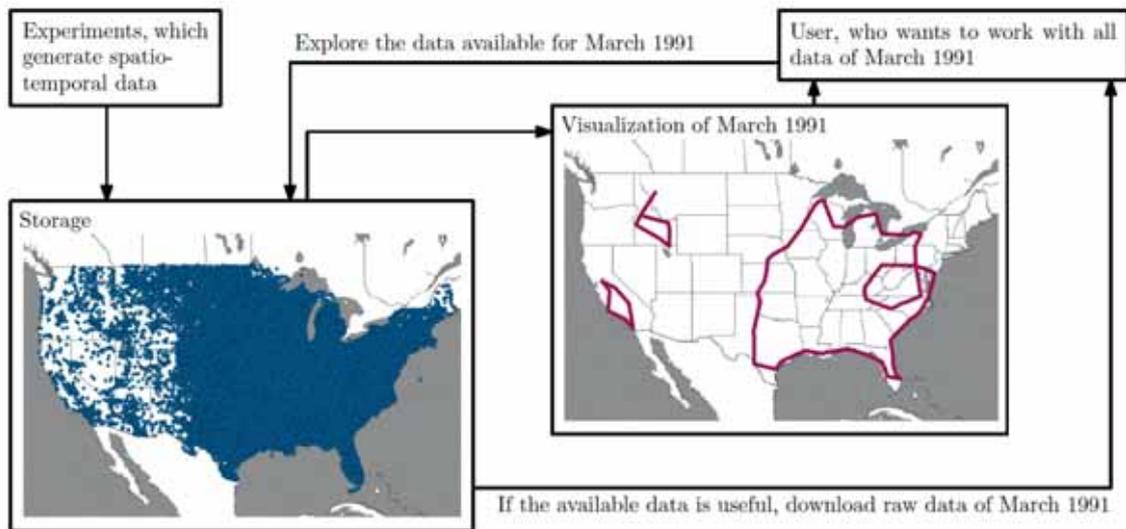


Fig. 2: Scenario for the case that the user queries a simplified visualization for all storm events in March 1991.

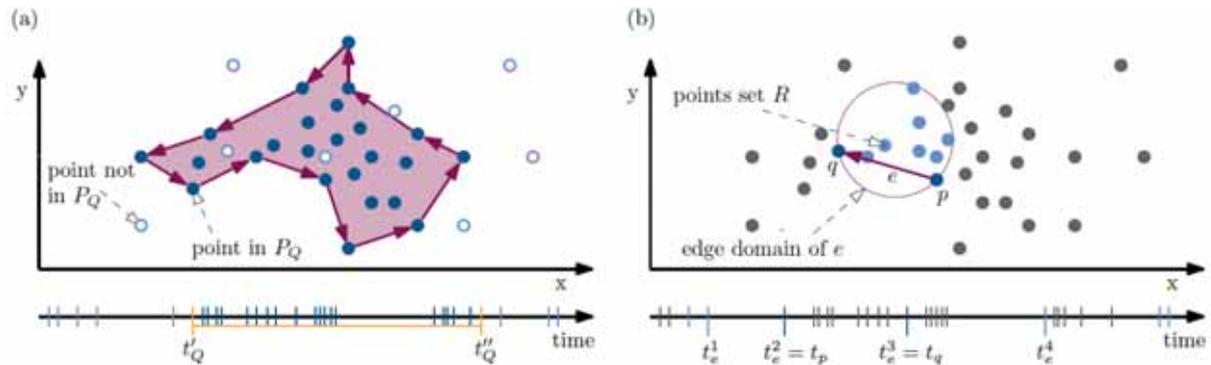


Fig. 3: (a)  $\alpha$ -shape (lilac arcs) of a queried point set  $P_Q$  (filled blue discs) and (b) an edge with its edge domain and time attributes.

In order to apply the definition of  $\alpha$ -shapes to our problem setting, we first introduce two more concepts. For a point  $p$  we denote its *time stamp* by  $t_p \in \mathbb{R}$ . Further, a *time window query*  $Q$  is a query for a temporal range  $[t'_Q, t''_Q]$  and  $P_Q = \{p \in P \mid t_p \in Q\}$  is the subset of  $P$  that is contained in  $Q$ . As described in the running example, the point set  $P$  is queried frequently for the  $\alpha$ -shape  $S_\alpha(P_Q)$  of some time window query  $Q$ . A straight-forward approach for a query  $Q$  first queries the set  $P$  obtaining  $P_Q$  and then computes the  $\alpha$ -shape  $S_\alpha(P_Q)$ . Utilizing a balanced binary search-tree for finding  $P_Q$  and additionally computing the  $\alpha$ -shape, we obtain  $O(\log n + |P_Q| \log |P_Q|)$  running time. For our use-case, we aim at a better running time per query. In particular, we propose to use a data structure for filtering search (CHAZELLE 1986) that allows us to answer a query in  $O(\log n + k)$  where  $k$  is the size of the returned  $\alpha$ -shape. In a pre-processing phase, we compute a data structure that aggregates the  $\alpha$ -shapes of all possible queries; we call it the  $\alpha$ -structure of  $P$ . We use this data structure in the query phase to obtain the  $\alpha$ -shapes of the incoming queries. Note that data structures, like the  $\alpha$ -structure that preprocess temporal data so that specific queries for a time window can be handled efficiently are called *time windowed data structures*. In Section 2, we provide an overview about existing time windowed data structures.

We have published two works concerning the  $\alpha$ -structure, one as a preprint at the workshop EuroCG'19 and an extended conference version at ACM SIGSPATIAL'19 (BONERATH et al. 2019a, BONERATH et al. 2019b). For the conference version we extended our approach to also provide schematized  $\alpha$ -shapes. In particular, we can provide octilinear  $\alpha$ -shapes where each edge has a direction that stems from a predefined set of eight directions, namely horizontal, vertical, and diagonal directions. Further, we provided a method, where the  $\alpha$  parameter of the queried  $\alpha$ -shape does not need to be fixed before preprocessing the  $\alpha$ -structure.

## 2 Related Work

In general, a time windowed data structure answers queries concerning specific properties of the input data set for time windows. This concept was first considered by BANNISTER et al. (2013). They considered social network data as their input. On this data a *relational event graph* is built, where each vertex represents an entity of the social network and each edge represents a

communication between two entities. Each edge is associated with a timestamp. In their work they provide data structures that handle counting problems for both general graph properties such as the number of connected components, and properties for social network analysis. Further improvements, extensions and results on time windowed data structures for relational event graphs are presented by CHANCHARY & MAHESHWARI (2019) and CHANCHARY et al. (2017).

Time windowed data structures are also considered for sets of geometric objects where each geometric object is associated with a point in time. For example, closely related to our work, BANNISTER et al. (2014) discussed the problem of reporting the convex hull, the skyline, and other basic problems from computational geometry for time window queries. Especially, closely related to our work is the problem of reporting the convex hulls for points with time stamps since the  $\alpha$ -shape is a generalization of the convex hull. As a basic concept for their data structure, they build a decomposition tree in which each node contains the convex hull of its descendants. Their data structure reports the convex hull in  $O(h \log^2 w)$  time where  $h$  is the number of edges of the convex hull and  $w$  is the number of points contained in the queried time window. In contrast, the running time of our method depends not directly on the number of points in the queried time window but only on the size of the  $\alpha$ -shape for those points

RUDI (2018) discusses the design of a time windowed data structure for answering hotspot queries on trajectories. The hotspot of a trajectory is the square of fixed side length that contains the longest contiguous part of the trajectory. As input data each trajectory point is associated with a time stamp. Answering a time window query takes  $O(\log^2 n)$  time. The algorithm provides only an approximation. The idea of the data structure is to enrich each vertex with additional attributes with which the result can be computed. This idea is similar to our approach, since we also enrich the edges of the  $\alpha$ -structure with additional attributes that we use in the query.

Further work on time windowed data structures for sets of geometric objects with time stamps concerns decision problems such as the problem if the convex hull area is larger than some threshold (BOKAL et al. 2015, CHAN & PRATT 2016), intersection decision problems and counting problems of extreme points (CHANCHARY et al. 2018), and geometric properties like reporting the closest pair (CHAN & PRATT 2015).

### 3 On $\alpha$ -Structures

In the following, we define the  $\alpha$ -structure of  $P$ . We say that an edge  $pq \in P$  is *active* for a temporal query  $Q$  if the  $\alpha$ -shape  $S_\alpha(P_Q)$  contains  $pq$ . We observe that an edge  $pq$  can be active for an infinite set of temporal queries, but it can only be active for  $O(n^2)$  different subsets of  $P$ . To characterize this set, we introduce the following notation; see Figure 3(b).

Let  $e = pq \in P \times P$  with  $t_p < t_q$  and let  $R \subseteq P$  be the set of points contained in the edge domain of  $pq$ . Further, let  $t_r$  with  $r \in R$  be the largest time stamp that is smaller than  $t_p$ ; if  $r$  does not exist, we set  $t_r = -\infty$ . Similarly, let  $t_s$  with  $s \in R$  be the smallest time stamp that is greater than  $t_q$ ; if  $s$  does not exist, we set  $t_s = \infty$ . We call  $t_e^1 = t_r$ ,  $t_e^2 = t_p$ ,  $t_e^3 = t_q$ ,  $t_e^4 = t_s$  the *time attributes* of  $pq$ . Using the time attributes we can formulate the *conditions of an active edge* for query  $Q$

- (1) the distance between  $p$  and  $q$  is smaller than  $\alpha$ ,
- (2)  $\forall r \in R : t_r \notin [t_p, t_q]$ , and
- (3)  $t'_Q \in [t_e^1, t_e^2]$  and  $t''_Q \in [t_e^3, t_e^4]$ .

We show that the conditions of an active edge are necessary and sufficient for an edge  $e$  to be active for  $Q$ . Assume that  $e$  is active for  $Q$ . This is equivalent to the following three conditions; (i)  $p, q \in P_Q$ , which is equivalent to  $t_p, t_q \in Q$ , (ii)  $e$  is shorter than  $\alpha$  (equivalent to Condition (1)) and (iii) no point  $r \in P_Q$  is contained in  $R$ , which is equivalent to  $\forall r \in R : t_r \notin Q$ . Applying the definition of the time attributes  $t_e^1, t_e^2, t_e^3, t_e^4$ , Condition (i) and (iii) are equivalent to Condition (2) and (3).

The  $\alpha$ -structure  $S_\alpha(P) \subseteq P \times P$  of  $P$  is the set of all active edges over all possible temporal queries. We show that Condition (1) and (2) are necessary and sufficient for an edge  $pq$  to be contained in  $S_\alpha(P)$ .

**Lemma 2.1.** *The edge  $e = pq \in P \times P$  is contained in  $S_\alpha(P)$  if and only if*

- (1) the distance between  $p$  and  $q$  is smaller than  $\alpha$ , and
- (2)  $\forall r \in R : t_r \notin [t_p, t_q]$ .

**Proof of Lemma 2.1.** Let  $e \in S_\alpha(P)$ , and let  $Q = [t'_Q, t''_Q]$  be a query for which  $e \in S_\alpha(P_Q)$ . Then  $e$  fulfills the conditions of an active edge for query  $Q$  and therefore the conditions of Lemma 2.1. Conversely, let  $e$  be shorter than  $\alpha$  and all points  $r \in R$  be temporally not in  $[t_p, t_q]$ . Then the  $\alpha$ -shape of the query  $Q$  with  $t'_Q = t_p$  and  $t''_Q = t_q$  contains the edge  $e$ .  $\square$

For our use-case of a database, the memory consumption of our approach is decisive for being deployed in practice. We first observe that  $O(n^2)$  is an upper bound for the size of an  $\alpha$ -structure. The following theorem shows that this is also a lower bound in the worst case.

**Theorem 2.2.** *For a set  $P$  of  $n$  points the  $\alpha$ -structure has size  $\Omega(n^2)$  in the worst case.*

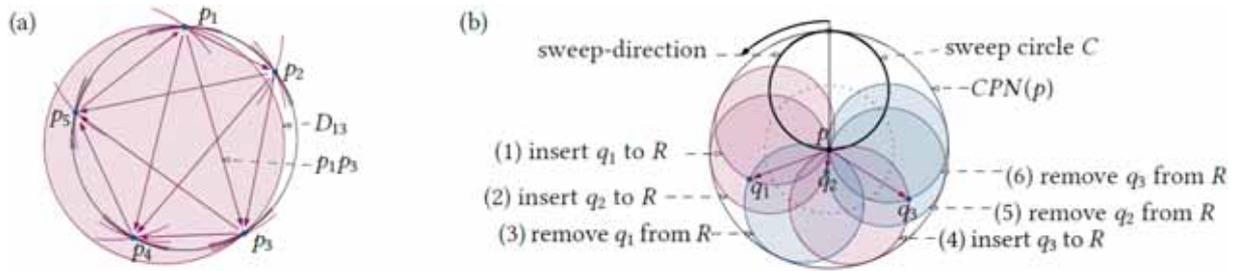


Fig. 4: (a) worst-case example for the size of the  $\alpha$ -structure as described in Theorem 2.2 and (b) the rotational sweep CPN-check for a point  $p$  and  $CPN T_q = \{q_1, q_2, q_3\}$ .

**Proof of Theorem 2.2.** Let  $P = \{p_1, p_2, \dots, p_n\}$  be a point set with time stamps  $t_1 < t_2 < \dots < t_n$  such that the points lie on a circle  $C$  of radius  $r < \alpha/2$  ordered clockwise according to their time stamps; see Figure 4(a). Let  $p_i, p_j \in P$  be two points with  $i < j$ . We show that  $p_i p_j$  is contained in the  $\alpha$ -structure  $S_\alpha(P)$  by proving the two conditions of Lemma 2.1. Due to  $r < \alpha/2$  the points  $p_i p_j$  have distance smaller than  $\alpha$ . Hence, Condition (1) of Lemma 2.1 is satisfied. For the second condition let  $R_{ij}$  be the set of points contained in edge domain  $D_{ij}$  of  $p_i p_j$ .

We observe that  $D_{ij}$  and  $C$  intersect in  $p_i$  and  $p_j$ . Since the radius of  $C$  is smaller than the radius of  $D_{ij}$  the boundary of  $D_{ij}$  partitions  $C$  into two parts. One part is contained in  $D_{ij}$  and the other lies outside of  $D_{ij}$ . Since the points  $p_1, p_2, \dots, p_n$  appear in clockwise order on  $C$ , and since the center  $D_{ij}$  lies to the right of  $p_i p_j$  by definition, we obtain  $R_{ij} = \{p_1, \dots, p_{i-1}, p_{j+1}, \dots, p_n\}$ . Consequently, Condition (2) is satisfied.

Hence, the database may exceed a size that is applicable in practice. However, the example is rather unlikely to occur in practice. Generally, the size of the database is bounded by  $O(nm)$  where  $m$  is the largest number of points in a distance  $\alpha$  to a point in  $P$ . Assuming that the point density is bounded by a constant and  $\alpha$  is fixed,  $m$  is also constant. If, on the other hand, the density increases, it becomes more likely that an edge gets destroyed. In our experiments, we observe a linear relation between the number of points and  $\alpha$ -structure size; see Section 4.

## 4 Constructing and Querying $\alpha$ -Structures

We introduce an algorithm that constructs the  $\alpha$ -structure of a point set  $P$  in  $O(n(\log n + \bar{m} \log \bar{m}))$  time, where  $\bar{m}$  is the maximum number of points in a square with width  $2\alpha$ . Further, we describe how to query this data structure.

The construction algorithm consists of two parts; see Algorithm 1. Each part is applied for each point  $p \in P$ . We call the first part *CPN-Search*. In this part we compute the set of all points  $T_p \subseteq P$  that fulfill Condition (1) of Lemma 2.1, i.e., all points that lie in a circle with center at  $p$  and radius  $\alpha$ . We call this circle the *circle of potential neighbors (CPN)* of  $p$ . We implement the CPN-Search with a sweep line approach to find  $T_p$  in  $O(\log n + \bar{m})$  time (PENG 2014).

**Algorithm 1.** Computation of the  $\alpha$ -structure

**Input:** point set  $P$ , parameter  $\alpha$   
**Output:**  $\alpha$ -structure  $S_\alpha(P)$   
**foreach**  $p \in P$  **do**  
     *CPN-Search:* Find all points  $T_p \subseteq P$  in the CPN of  $p$   
     *CPN-Check:* Check for each edge  $pq$  with  $q \in T_p$  whether it fulfils Condition (2) of Lemma 2.1, possibly compute the time attributes and add to  $S_\alpha(P)$

The second part of the algorithm, which we call *CPN-Check*, checks for each point  $q \in T_p$  whether the edge  $pq$  fulfills Condition (2) of Lemma 2.1. If this is the case, it computes the time attributes of  $pq$ . We use a rotational sweep for the CPN-Check. More precisely, we use a circle  $C$  of radius  $\alpha/2$  which sweeps around  $p$  such that the center of  $C$  moves along the circle with center  $p$  and radius  $\alpha/2$ ; see Figure 4(b). We call  $C$  the *sweep circle* of  $p$ . Let  $\hat{R}$  be the points contained in  $C$ ; we represent  $\hat{R}$  using a binary search tree ordered by the time stamps of the points. The sweep circle  $C$  stops its rotation whenever its boundary intersects with a point  $q \in T_p$ . Two kind of events are possible; either the point  $q$  enters  $C$ , or it leaves  $C$ . Whenever a point  $q$  enters  $C$ , the sweep circle equals the edge domain of  $pq$ . Utilizing the properties of the binary search tree  $\hat{R}$ , Condition (2) of Lemma 2.1 can be checked in  $O(\log \bar{m})$  time. If this is the case, the time attributes of  $pq$  can be computed using the temporal order of  $\hat{R}$  in  $O(\log \bar{m})$  time. This rotational sweep can be done in  $O(\bar{m} \log \bar{m})$  time.

**Theorem 2.3.** *Computing the  $\alpha$ -structure of a set of  $n$  points takes  $O(n(\log n + \bar{m} \log \bar{m}))$  time.*

For the query phase we represent each edge  $e$  with  $t_e^1, t_e^2, t_e^3, t_e^4$  of the  $\alpha$ -structure by a rectangle  $[t_e^1, t_e^2] \times [t_e^3, t_e^4]$ . A query  $[t'_Q, t''_Q]$  corresponds to finding all rectangles containing  $(t'_Q, t''_Q)$ . For these rectangles we propose to use a data structure for filtering search (CHAZELLE 1986) that allows us to answer a query in  $O(\log n + k)$ , where  $k$  is the size of the returned  $\alpha$ -shape.

## 5 Experimental Evaluation

We analyze the performance of  $\alpha$ -structures using on the one hand a data set of storm events in the United States in the years 1991-2000 obtained from Data.gov; see Figure 1 for the year 1991 and on the other hand a set of synthetic generated data with uniform distribution in time and space with density of 0.001 points per  $m^2$ . We performed the experiments on a 4-core Intel Core i7-7700T CPU with 16 GiB RAM using an implementation in Java.

The experiments indicate for both data sets that the memory consumption is linear in  $n$ ; see Figure 5(a). The construction time for a point set of size  $n = 70000$  varies between several seconds and hours depending on the value of  $\alpha$ . For smaller  $\alpha$  we obtain a shorter running time than for larger  $\alpha$ . We assume this to be acceptable, since it is a pre-processing step.

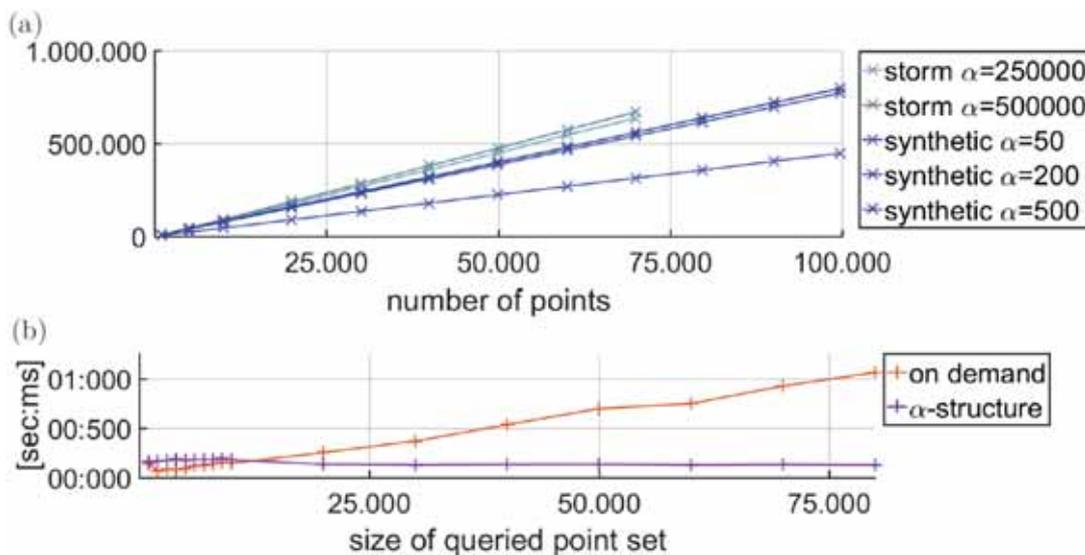


Fig. 5: (a) memory consumption of an  $\alpha$ -structure (parameter  $\alpha$  in meter) and (b) query phase time of the  $\alpha$ -structure compared to an on demand approach for the synthetic dataset and  $\alpha = 200$ .

Figure 5(b) illustrates the query time. For this experiment, we queried the data structure for different window sizes. The x-axis shows the size of the queried point set  $P_Q$ , the y-axis the query time. The experiments show that the query time using the  $\alpha$ -structure is nearly constant around 250[ms] with respect to the size of  $P_Q$ ; see Figure 5(b). In contrast, the results for an implemented on demand approach indicate a dependency to the subset size. For this data set the query times for the  $\alpha$ -structure are smaller than for the on demand approach for all subsets of size 20.000 or larger.

## 6 Conclusion and Outlook

Overall, we presented the design and construction of a data structure that provides the edges of  $\alpha$ -shapes for time window queries on point sets. Using this data structure the query time depends not on the number of points in the time window, but on the size of the  $\alpha$ -shape. The experiments indicate that with respect to preprocessing time and memory consumption this data structure is applicable in practice. Further, the experiments show that the  $\alpha$ -structure performs better than the on demand approach with respect to the query time.

For future work we plan to extend our approach of  $\alpha$ -structures such that we can handle also other geometric objects than points, e.g. lines and polygons. Further, we want to consider other aggregated representations of geographic objects for time windowed data structures. Even more general, we are interested in the concept of time windowed data structures for interactive visualization such as labeling problems or cartograms.

## References

- ATTALI, D., 1997:  $r$ -regular shape reconstruction from unorganized points. Proceedings of the Thirteenth Annual Symposium on Computational Geometry, Nice, France, June 4-6, 248-253.
- BANNISTER, M.J., DEVANNY, W.E., GOODRICH, M.T., SIMONS, J.A. & TROTT, L., 2014: Windows into geometric events: Data structures for time-windowed querying of temporal point sets. Proceedings of the 26th Canadian Conference on Computational Geometry, CCCG 2014, Halifax, Nova Scotia, Canada, 11-19.
- BANNISTER, M.J., DUBOIS, C., EPPSTEIN, D. & SMYTH, P., 2013: Windows into relational events: Data structures for contiguous subsequences of edges. Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2013, New Orleans, Louisiana, USA, January 6-8, 2013, 856-864.
- BOKAL, D., CABELLO, S. & EPPSTEIN, D., 2015: Finding all maximal subsequences with hereditary properties. Proceedings of the 31st International Symposium on Computational Geometry, SoCG 2015, June 22-25, 2015, Eindhoven, Netherlands, 240-254.
- BONERATH, A., HAUNERT, J.-H. & NIEDERMANN, B., 2019: Computing alpha-shapes for temporal range queries on point sets. Proceedings of the 35th European Workshop on Computational Geometry, EuroCG 2019, March 18-20, Utrecht, Netherlands,
- BONERATH, A., NIEDERMANN, B. & HAUNERT, J.-H., 2019: Retrieving alpha-shapes and schematic polygonal approximations for sets of points within queried temporal ranges. Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, SIGSPATIAL 2019, Chicago, IL, USA, November 5-8, 249-258.
- BERG, M., MEULEMANS, W. & SPECKMANN, B., 2011: Delineating imprecise regions via shortest-path graphs. Proceedings of the 19th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems, SIGSPATIAL 2011, November 1-4, Chicago, IL, USA, 271-280.

- CHAN, T.M. & PRATT, S., 2015: Time-windowed closest pair. Proceedings of the 27th Canadian Conference on Computational Geometry, CCCG 2015, Kingston, Ontario, Canada, August 10-12, 141-144.
- CHAN, T.M. & PRATT, S., 2016: Two approaches to building time-windowed geometric data structures. Proceedings of the 32nd International Symposium on Computational Geometry, SoCG 2016, June 14-18, Boston, MA, USA, 28:1-28:15.
- CHANCHARY, F. & MAHESHWARI, A., 2019: Time windowed data structures for graphs. Journal of Graph Algorithms and Applications, **23**(2), 191-226.
- CHANCHARY, F., MAHESHWARI, A. & SMID, M., 2017: Querying Relational Event Graphs Using Colored Range Searching Data Structures. Discrete Applied Mathematics, CALDAM 2017, Lecture Notes in Computer Science, **10156**, Springer, 83-95.
- CHANCHARY, F., MAHESHWARI, A. & SMID, M., 2018: Window Queries for Problems on Intersecting Objects and Maximal Points\*. Algorithms and Discrete Applied Mathematics, CALDAM 2018. Lecture Notes in Computer Science, **10743**, Springer, 199-213.
- CHAZELLE, B., 1986: Filtering Search: A New Approach to Query-Answering. SIAM Journal on Computing, **15** (3), 703-724.
- DUCKHAM, M., KULIK, L., WORBOYS, M. & GALTON, A., 2008: Efficient generation of simple polygons for characterizing the shape of a set of points in the plane. Pattern recognition, **41**(10), 3224-3236.
- EDELESBRUNNER, H., 1992: Weighted alpha shapes. Technical Report UIUCDCS-R-92-1760, University of Illinois at Urbana-Champaign, Department of Computer Science.
- EDELESBRUNNER, H., 2010: Alpha shapes - a survey. Tessellations in the Sciences, **27**, 1-25.
- EDELESBRUNNER, H., KIRKPATRICK, D., & SEIDEL, R., 1983: On the shape of a set of points in the plane. IEEE Transactions on Information Theory, **29**(4), 551-559.
- JONES, C., PURVES, R., CLOUGH, P. & JOHO, H. 2008: Modelling vague places with knowledge from the web. International Journal of Geographical Information Science, **22**(10), 1045-1065.
- LIANG, J., EDELSBRUNNER, H., FU, P., SUDHAKAR, P. & SUBRAMANIAM, S., 1998: Analytical shape computation of macromolecules: I. molecular area and volume through alpha shape. Proteins: Structure, Function, and Bioinformatics, **33**(1), 1-17.
- PENG, D. & WOLFF, A., 2014: Watch Your Data Structures! Proceedings of the GIS Research UK 22nd annual conference, GISRUK 2014. Glasgow, Scotland.
- PURVES, R., CLOUGH, P. & JOHO, H., 2005: Identifying imprecise regions for geographic information retrieval using the web. Proceedings of the 13th Annual GIS Research UK Conference, Glasgow, Scotland, 313-318.
- RUDI, A.G., 2017: Answering Time-Windowed Contiguous Hotspot Queries. Proceedings of the 1st Iranian Conference on Computational Geometry, ICCG, February 27, Teheran, Iran.
- VAUHKONEN, J., KORPELA, I., MALTAMO, M. & TOKOLA, T., 2010: Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. Remote Sensing of Environment, **114**(6), 1263-1276.

# Estimation of River Discharge Using Satellite Altimetry and Optical Remote Sensing Images

DANIEL SCHERER<sup>1,2</sup>, CHRISTIAN SCHWATKE<sup>2</sup> & PETER KRZYTEK<sup>1</sup>

*Abstract: In this paper, a new approach for estimating the discharge of large rivers based on long-term remote sensing data and using the Manning equation is presented. The key idea is to observe the river's cross-sectional geometry from the combination of satellite altimetry and water masks extracted from optical remote sensing imagery. The water surface heights measured by the satellite altimeter missions (Envisat, Jason-2/-3) are combined with monthly water surface masks from Landsat and Sentinel-2 satellite imagery. By fusing both observations, the river bathymetry is estimated based on an adapted hypsometric function and a predicted river bed height. A linear adjustment of altimetry data measured at various virtual stations provides the required flow gradient. Finally, the roughness coefficient is calculated from substantiated compensation factors for geomorphologic features observable with satellite imagery. Note that all mandatory parameters are estimated based on remote sensing data. The validation of the method at the Lower Mississippi River shows that uniform and straight river sections are the most suitable for this new methodology. At six of seven suitable river sections, the Normalized Root Mean Square Error (NRMSE) varied between 12.59% and 26.16%. The Nash-Sutcliffe Efficiency (NSE) was in a range from 0.710 to 0.923. The NRSME deteriorates at nine non-uniform river sections between 38.17% and 108.25%.*

## 1 Motivation

Water is an essential element on Earth. The global water cycle influences the climate decisively. Aside from its use for drinking and hygiene, most of the total surface freshwater withdrawals are used for thermoelectric power, agricultural irrigation, mining and other industries (MAUPIN ET AL 2014). However, freshwater is only 2.5% of the Earth's water reserve and only 0.3% is stored in lakes or flows in rivers, while the rest is bound in permanent ice or groundwater (GLEICK 2012). While the water cycle is affected by global warming, it also influences the climate. As water cycles between the land, oceans, and atmosphere, it changes the dynamics of the climate system (CHAHINE 1992). Hence, changes in the water cycle result in climate change. River discharge measurements are essential for water management purposes, such as watershed protection and computing storm runoff volumes, runoff predictions, or reservoir storage and using them as an input for hydrological models (HUNGER & DÖLL 2008; NRCS 2015). Despite the need for increased attention to the global water cycle and resources, continuous in-situ river discharge measurements declined from around 6000 stations in 1979 to 1000 stations in 2009 (HANNAH et al. 2011). Because of the increasing lack of in-situ measurements, there is a strong motivation to derive river discharge from remote sensing data. Based on previously sampled in-situ discharge

---

<sup>1</sup> Hochschule für angewandte Wissenschaften München, Fakultät für Geoinformation, Karlstr. 6, 80333 München, E-Mail: [dscherer, peter.krzystek]@hm.edu

<sup>2</sup> Technische Universität München, Deutsches Geodätisches Forschungsinstitut DGFI-TUM, Arcisstr. 21, 80333 München, E-Mail: [daniel.scherer, christian.schwatke]@tum.de

measurements, TOURIAN et al. (2017) predicted the discharge using satellite altimetry water level measurements. However, it is a challenge to estimate discharge purely from remote sensing data, as in-situ data for calibration is not available everywhere and long-term changes of the river morphology are not represented by the rating curve. BJERKLIE et al. (2003) demonstrated the main problems of measuring discharge from remote sensing data, which is the estimation of the river's flow velocity and depth. From satellite altimetry, only water level variations above the baseflow, the lowest occurred water level, are observable. However, the flow velocity depends on the whole geometry of the river cross-section and thus, the depth is required for the discharge estimation.

## 2 Study Area and Data

Sixteen locations along the Lower Mississippi River between river mile 276 and 340 were chosen as the study area. Figure 1 shows a map of the gage locations and satellite altimetry.

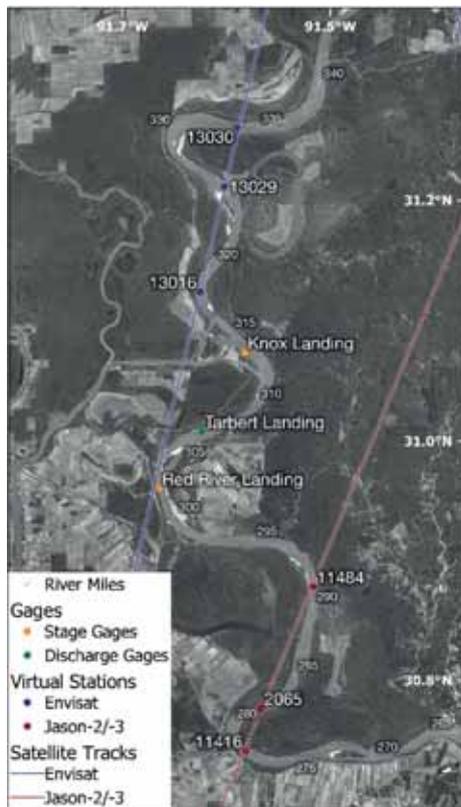


Fig. 1: Overview of gage locations and satellite altimetry

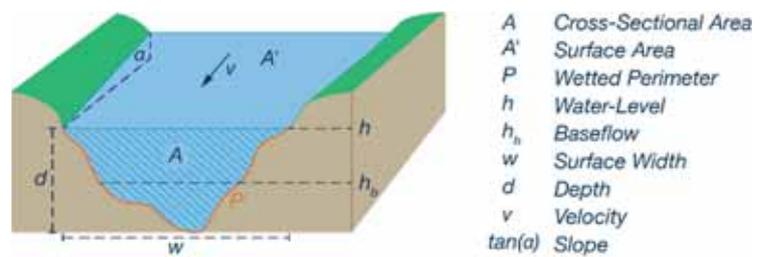


Fig. 2: At-a-station hydrographic parameters

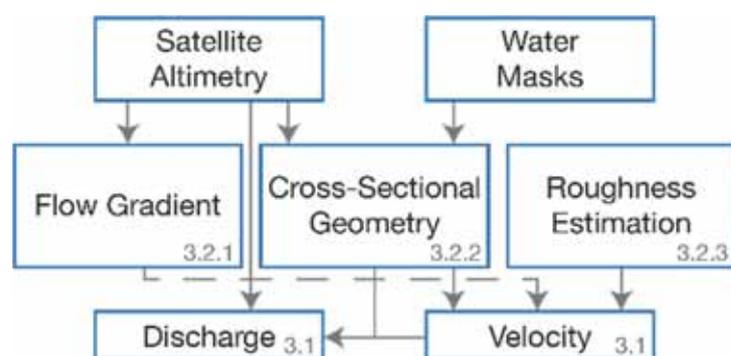


Fig 3: Methodology flowchart

## 2.1 Remote Sensing Data

Satellite altimetry data from the Jason-2/-3 and Envisat missions were downloaded from the Database for Hydrological Time Series of Inland Waters (DAHITI) for six virtual stations in the study area where the satellite tracks cross the Mississippi River (SCHWATKE et al. 2015). Two additional upstream stations were used for the estimation of the flow gradient. Jason-2/-3 data was available from 2008 to 2019 with a repeat cycle of 10 days and Envisat data was available from 2002 to 2010 with a repeat cycle of 35 days.

Monthly water masks and surface area time series from the Automated Water Area Extraction Tool (AWAX) were used to determine the river widths and surface area. AWAX uses satellite imagery from Landsat and Sentinel-2 missions to measure the time-variable water surface of a given stretch of water. The spatial resolution of AWAX depends on the image availability. Landsat data is available since 1984 and has a resolution of 30 m. The Sentinel-2 mission launched in 2015 and provides imagery with a spatial resolution of 10 m. Cloud obscured pixels are reconstructed using a derived long-term water probability mask (SCHWATKE et al. 2019).

## 2.2 In-Situ Data

Within the study area, river discharge is measured at the Tarbert Landing discharge range and the water level is measured at Red River Landing and Knox Landing. The time series data of those gages were used for validation. The main reason for selecting this study area is that important bathymetric data is available in order to validate the estimated cross-sectional geometry. To evaluate the estimated flow gradient, the Low Water Reference Plane (LWRP) was used, which establishes a common hydraulic based datum along the river for civil engineering projects. Additionally, measured river velocities were utilized for validation. All the in-situ data were available on the U.S. ARMY CORPS OF ENGINEERS (2019) Mississippi Valley Division website.

## 3 Methodology

Discharge cannot be measured directly but depends on several hydrographic parameters, which are shown in Figure 2. All cross-sectional parameters that are functions of stage are elements of the at-a-station hydraulic geometry (JULIEN 2018). For discharge estimation, it is key to know the water velocity  $v$  and the cross-sectional area  $A$  of the river at the study site. Some hydrographic variables like river width  $w$  and surface area  $A'$  can be measured from satellite imagery, and the water level  $h$  can be derived from satellite altimetry. The flow velocity cannot be measured by these instruments. Therefore, it must be calculated from hydrologically established formulas like the Gauckler-Manning-Strickler formula. This requires a roughness parameter, the slope of the river  $\tan(\alpha)$  and the cross-sectional geometry expressed as the relationship of the wetted perimeter  $P$  and  $A$ . Yet, the cross-sectional geometry cannot be observed below the minimum water level, which is the baseflow  $h_b$  and has to be predicted by a hypsometric relation.

Figure 3 shows a flowchart of the processing steps, the input data and the describing sections.

### 3.1 Fundamental Equations

As described by AIGNER & BOLLRICH (2015), the fundamental equation for calculating the discharge  $Q$  at a river cross-section is:

$$Q = \bar{v} \cdot A \quad (1)$$

where  $\bar{v}$  is the depth-averaged flow velocity [m/s], and  $A$  is the cross-sectional area [m<sup>2</sup>]. Both parameters cannot be measured from remote sensing data directly. In order to estimate parameter  $A$ , the cross-sectional geometry is constructed from numerous combinations of water level measurements from satellite altimetry and river widths extracted from remote sensing imagery (Section 3.2.2). Parameter  $\bar{v}$  is estimated using the Gauckler-Manning-Strickler formula:

$$\bar{v} = k_{st} \cdot R^{\frac{2}{3}} \cdot I^{\frac{1}{2}} \quad (2)$$

where  $k_{st}$  is the roughness coefficient [m<sup>1/3</sup>/s],  $I$  is the flow gradient and  $R$  is the hydraulic radius [m] which is expressed as  $R = A/P$ , with  $P$  as the wetted perimeter [m].

### 3.2 Parameter Estimation

#### 3.2.1 Flow Gradient

The flow gradient  $I$  is determined from multiple satellite altimetry measurements at eight virtual stations along the river by a linear adjustment. The time differences between the measurements are used for the weighting function.

#### 3.2.2 Cross-Sectional Geometry

Cross-sectional geometry is required for estimating the parameters  $A$ ,  $P$  and  $R$ . Void-free monthly water masks from the AWAX algorithm are combined with monthly averaged water level measurements from satellite altimetry. For every water pixel, the lowest water level is used to set up a bathymetric raster for the extent of the water masks, representing the bathymetry down to the lowest observed water level, the observed baseflow. Because the acquisition time of Landsat imagery is backwards the satellite altimetry missions numerous water masks lack synchronized water level measurements. Therefore, a hypsometric curve is fitted to the available data to use this additional data as well. This requires the estimation of the river depths using an empirical width  $w$  to depth  $d$  relationship based on a study by MOODY & TROUTMAN (2002). By reshaping their obtained relations:

$$\begin{aligned} \bar{w} &= 7.2Q^{0.50 \pm 0.02} \Leftrightarrow Q = w^2/7.2^2 \\ \bar{d} &= 0.27Q^{0.39 \pm 0.01} \end{aligned} \quad (3)$$

The depth can be estimated with:

$$\bar{d} = 0.27(w^2/7.2^2)^{0.39} \quad (4)$$

To estimate the river bed elevation  $h_0$ , synchronized data from satellite altimetry and water masks are used to get the monthly width  $w_m$  and water level  $h_m$  and calculate the monthly depth  $d_m$ . Finally, the river bed elevation is averaged from all observed differences between  $h_m$  and  $d_m$  and subtracted from all the water level measurements. The following logistic function is fitted to the observed synchronized water level and surface area  $a$  data:

$$h(a) = \frac{y_1 - y_0}{1 + e^{-m \cdot (a-b)}} + y_0 \quad (5)$$

where  $y_1$  is the maximum and  $y_0$  is the minimum of the curve,  $e$  is the Euler's number,  $m$  is the slope and  $b$  represents the position of the curves' midpoint. The data is weighted by the errors estimated from the AWAX algorithm to improve the fitting. Using the predicted water levels, the bathymetric raster can be expanded below the observed baseflow down to the lowest predicted water level, the predicted baseflow. The remaining gap between the predicted baseflow and the estimated bed elevation is filled with a triangular shape.

### 3.2.3 Roughness Coefficient

It turned out that the estimation of the roughness coefficient  $k_{st}$  is the most challenging part of the method. In this study, a constant value is used, which is actually changing with the water level. ARCEMENT & SCHNEIDER (1989) described a method to determine the roughness coefficient based on different adjustment factors:

$$k_{st} = ((n_b + n_1 + n_2 + n_3 + n_4)m)^{-1} \quad (6)$$

where  $n_1$  is a correction factor for surface irregularities,  $n_2$  is a value for variations in shape and size of the channel cross-section,  $n_3$  is a value for obstructions,  $n_4$  is a value for vegetation and flow conditions, and  $m$  is a correction factor for the channel meandering. The base value  $n_b$  depends on the channel type, its bed material and the grain size. It is too ambitious to determine the grain size from remote sensing, but the channel material can be estimated from high-resolution satellite imagery. Quantitative geomorphologic methods, as described by SCHUMM (1977) and a decision guide by ARCEMENT & SCHNEIDER (1989), help to determine the adjustment factors. The discharge estimation can be calibrated by adjusting the roughness coefficient.

## 4 Case Study

### 4.1 Closed Loop Test

A closed loop test was performed at Tarbet Landing to validate the methodology and the significance of the parameters, thereby showing that they can be substituted by remote sensing data. As different water-level data products were used at shifted locations along the river, all heights were adjusted along the LWRP.

#### 4.1.1 Fundamental Equation

First, Equation 1 was used to calculate the discharge based on the in-situ water level. The multibeam bathymetric data was used to determine the cross-sectional area  $A$ . The in-situ velocity to water level relationship was used as mean velocity  $\bar{v}$ . Compared to the in-situ discharge time series, the NRMSE was 9.72%, and the NSE was 0.963.

#### 4.1.2 Satellite Altimetry

The in-situ water level time series were substituted by satellite altimetry data to evaluate the usability of inland satellite altimetry. Depending on the virtual station used as input time series, the NRMSE increased up to 23.87%.

#### 4.1.3 Gauckler-Manning-Strickler Formula

Next,  $\bar{v}$  was estimated using Equation 2. A constant  $k_{st}$  of 36.23 was calculated using the adjustment factors  $n_b=0.02$ ,  $n_1=0.001$ ,  $n_2=0.001$ ,  $n_3=0$ ,  $n_4=0.002$  and  $m=1.15$ . The NRMSE was 8.94% with in-situ water levels as input and the flow gradient  $I$  from adjusted altimetry measurements, which was  $50 \cdot 10^{-6}$  at Tarbert Landing and steeper than the gradient from the LWRP ( $22 \cdot 10^{-6}$ ) and the gradient from adjusted in-situ water level measurements ( $39 \cdot 10^{-6}$ ).

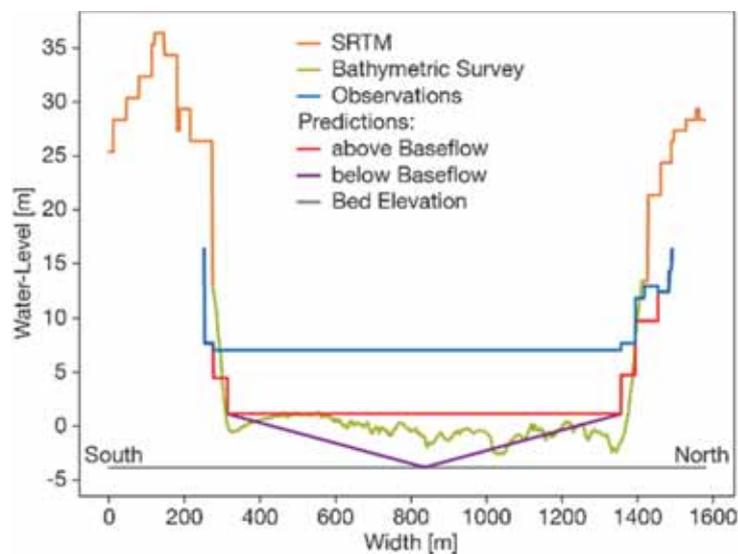


Fig. 4: Observed, predicted and surveyed cross-sectional geometry at Tarbert Landing

#### 4.1.4 Cross-Sectional Geometry

Using cross-sectional geometry derived from in-situ water level data and AWAX water masks to estimate  $A$ , the NRMSE was 10.08%. With a geometry based on altimetry water level data, the NRMSE was between 9.18% and 34.31% depending on the used virtual station. Figure 4 shows the geometry constructed from water masks combined with Jason-2/-3 altimetry observations and predictions and the in-situ bathymetry from multi-beam bathymetric data.

### 4.2 Full Remote Sensing Results

Finally, discharge was estimated for 16 cross-sections solely using remote sensing data. The cross-sections were chosen by the location of in-situ gages, virtual stations or geomorphologic features, such as straight or widening river sections. The cross-sectional geometry was estimated as described in Section 3.2.2, and the flow velocity was calculated with the Gauckler-Manning-Strickler formula based on the flow gradient from adjusted altimetry measurements and the constant  $k_{st}$  of 36.23. In total, the NRMSE ranged from 12.59% to 108.25%. However, it was possible to classify the cross-sections by channel form based on the bathymetric survey data. For

cross-sections that were heavily dredged by the river, the NRMSE was 38.17% or greater. These sections were all located in curved river sections, which can be avoided in future studies. The best results were achieved in straight and widened river segments, where the cross-sectional form was uniform or the river sections had multiple channels. For seven of eight such river segments, the NRMSE was 26.16% or less. Figure 5 shows an exemplary discharge time series for Tarbert Landing.

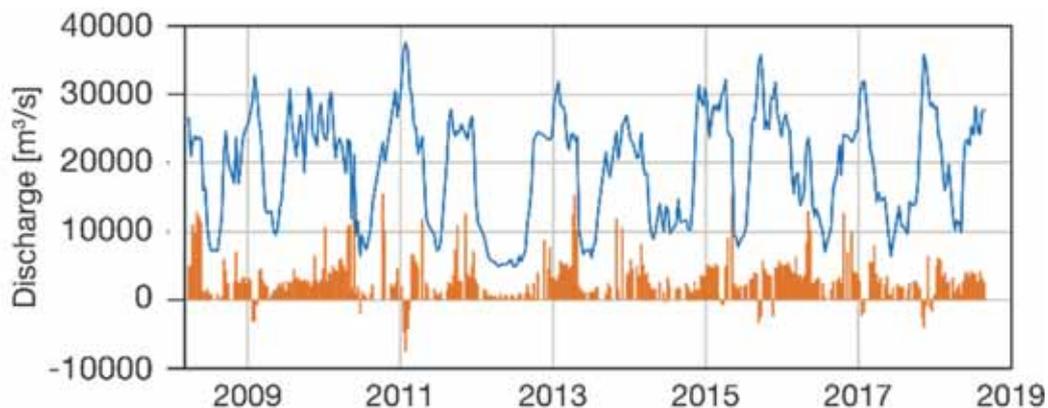


Fig. 5: Resulting discharge (blue) and errors (orange) based on remote sensing data for Tarbert Landing

## 5 Conclusion and Outlook

The study showed that estimation of river discharge solely using satellite data is possible. The best results could be achieved at straight and widened river segments, which can be identified using satellite imagery. At six of seven suitable river sections, the NRMSE varied between 12.59% and 26.16%. The NSE ranged from 0.710 to 0.923. The roughness coefficient was the weakest parameter. In further studies, the Gauckler-Manning-Strickler equation should be exchanged for other methods with dimensionless roughness coefficients. Calibration would improve the resulting discharge time series, e.g. adjusting the roughness parameter to minimize the RMSE.

## 6 References

- AIGNER, D. & BOLLRICH, G., 2015: *Handbuch der Hydraulik. Für Wasserbau und Wasserwirtschaft*. 1. Aufl. Beuth Wissen. s.l.: Beuth Verlag GmbH.
- ARCEMENT, G.J. & SCHNEIDER, V. R., 1989: *Guide for selecting Manning's roughness coefficients for natural channels and flood plains*. USGS.
- BJERKLIE, D.M., DINGMANN, S.L., VOROSMARTY, C.J., BOLSTER, C.H. & CONGALTON, R., 1989: Evaluating the potential for measuring river discharge from space. *Journal of Hydrology* **278**(1-4), 17-38.
- CHAHINE, M.T., 1992: The hydrological cycle and its influence on climate. *Nature* **359**(6394), 373-380.

- GLEICK, P.H., 2012: Water Resources. Encyclopedia of climate and weather. 2. ed. Vol. 2. S.H. Schneider (Editor), New York: Oxford Univ. Press, 817-823.
- HUNGER, M. & DÖLL, P., 2008: Value of river discharge data for global-scale hydrological modeling. *Hydrology and Earth System Sciences* **12**(3), 841-861.
- HANNAH, D.M., DEMUTH, S., VAN LANEN, H.A.J., LOOSER, U., PRUDHOMME, C., REES, G., STAHL, K. & TALLAKSEN, L.M., 2011: Large-scale river flow archives: importance, current status and future needs. *Hydrological Processes* **25**(7), 1191-1200.
- JULIEN, P.Y., 2018: River Mechanics. Second Edition. New York: Cambridge University Press.
- MAUPIN, M.A., KENNY, J.F., HUTSON S.S., LOVELANCE, J.K., BARBER, N.L. & LINSEY, K.S. 2014: Estimated use of water in the United States in 2010, **1405**. Circular. Reston Virginia: U.S. Geological Survey.
- MOODY, J.A. & TROUTMAN, B.M., 2002: Characterization of the spatial variability of channel morphology. *Earth Surface Processes and Landforms* **27**(12), 1251-1266.
- NRCS (NATURAL RESOURCES CONSERVATION SERVICE), 2015: National Engineering Handbook Part 630 Hydrology. Chapter 5: Streamflow Data. Fort Worth, Texas: U.S. Department of Agriculture.
- SCHWATKE, C., DETTMERING, D., BOSCH, W. & SEITZ, F. 2015: DAHITI - an innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry. *Hydrology and Earth System Sciences* **19**(10), 4345-4364.
- SCHWATKE, C., SCHERER, D. & DETTMERING, D., 2019: Automated Extraction of Consistent Time-Variable Water Surfaces of Lakes and Reservoirs Based on Landsat and Sentinel-2. *Remote Sensing* **11**(9), 1010.
- SCHUMM, S.A., 1977: The fluvial system. A Wiley-Interscience publication. New York: Wiley.
- TOURIAN, M.J., SCHWATKE, C. & SNEEUW, N., 2017: River discharge estimation at daily resolution from satellite altimetry over an entire river basin. *Journal of Hydrology* **546**, 230-247.
- US ARMY CORPS OF ENGINEERS, 2019: Mississippi Valley Division url: [www.mvd.usace.army.mil](http://www.mvd.usace.army.mil)

# Monitoring of Alpine Snow Conditions Using C-Band SAR

CLAUDIO NAVACCHI<sup>1</sup>, BERNHARD BAUER-MARSCHALLINGER<sup>1</sup> & WOLFGANG WAGNER<sup>1</sup>

*Abstract: Information about the state of a snow pack is valuable in many geoscientific applications, e.g., for hydrological run-off models. This study aimed to retrieve such information by linking C-band Sentinel-1 Synthetic Aperture Radar (SAR) data with snow profile measurements in an alpine region. A new technique for normalising the Sentinel-1 backscatter measurements was developed, accounting for the varying observation geometry. With this, a strong relationship between radar backscatter and snow wetness could be established. Moreover, the normalisation method allows to deduce snow parameter information even without well-known change detection techniques, since systematic observation effects cancel out. Using these findings for wet snow mapping, a good agreement with optical imagery could be achieved.*

## 1 Introduction

Aperture synthesis in radar imaging enables the acquisition of high-resolution Earth observation imagery at microwave frequencies (MOREIRA et al. 2013). Most Synthetic Aperture Radar (SAR) systems operate in the microwave C-band (4-8 GHz), which exhibits a significant sensitivity to water, one of the most abundant molecules on Earth. It has proven to be an indispensable way of monitoring processes taking place within the cryosphere, e.g. iceberg monitoring and wet snow extent mapping (POWER et al. 2001, NAGLER & ROTT 2000). Thanks to the recently much improved availability of data from SAR missions and ground based snow observations in terms of spatial and temporal resolution, the relationship between high-resolution SAR backscatter data and snow properties can be studied with unprecedented quality and details.

The aim of this study was to compare snow profile measurements (e.g., snow height, grain size, snow wetness, ...) collected by the *Lawinenwarndienst Tirol* with Sentinel-1 SAR time series. The major hurdle when comparing these two data sources is the pronounced impact of the topography on the Sentinel-1 measurements. Thus, before snow properties can be derived from dense Sentinel-1 time series, representations of backscatter data need to be identified that minimise the impact of orbit geometry and local topography.

This paper starts with an introduction into the state of the art in snow-backscatter modelling and common approaches for radar backscatter normalisation in Section 2. Subsequently, a short overview of the data sets serving as an input for this research is provided. In Section 4, the methodology for preparing and normalising backscatter data to enhance its comparability with respect to snow parameter measurements is described. Finally, obtained results are discussed and the paper concludes with an outlook to future research topics.

---

<sup>1</sup> TU Wien, Department of Geodesy and Geoinformation, Microwave Remote Sensing Research Group, Wiedner Hauptstraße 8, A-1040 Wien, E-Mail: [claudio.navacchi, bernhard.bauer-marschallinger, wolfgang.wagner]@geo.tuwien.ac.at

## 2 State of the Art

### 2.1 Snow-Backscatter Modelling

In microwave remote sensing the interaction of the microwave pulses with the land surface is most commonly modelled using radiative transfer theory. This theory is very complex in the presence of a dense medium as it is the case for an aged snow pack with larger grains, fluid retention and differing layers (cf. Fig. 1).

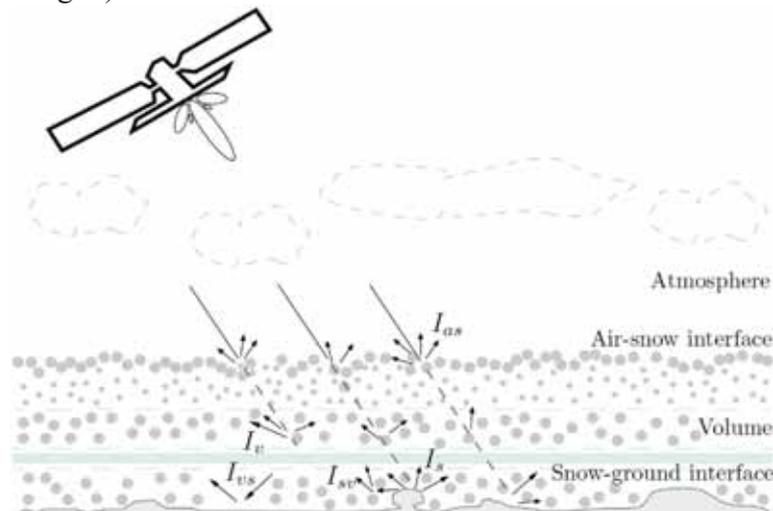


Fig. 1: Interaction of microwave radiation with a multi-layered snow pack

Yet, assumptions about the composition of the snow pack or the propagation characteristics of the radiation allow to derive certain snow parameters. This has been shown in past studies, e.g. an estimation of grain size, snow density and snow depth using the polarimetric properties of SIR-C's X/C-band sensor (SHI & DOZIER 2000), or deriving the snow water equivalence (SWE) from C-band SAR backscatter data (BERNIER & FORTIN 1998). Mapping extents of a wet snow pack is of high relevance in literature primarily being initiated by NAGLER & ROTT (2000). Until recently, change detection with an empirically defined threshold of -3 dB between wet and dry snow persisted as a suitable method for accomplishing this task (NAGLER et al. 2016). It relies on the fact, that dry snow is nearly a transparent medium at C-band frequencies, while wet snow absorbs the microwaves very quickly. Thus, the presence of wet snow, causing scattering at the air-snow interface, can be revealed by choosing a reference image at dry snow or snow-free conditions.

### 2.2 Radar Backscatter Normalisation

Radar backscatter normalisation is about the elimination of the dependency of backscatter on incidence angles. Diverse remote sensing satellite systems collecting multi-angle data made it possible to apply data-driven normalisation techniques. Some of those techniques rely on a high temporal resolution or a special measurement configuration. Scatterometers, such as ERS-SCAT or ASCAT, allow for a direct (i.e. in one overpass) normalisation, since the same area is measured almost simultaneously with multiple beams under different observation angles (NAEMI et al. 2009). For other sensors like ASAR on ENVISAT, dense backscatter and incidence angle time series data is used to estimate a slope and intercept parameter over the whole observation archive, based on

linear regression (LR). These parameters can then be used to tilt the backscatter distribution with respect to a certain reference incidence angle, which is generally chosen to minimise extrapolation errors. Regarding Sentinel-1, this estimation is more difficult due to a smaller range of incidence angles covered by the measurements resulting from the more stringent orbital configuration. Recent attempts have shown success when resampling Sentinel-1 data to 500 m and to estimate the slope parameter through a multivariate linear regression model (BAUER-MARSCHALLINGER et al. 2018). However, this approach could not be transferred to higher-resolution backscatter which is still an open research topic.

### 3 Region of Interest and Data Sets

The region of interest for this study covers the western part of North Tyrol and a small part of South Tyrol in the Austrian/Italian Alps. Based on the availability of snow profile and satellite data, the time span ranges from July 2015 to September 2017, thus comprising two winter seasons. The satellite data consists of Level-1 IW GRDH data from Sentinel-1A/B. It was preprocessed to get backscatter over ground (sigma nought,  $\sigma^0$ ), and radiometric terrain flattened gamma backscatter ( $\gamma_{rtf}^0$ ), which will be explained in more detail in the next chapter. Incidence angles data was also stored during the preprocessing routine to allow backscatter normalisation in downstream processing, which requires information about the topography and the observation geometry.

For two study years, approximately 300 manual snow profile measurements were provided by the *Lawinenwarndienst Tirol*, containing information about snow wetness, snow height, snow hardness, grain size, grain shape, grain type and snow temperature. Measurement locations usually vary, because the data collection is performed by different employees of *Lawinenwarndienst Tirol*, skiers or ski tourers on a voluntary basis. This leads to a broad distribution of data samples in space with the drawback of reducing the temporal comparability. Beside this data set, weather data from *Wetteronline.de* (i.e., rainfall and maximum temperatures) was also considered within our analysis.

## 4 Methodology

### 4.1 Radar Backscatter Normalisation

Radiometric terrain flattening is a relatively new approach to represent backscatter as  $\gamma^0$ , but additionally correcting for terrain, i.e. regions affected by foreshortening (SMALL 2011). SMALL'S (2011) method tries to resolve the one-to-many and many-to-one relationship between the slant and ground range geometry and gives preference to an estimate of the illuminated area instead of applying an incidence angle normalisation. The proposed algorithm starts by integrating the local illuminated area for each point on a DEM. This area is estimated in a plane perpendicular to the slant range, where overlapping regions (e.g., when foreshortening is present) sum up, thus yielding a better estimate of  $\gamma^0$ . Shadow regions are excluded and masked during this procedure as they do not contribute any information.

Radar backscatter normalisation with respect to the incidence angle is another way to account for the influence of terrain, and for some extent, also of land cover. Within the frame of this research, a focus was laid on two methods: linear regression and a novel approach, the Piecewise Linear Percentile Slope (PLPS) method.

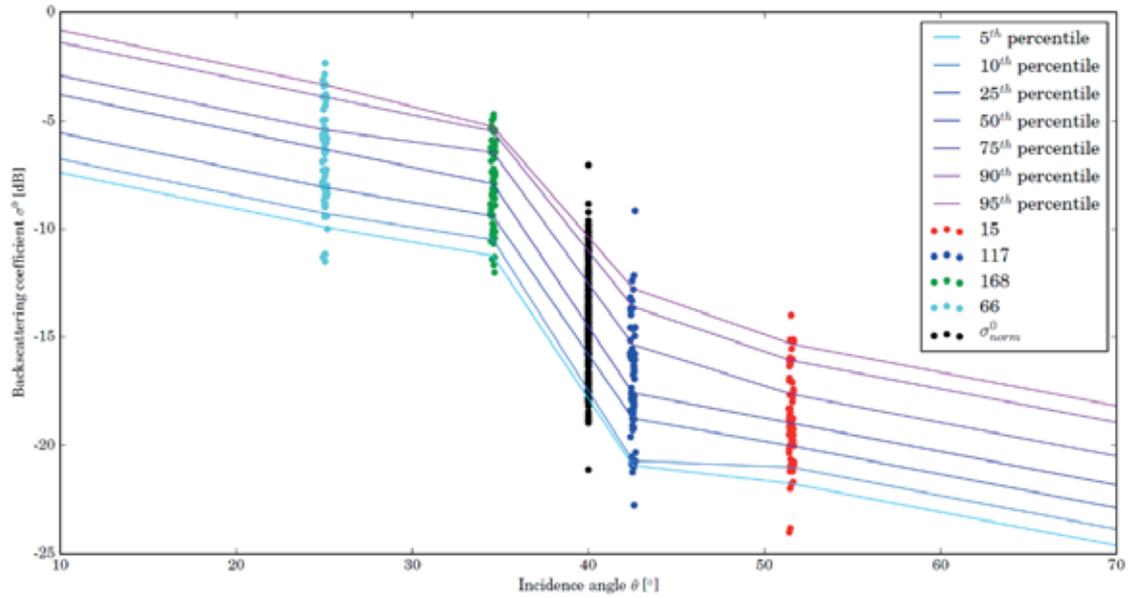


Fig. 2: PLPS normalised backscatter at a reference incidence angle of  $40^\circ$  (black) for a pixel located on a steep slope. Each point colour refers to a certain Sentinel-1 orbit and each line colour to a percentile

The idea of the here presented PLPS method is to discretise the backscatter distribution per orbit, instead of relying only on one single slope estimate for all orbits as it is the case for LR. Percentiles are thought to be the best choice for sampling the distribution, since they offer to derive a slope being dependent on the given backscatter distribution at a given incidence angle and are less influenced by outliers. However, a linear regression between equal percentiles and their related incidence angles does not work for a complex behaviour of backscatter, i.e. a non-linear behaviour along the range of incidence angles. This issue can be solved by going one step further in discretisation and connect each pair of percentiles between neighbouring orbits, which is depicted in Fig. 2. To minimise the necessity of extrapolation,  $40^\circ$  seems to be an appropriate choice for the reference incidence angle as the mean value of the Sentinel-1 incidence angles tends to be around  $40^\circ$ . For further analysis and comparison,  $\sigma^0$  was once normalised with the LR method ( $\sigma_{LR}^{40}$ ) and once with the PLPS method ( $\sigma_{PLPS}^{40}$ ).

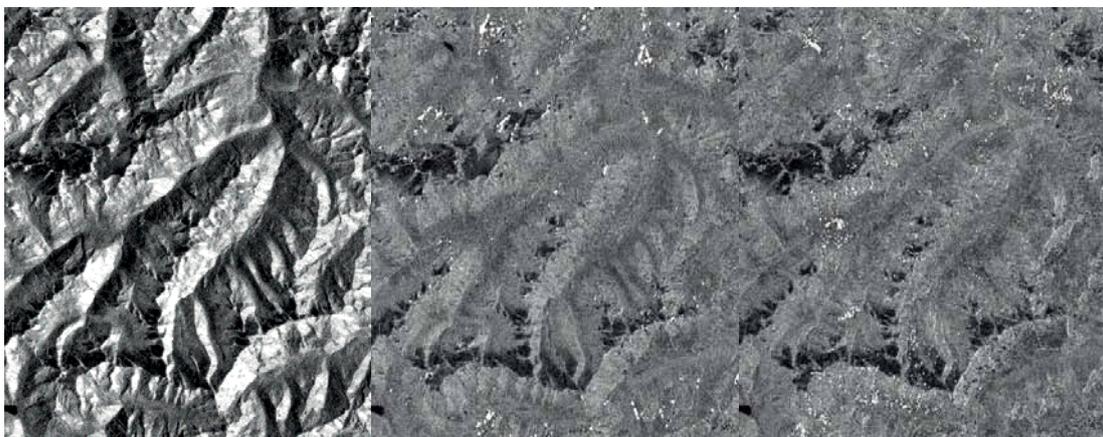


Fig. 3: Comparison between  $\sigma^0$  (left),  $\gamma_{rtf}^0$  (center) and  $\sigma_{PLPS}^{40}$  (right).

Fig 3. underlines the capability of both radiometric terrain flattening and PLPS to reduce the impact of terrain on backscatter. In the left image, the typical illumination pattern of a SAR sensor is visible. Some regions appear very bright, because they are facing towards the sensor (foreshortening), while others are dark as they are situated on a slope turned away from the sensor. This behavior is very well compensated by means of backscatter normalisation as shown in the other two images. Shadow regions are masked for  $\gamma_{rtf}^0$ , whereas  $\sigma_{PLPS}^{40}$  excludes areas where no percentile slope estimation is possible, e.g., when only measurements from two orbits covering a small range of incidence angles are available.

## 4.2 Change Detection

Another way to deal with the influence of the observation geometry is to eliminate systematic effects by performing a subtraction between two backscatter values from the same orbit. By doing so, changes with respect to the chosen reference backscatter value can be detected and related to a geophysical variable, e.g., snow wetness. NAGLER & ROTT (2000) have used change detection for wet snow mapping relying on a selection of a suitable reference image (e.g., at dry-snow or snow-free conditions) to separate backscatter of a snow pack from surface backscatter.

The herein presented reference image selection follows recommendations of NAGLER & ROTT (2000) and NAGLER et al. (2016). In addition, the selection is explicitly assisted by continuously measured weather data. Past studies relied on a selection of one image, which is reasonable for a smaller test site. However, as this paper aims to compare backscatter measurements with snow profile data on a larger scale, a selection of a reference backscatter value is thought to be most reliable at pixel level, in contrast to image level. To detect the most appropriate backscatter values in a time series, meteorological data was interpolated at each pixel. Thereby, Inverse Distance Weighting (IWD) was applied for rainfall data and linear interpolation by height for temperature data. The final selection of a reference backscatter value is based on filtering the interpolated weather data in time, i.e. to find the pixel value with the coldest and driest conditions.

## 5 Results and Discussion

A subset of snow parameters was finally compared to the different backscatter representations, which are  $\sigma^0$ ,  $\gamma_{rtf}^0$ ,  $\sigma_{LR}^{40}$  and  $\sigma_{PLPS}^{40}$ , for both polarisations VV and VH. From Fig. 4, one can conclude that normalised backscatter (lower rows) leads to superior results, with  $\sigma_{LR}^{40}$  performing best. This confirms the expectations of normalised backscatter not only being less dependent on terrain variations but also on the land cover type. Moreover, VH polarisation has a higher correlation for all snow parameters, which underlines the fact of multiple scattering and perhaps also the greater sensitivity with respect to fluid retention in a snow pack. Snow parameters such as air temperature, maximum snow wetness, mean grain size and mean snow wetness are characterised by a larger negative correlation, whereas snow height and the number of layers are positively correlated with backscatter.

However, one has to keep in mind, that there is a large inter-dependence between the individual snow parameters, and hence might influence the backscatter dynamics in a coupled way. This could be the case for snow parameters with a negative correlation, all thought to be dependent on snow wetness. In general, larger grains occur when the snow pack contains wet snow or is governed by past melt-freeze cycles. The same is true for air temperature with a maximum correlation

of -0.55: the warmer, the higher the probability of wet snow. A positive correlation with snow depth is meaningful (more volume scattering), but is not necessarily the only cause of the depicted correlations. It could be the case that the results are affected by the site and the land cover itself, i.e. the general level of the surface backscatter, because of the varying measurement location. Correlations shown in Fig. 4 are based on a single value retrieval, which is disturbed to some extent by speckle, noise or the snow profile measurement procedure, which impacts on the snow pack structure. Building a local average over a certain neighbourhood around the measurement site (around 30m) increased the correlations significantly. Furthermore, by applying change detection based on the generated reference image (referred to as backscattering difference), nearly all backscatter representations at VH polarisation reached a correlation of -0.64 with snow wetness. This behaviour is caused by the difference formation, where steady effects (e.g. terrain) cancel out (cf. Fig. 5, bottom).

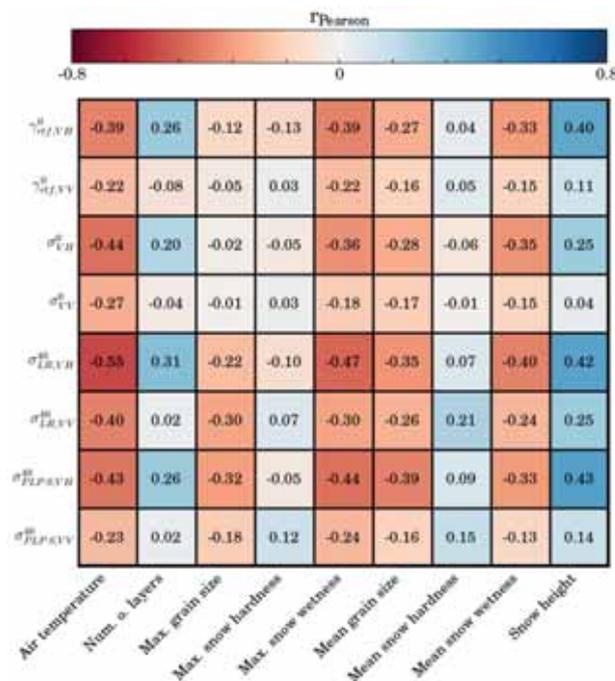


Fig. 4: Pearson's correlation coefficient resulting from relating backscatter to snow parameter data. Visual guidance is given through a colourisation of the correlation, where negative correlation is coloured as red and positive as blue

Fig. 5 depicts a time series of snow and C-band SAR backscatter parameters for a part of the winter season in 2017. One can identify the large variation of  $\sigma^0$  caused by the different viewing geometries (Fig. 5, top). All other backscatter representations line up very well and confirm the success of each normalisation method. As mentioned before, the separation of all backscatter parameters is further reduced when applying change detection (Fig. 5, bottom). Both figures (Fig. 5 top and bottom) show a rather strong, positive correlation between air temperature and maximum snow wetness, with dominant peaks in March and April. They are delineated well by the troughs of the backscatter curves being even more distinctive for the backscattering differences. Applying a harsh threshold of -3 dB as proposed by NAGLER & ROTT (2000) would exclude some measured snow melt events. We therefore suggest to create a continuous classification of backscattering differences below 0 dB.

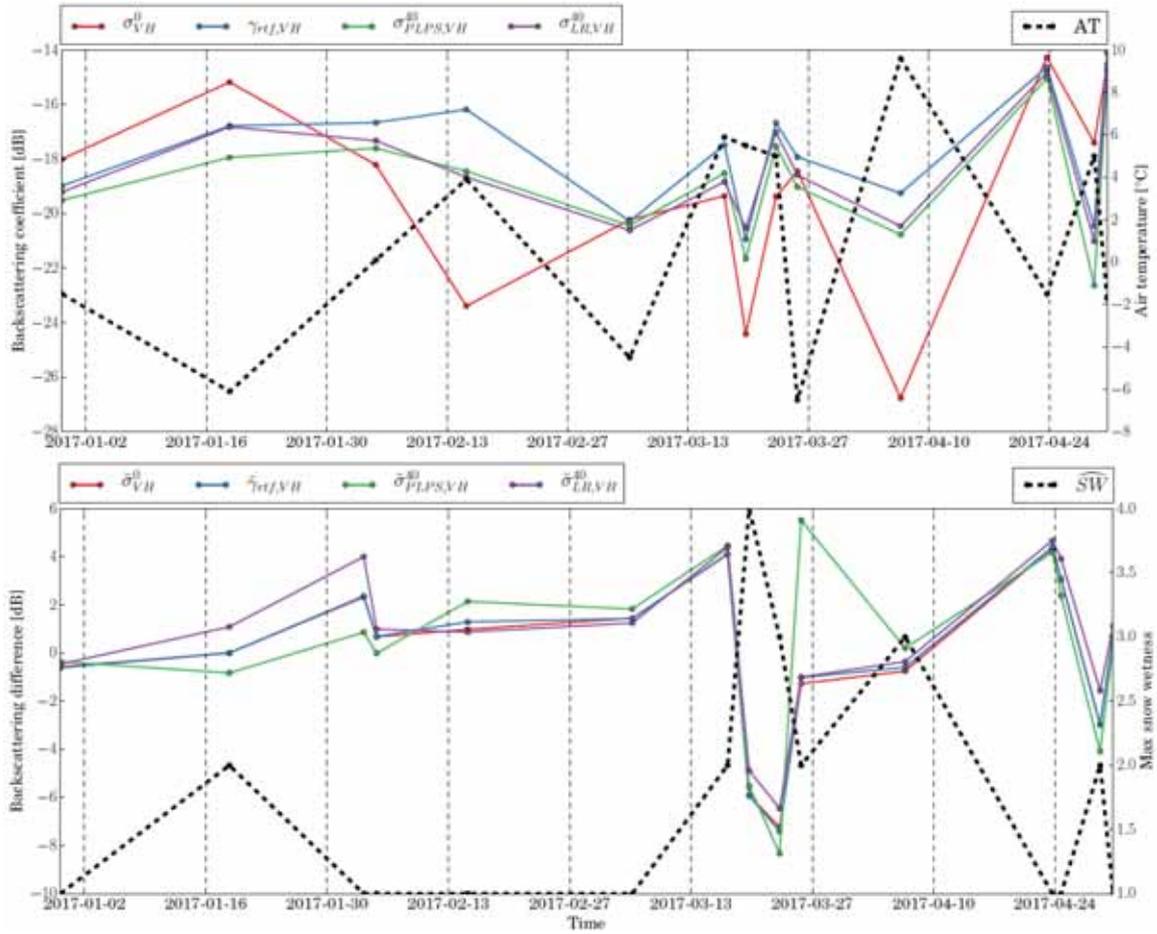


Fig. 5: Comparison between  $\sigma^0$  (red),  $\gamma_{rtf}^0$  (blue),  $\sigma_{PLPS}^{40}$  (green), and  $\sigma_{LR}^{40}$  (violet) once for VH polarized backscatter and air temperature (top) and once for backscattering differences and maximum snow wetness (bottom), respectively

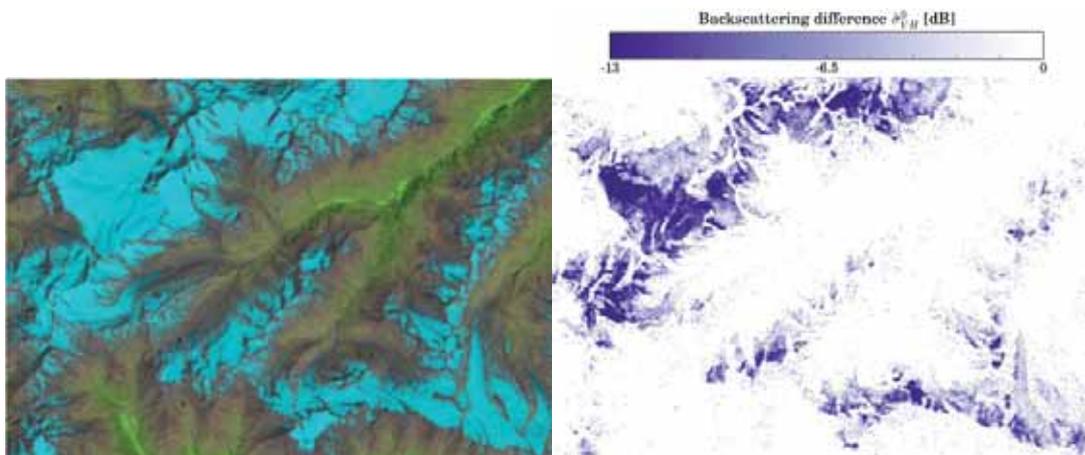


Fig. 6: Comparison between a Landsat 8 false-colour composite (left) and a map indicating wet snow (right)

The aforementioned findings can now be used to apply change detection on the backscatter data resulting in a map, which indicates snow wetness. In Fig. 6, such a map is compared to a Landsat 8 false-colour composite showing good agreement concerning the extent of the snow pack.

## 6 Conclusions and Outlook

In this study the relationship between the properties of alpine snow and C-band SAR backscatter observations from Sentinel-1 was investigated. To minimise the impacts of topography and imaging geometry, Sentinel-1 backscatter measurements were modified by means of radiometric normalisation, a novel incidence angle normalisation method and change detection. Correlations with snow parameters seem to be highest when applying a linear regression model, as it appears to be more robust concerning critical backscatter distributions at high and low incidence angles. Yet, the here presented piecewise linear percentile slope normalisation is considered to be superior if the Sentinel-1 orbits cover a broader range of incidence angles.

The results feature that Sentinel-1 backscatter shows the strongest (negative) correlation to manually measured snow wetness. The relation can be made even more pronounced by using change detection and spatial filtering (-0.64). Other snow parameters are also characterised by a high correlation, but being rather insignificant due to inter-dependencies and a varying land cover.

This work opens up further research questions, e.g., how to model additional orbit effects resulting from different viewing directions in azimuth or how to set up an objective snow parameter measurement configuration to reduce the influence of the surveyor.

## 7 References

- BAUER-MARSCHALLINGER, B., FREEMAN, V., CAO S., PAULIK, C. SCHAUFLER, S., STACHL, T., MODANESI, S., MASSARI, C., CIABATTA, L., BROCCA L., WAGNER, W., 2018: Toward Global Soil Moisture Monitoring With Sentinel-1: Harnessing Assets and Overcoming Obstacles. *IEEE Transactions on Geoscience and Remote Sensing*, **99**, 1-20.
- BERNIER, M. & FORTIN, J-P., 1998: The Potential of Times Series of C-Band SAR Data to Monitor Dry and Shallow Snow Cover. *IEEE Transactions on Geoscience and Remote Sensing*, **36.1**, 226-243.
- MOREIRA, A., PRATS-IRAOLA, P., YOUNIS, M., KRIEGER, G., HAJNSEK, I. & PAPATHANASSIOU, K., 2013: A Tutorial on Synthetic Aperture Radar. *IEEE Geoscience and remote sensing magazine*, **1.1**, 6-43.
- NAEIMI, V., SCIPAL, K., BARTALIS, Z., HASENAUER, S., & WAGNER, W., 2009: An Improved Soil Moisture Retrieval Algorithm for ERS and METOP Scatterometer Observations. *IEEE Transactions on Geoscience and Remote Sensing*, **47.7**, 1999-2013.
- NAGLER, T. & ROTT, H., 2000: Retrieval of wet snow by means of multitemporal SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, **38.2**, 754-765.
- NAGLER, T., ROTT, H., RIPPER, E., BIPPUS, G., & HETZENECKER, M., 2016: Advancements for Snowmelt Monitoring by Means of Sentinel-1 SAR. *Remote Sensing*, **8.4**, 348.

- POWER, D., YODEN, J., LANE, K., RANDELL, C. & FLETT, D., 2001: Iceberg detection capabilities of RADARSAT synthetic aperture radar. *Canadian Journal of Remote Sensing*, **27.5**, 476-486.
- SHI, J. & DOZIER, J., 2000: Estimation of snow water equivalence using SIR-C/X-SAR: II. Inferring snow depth and particle size. *IEEE Transactions on Geoscience and Remote Sensing*, **38.6**, 2475-2488.
- SMALL, D., 2011: Flattening Gamma: Radiometric Terrain Correction for SAR Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, **49.8**, 3081-3093.

# Die Viermastbark Peking in Virtual Reality als neue Form der Wissensvermittlung

DOMINIK TRAU<sup>1,2</sup>, THOMAS P. KERSTEN<sup>1</sup> & FELIX TSCHIRSCHWITZ<sup>3</sup>

*Zusammenfassung: Virtual Reality (VR) hat sich durch die immersive Visualisierung raumbezogener Daten in den letzten Jahren in der Geoinformatik etabliert. Dadurch bietet VR dem Anwender neue Möglichkeiten, sich durch Interaktion mit der VR-Applikation Wissen anzueignen und auch spielerisch zu erarbeiten. In diesem Beitrag wird die Entwicklung einer VR-Applikation vorgestellt, mit der eine neue Form der Wissensvermittlung durch Interaktivität in einem Virtual Reality-System untersucht und umgesetzt wurde. Als Beispiel für diese Umsetzung dient die an der HafenCity Universität Hamburg entwickelte VR-Applikation der Viermastbark „Peking“, welche ab 2023 als restauriertes Museumsschiff ein wichtiges Exponat des zukünftigen Deutschen Hafenmuseums in Hamburg verkörpern wird. In der Applikation ist es möglich, die „Peking“ zu betreten und zu erkunden sowie am Beispiel von drei Szenarien (3D-Modellerstellung, Segel- sowie Flaggenkunde) mit dem VR-Modell zu interagieren, um so viele theoretische Aspekte der Wissensvermittlung durch Interaktivität in die entwickelte VR-Applikation zu integrieren und zu untersuchen. Damit soll die VR-Applikation einen entscheidenden Anteil zum Lernprozess des Anwenders beitragen. In fortgeschrittenen Ansätzen, wie bei durch den Anwender selbständig formulierten komplexeren Fragestellungen, besteht jedoch noch Entwicklungspotential.*

## 1 Einleitung

Die Erschaffung virtueller Realitäten war in der Anfangszeit lediglich größeren Forschungslaboren und militärischen Einrichtungen vorbehalten, wodurch nur beschränkte Einsatzmöglichkeiten existierten (DÖRNER 2013). Durch das Aufkommen preiswerter Virtual Reality Systeme Mitte der 2010er Jahre (Oculus Rift, HTC Vive) sowie verbesserter Verfügbarkeit und Lizenzierungsmöglichkeiten von Spiele-Engines ist diese Technologie heutzutage für die breite Öffentlichkeit zugänglich und soll einen wichtigen Beitrag für die zukünftige Entwicklung der Mediennutzung leisten.

Immer mehr VR-Entwicklungen beschäftigen sich mit nicht mehr existierenden Bereichen oder Objekten, z.B. historische Stadtzustände oder Kulturstätten (DEGGIM et al. 2017; TSCHIRSCHWITZ et al. 2019; WALMSLEY & KERSTEN 2019). Allerdings besitzen sowohl die generierten virtuellen Umgebungen als auch die zur Erzeugung verwendeten Spiele-Engines viel mehr Potential als die bloße Darstellung und Präsentation von Kulturgütern für den Anwender. Beispielsweise kann unter dem Einsatz von interaktiven Elementen auch das Edutainment für ein breiteres, fachlich nicht so versiertes Publikum oder Nutzergruppe gefördert werden. Weiterhin bietet die

---

<sup>1</sup> HafenCity Universität Hamburg, Labor für Photogrammetrie & Laserscanning, Überseeallee 16, D-20457 Hamburg, E-Mail: [Dominik.Trau, Thomas.Kersten]@hcu-hamburg.de

<sup>2</sup> Landesbetrieb Geoinformation und Vermessung, Neuenfelder Straße 19, D-21109 Hamburg, E-Mail: Dominik.Trau@gv.hamburg.de

<sup>3</sup> IB&T Software GmbH, An'n Slagboom 51, D-22848 Norderstedt, E-Mail: Felix.Tschirschwitz@card-1.com

Implementierung von dynamischen Elementen eine Steigerung des immersiven Erlebnisses für den Betrachter.

Am Beispiel der Viermastbark „Peking“ werden unterschiedliche Formen der interaktiven Wissensvermittlung mithilfe verschiedener Elemente entwickelt und untersucht. Diese neue Form des Wissenstransfers beinhaltet sowohl historische Kennwerte des Schiffes als auch Informationen über den Entstehungsprozess der VR-Applikation, denn als Zielgruppe stehen die breite Öffentlichkeit und nicht nur ausgewählte Experten im Fokus.

## 2 Methode

Für die Visualisierung der Viermastbark „Peking“ und zur Erstellung der interaktiven VR-Anwendung wurde die Spiele-Engine Unreal Engine 4.22.3 des Softwareunternehmens Epic Games Inc. eingesetzt. Als Hardwarelösung kommt das Head-Mounted-Display HTC Vive Pro der beiden Hersteller HTC und Valve Corporation zum Einsatz.

Zur Präsentation als Exponat im zukünftigen Deutschen Hafenmuseum wird die „reale“ Viermastbark „Peking“ (Abb. 1, links) in einen Zustand von 1927/1928 zurückversetzt. Für die Erstellung des „virtuellen“ Schiffes „Peking“ (Abb. 1, rechts) sind neben 3D-Modellen und Plänen aus der Schifffahrtsbranche auch durch geodätische Methoden (Photogrammetrie und terrestrisches Laserscanning (TLS)) erfasste Informationen vom Schiff und vom Inventar des Schiffes verwendet worden.



Abb. 1: Die "Peking" 2017 (FREUNDE DER VIERMASTBARK PEKING E.V. 2020) vor der Peters-Werft in Wewelsfleth (links) und die virtuelle Rekonstruktion des Schiffes (rechts)

Die „Peking“ ist eine Viermast-Stahlbark, welche von der Reederei F. Laeisz in Auftrag gegeben und in den Jahren 1910 und 1911 durch die Werft Blohm + Voss in Hamburg gebaut wurde. Das insgesamt über 115 m lange Schiff mit vier bis zu 50 m hohen Masten besitzt eine lange und wechselhafte Geschichte, die ausführlich durch den Verein FREUNDE DER VIERMASTBARK PEKING E.V. (2020) beschrieben wird. In der Zeit bis kurz nach 1927/1928 wurde das Schiff für den Salpetertransport von Chile nach Europa eingesetzt und diente dann nach einigen Umbauten als Schulschiff für Seeoffiziersanwärter. Seit 2017 ist die „Peking“ nach einigen Besitzerwechseln wieder in Deutschland und wird zurzeit in der Peters-Werft in Wewelsfleth (bei Glückstadt)

restauriert, um voraussichtlich ab 2023 eines der Aushängeschilder des zukünftigen Deutschen Hafenumuseums in Hamburg darzustellen.

Nachfolgend werden spezielle Schwerpunkte des Visualisierungsprozesses dargestellt, die sich signifikant von bisherigen Ansätzen in der VR-Visualisierung unterscheiden. Zunächst wird der gesamte Workflow besonders im Hinblick auf die verwendeten Datenformate vorgestellt, bevor die Implementierung von dynamischen und interaktiven Elementen in diesem Projekt näher beschrieben wird.

## 2.1 Entwicklung der VR-Applikation

Als „virtueller Liegeplatz“ der „Peking“ wurde der Hansahafen im Hamburger Stadtteil Kleiner Grasbrook gewählt, da dieser ein visuell geschlossenes Areal bietet und den historischen Ankerplatz während der aktiven Zeit des Schiffes darstellt. Umliegende Objekte wie Gebäude, Kaimauern und Vegetation sind nur rudimentär generiert, um nicht vom Hauptuntersuchungsgegenstand, der Viermastbark „Peking“, abzulenken. Eine zu detaillierte Ausführung kann den Fokus des Betrachters zu sehr auf diese Elemente lenken. Zur Darstellung der Gebäude wurden überwiegend Objekte des Hamburger 3D-Stadtmodells in der Detailstufe LoD 2 (Level of Detail) herangezogen, welche vor allem durch die Sichtbeschränkung in alle Himmelsrichtungen zum immersiven Erlebnis beitragen sollten. Diese Daten sowie das als Untergrund genutzte Digitale Geländemodell mit einer Rasterweite von einem Meter sind frei verfügbar und stammen vom Landesbetrieb Geoinformation und Vermessung Hamburg.

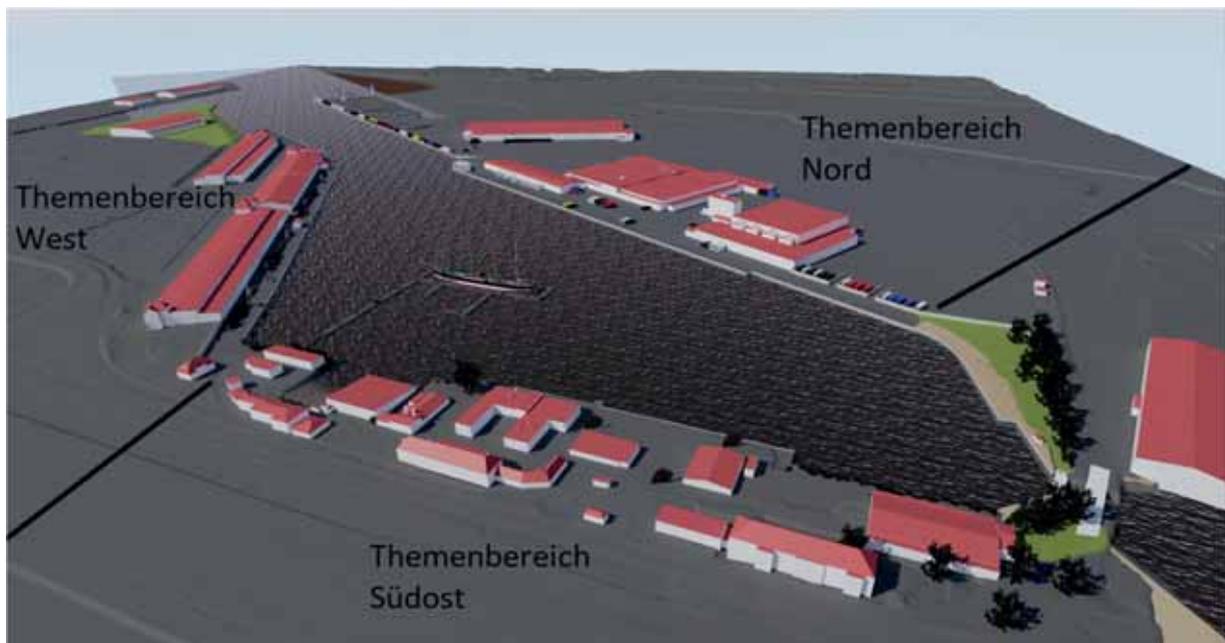


Abb. 2: Die drei im VR-Projekt festgelegten Themenbereiche im Hamburger Hansahafen

Wie in Abb. 2 dargestellt, ist das Areal des Hansahafens in drei Themenbereiche unterteilt. Es ist ausdrücklich nicht das Ziel dieser Gestaltung, exakt die Realität abzubilden, sondern einen Eindruck über die typische, in diesem Bereich vorhandene, Hafencharakteristik zu vermitteln.

Im nördlichen Umfeld wurden neben den vorhandenen Logistikkomplexen vor allem Container, Verladekräne und Mauern der Szene hinzugefügt. Der zweite Themenbereich im Süden umschließt die Zentrale der Hamburg Port Authority (HPA) und ist, gemäß des momentanen Zustandes, überwiegend mit Bäumen und Sträuchern gestaltet. Währenddessen befindet sich der dritte Themenbereich am westlichen Ende des Hafenbeckens und umschließt die historischen Schuppen 50 und 51, welche bereits als Umschlagplatz für Fracht Anfang des 20. Jahrhunderts dienten. Heute sind dort viele Krananlagen und Waggons zu finden, welche in diese VR-Applikation als Ausstattung der Szenerie platziert wurden.

Für dieses Projekt war eine der zentralen Zielsetzungen, die Viermastbark „Peking“ als detailliertes 3D-Modell zu erstellen, um dieses virtuell im FPV (First-Person-View) besuchen und erleben zu können. Zu diesem Zweck wurden für die Erstellung des Schiffsrumpfes zum einen digitale Modelle von an der Restauration der „realen“ Viermastbark „Peking“ beteiligten Unternehmen verwendet, die jedoch unterschiedliche Entwicklungszustände und Schwerpunkte besaßen. Zum anderen konnten auch eine TLS-Punktwolke des Rumpfes sowie der originale Konstruktionsplan aus dem Jahre 1910 herangezogen werden. Aus all diesen Datenquellen sind die jeweils umfangreichsten und detailtreuesten Bereiche selektiert und das Schiff über mehrere Entwicklungsabschnitte virtuell konstruiert worden.

Für die Modellierung der Decksaufbauten (z.B. Abb. 3, rechts) sind vor allem Anregungen vom baugleichen Schwesterschiff, der „Passat“ (Abb. 3, links), herangezogen worden, das als Museumsschiff in Travemünde vor Anker liegt. Zudem lieferte die originale Bauvorschrift der „Peking“ wichtige Informationen. Der generelle Ablauf der CAD-Modellierung auf Grundlage verschiedenster Daten wurde in der Literatur bzw. in weiteren Projekten bereits ausführlich behandelt. Beispielhaft wird dieser Workflow in KERSTEN & LINDSTAEDT (2012a) oder KERSTEN et al. (2013) umfänglich erläutert.



Abb. 3: Blick über das Hauptdeck der "Passat" (links) und über das visualisierte Hauptdeck der "Peking" (rechts)

Die „Peking“ war während der Erstellung dieses Projektes nicht zugänglich, da das Schiff zur Restauration in der Werft liegt (siehe ULRICH 2020). Die einzigen beiden in Hamburg verfügbaren Komponenten des Schiffes (das Hilfssteuerrad im Museum für Hamburgische Geschichte und die Krullgalion im Internationalen Maritimen Museum Hamburg) sind photogrammetrisch erfasst und für Interaktionen auf der virtuellen Viermastbark implementiert

worden. Insgesamt wurden für die Visualisierung des Schiffes und der Umgebung sowie interaktiven Elementen 544 Arbeitsstunden aufgewendet. Ein Großteil hiervon nahm die CAD-Modellierung mit 238 Stunden ein, was im Verhältnis zu der gesamten Arbeitszeit anderen Modellierungsprojekten ähnelt (z.B. das „Alt-Segeberger Bürgerhaus“ in KERSTEN et al. (2013) und der „Kaiserdom Königsutter“ in KERSTEN & LINDSTAEDT (2012b)). Der Anteil der Visualisierung der VR-Applikation (222 Stunden) fällt entsprechend umfangreich aus, da einzelne Konzepte erst entwickelt werden mussten. Bei zukünftigen ähnlichen Projekten ist davon auszugehen, dass der Beitrag am Gesamtarbeitsaufwand aufgrund der gewonnenen Erfahrung geringer sein wird.

Weiterhin werden mithilfe interaktiver Elemente dem Betrachter mehrere Visualisierungsformen in VR präsentiert. In diesem Projekt wurden zwei Objekte exemplarisch ausgewählt, welche als Punktwolke, Dreiecksvermaschung, CAD-Modell und Fotografie dargestellt werden. Hierfür sind die originale Krullgalion und das Hilfssteuerrad der „Peking“ photogrammetrisch mit einer Nikon D800 erfasst und zur Erzeugung der genannten Datenformate entsprechend prozessiert worden. Im Zuge der photogrammetrischen Auswertung konnte eine Genauigkeit der verwendeten Maßstäbe von 0,2 mm (Krullgalion) und 1,8 mm (Steuerrad) zur Skalierung der Punktwolken bestimmt werden, die aber auf die Modellierung des Schiffes keine Auswirkung hat.

Um die Rechenleistung der VR-Applikation durch z.B. animierte Segel zu reduzieren, sind alle Segel des Schiffes in der Standardansicht gerafft. Durch die Interaktion mit der VR-Applikation (siehe auch Kapitel 2.3.) kann ein Matinee (EPIC GAMES 2020), eine Animationssequenz, ausgelöst werden, wodurch das Hissen selektierter Segel aktiviert wird. Innerhalb von fünf Sekunden werden ausgewählte Segelgruppen iterativ skaliert, bis die Segel vollständig gesetzt sind, um so eine unterbrechungsfreie Bewegung zu simulieren (Abb. 4). Durch das Ansprechen und das interaktive Hissen der Segel soll der Anwender eigenständig Informationen bzw. Wissen über die Segel (Position und Bezeichnung) erlernen (siehe Kapitel 2.3.).

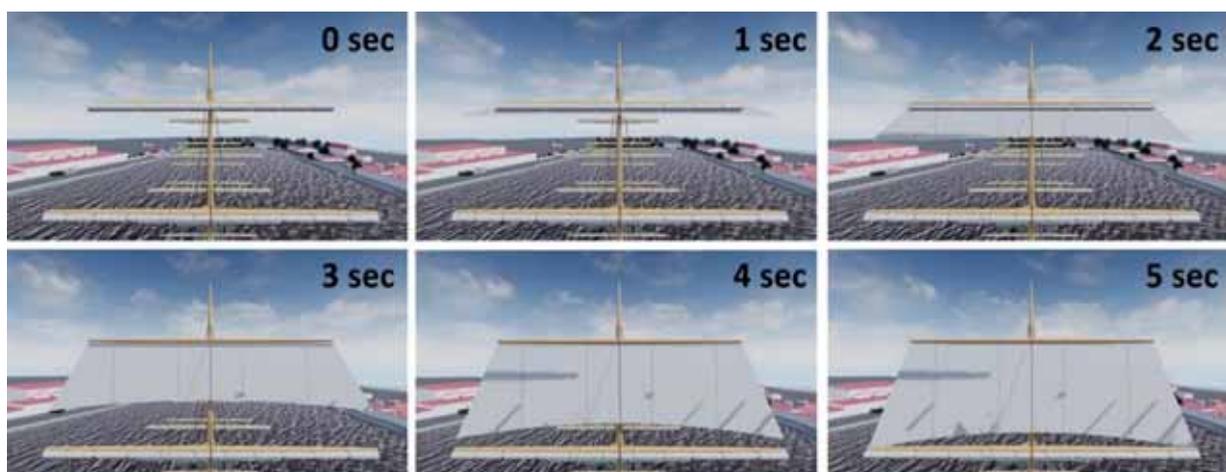


Abb. 4: Hissen des Vorroyalsegels als Matinee-Sequenz

## 2.2 Dynamische Komponenten des Projektes

In einer VR-Applikation ist es zudem möglich, natürliche Bewegungen der Umwelt in visueller Form darzustellen, um die Szenerie „lebendiger“ wirken zu lassen. Denn herkömmliche Anwendungen, die die Realität als visualisiertes 3D-Modell wiedergeben, besitzen zwar die wissenschaftliche Akkuratessse, jedoch ist die Immersion für den Nutzer eingeschränkt. In dieser Anwendung ist zum einen das Verhalten des Wassers realitätsnah imitiert. Durch die Anwendung eines Offsets können bestimmte Punkte der Wasseroberfläche eine andere vertikale Lage im Raum annehmen, sodass eine Bewegung vorgegeben wird. Auch visuell wird dieser Effekt unterstützt, indem je nach Blickwinkel des Betrachters mittels spezieller Beleuchtungstechnik die Spiegelung und Hervorhebung von Strukturen im Wasser unterstützt wurde.

Zum anderen wird an der Takelage, durch das Prinzip des „Verlet-Algorithmus“ (JAKOBSEN 2001) das Wirken externer Kräfte simuliert. So konnten Segel mit einem realistischen physikalischen Verhalten erzeugt werden. Hierfür wurden Ebenen mit jeweils tausenden Teilflächen generiert. Jede dieser Teilflächen wurde klassifiziert und so die an den Rahen befestigten Bereiche als unbeweglich festgelegt, während die restlichen Regionen unter Einwirkung äußerer Kräfte (Wind, Schwerkraft) von der zugewiesenen Position abweichen konnten. Anhand dieses Vorgehens konnten wehende Segel erschaffen werden.

## 2.3 Implementierung interaktiver Elemente

Zusätzlich zu der Visualisierung von Objekten wurden in diesem VR-Projekt noch interaktive Vorgänge eingebunden, um die Immersion des Nutzers mit der Umgebung zu fördern. Die implementierten Interaktionen sowie die resultierenden Prozesse sind im Hinblick sowohl unter Berücksichtigung verschiedenster Formen der Wissensvermittlung als auch auf die Anforderungen an Interaktivität konzipiert worden, um die Eignung von VR-Systemen dafür zu untersuchen.

Im Folgenden werden alle drei entwickelten interaktiven Menüs erläutert. In der VR-Anwendung kann sich der Nutzer durch Teleportation mittels Controller fortbewegen und durch das VR-Modell navigieren (Abb. 5, rechts). Beim Eintreten in markierte Bereiche erscheinen an festgelegten Stellen interaktive Auswahlmöglichkeiten, bei welchen der Betrachter mithilfe der Controller einzelne Button auswählen kann, um an Informationen und damit verbunden an Wissen zu gelangen. Als Reaktion resultiert immer eine Veränderung der Umgebung und die Anzeige von zusätzlichen Informationen in Form von Bildern, 3D-Modellen und Animationen. Dieser Vorgang stellt einen erheblichen Vorteil gegenüber der Informationsaufnahme in der Realität dar, da auf einem echtem Segelschiff solche Auswirkungen bzw. Animationen nicht „per Knopfdruck“ möglich sind.

In Abb. 5 (links) ist das erste interaktive Menü dargestellt. Mit dieser Interaktion kann der Nutzer alle mitgeführten Flaggen der „Peking“ betrachten und nähere Informationen darüber erlangen. Der Aufbau des Layouts ähnelt den beiden nachfolgenden interaktiven Elementen stark, lediglich ein einleitender Text ist nicht vorhanden. Mit der Auswahl eines Button wird die jeweilige Flagge am Fahnenmast am Heck des Schiffes gehisst und im rechten Bereich des interaktiven Menüs erscheint ein erläuternder Text speziell zu dieser Flagge der „Peking“.



Abb. 5: Das interaktive Menü "Flaggen der Peking" (links) und die gezielte Fortbewegung durch Teleportation im VR-Modell (rechts)

Das zweite implementierte Menü hat Informationen über die Segel der „Peking“ (Abb. 6) zum Thema. Wie alle weiteren Interaktionsmöglichkeiten ist das Layout in drei Abschnitte eingeteilt. Zunächst zur Orientierung befindet sich eine Überschrift im oberen Bereich der Komponente. Daraufhin ist immer ein einleitender und erklärender Text gegeben, bevor interaktive Elemente einen Großteil der Aufmerksamkeit einnehmen. Neben der Auswahl einiger historischer Fotografien auf der linken Seite kann der Anwender einen, mehrere oder alle Checkboxes der rechten Grafik auswählen. Daraufhin werden die Segel des jeweiligen Mastes zur Wissensvermittlung farblich hervorgehoben, so dass Bezeichnung und Position der Segel erlernt werden können. Zeitgleich beginnt der Hissvorgang, wie in Kapitel 2.1. beschrieben, am visualisierten Modell. Durch nochmalige Interaktion mit einer ausgewählten Checkbox kann der Prozess auch invers durchgeführt werden.



Abb. 6: Die "Peking" unter vollen Segeln (links) und das interaktive Menü "Segel der Peking" (rechts)

Bei dem dritten vorgestellten Menü handelt es sich um die Möglichkeit, Objekte des Schiffes (Krullgalion und Hilfssteuerrades) in verschiedenen Visualisierungsformen darzustellen. Die Interaktion ist so gestaltet, dass der Benutzer beim Betreten der näheren Umgebung der Krullgalion (Abb. 7, rechts) vor dem Schiff und des Hilfssteuerrades auf dem Mittschiffsdeck (Abb. 7, links) unterschiedliche Auswahlmöglichkeiten aktiviert. Der Aufbau des Layouts gleicht dem der anderen interaktiven Menüs. Mithilfe mehrerer Checkboxes kann der Betrachter eine favorisierte Visualisierungsform des Objektes selektieren, woraufhin das Objekt entsprechend vor

dem Nutzer erscheint. Bei dem Format der Punktwolke und der Vermaschung ist es möglich, auch den Detaillierungsgrad zu bestimmen. Zusätzlich erscheint eine erklärende Einheit (Tafel) links neben dem Objekt, das in einen textlichen und visuellen Part eingeteilt ist und über mehrere Seiten den Entstehungsprozess des jeweiligen Datenformates erläutert. Damit erhält der Anwender Erklärungen über Photogrammetrie und CAD-Modellierung.



Abb. 7: Interaktive Menüs "Steuerrad der Peking" (links) und "Krull der Peking" (rechts)

### 3 Ergebnis

Im Anschluss an die Entwicklung der VR-Applikation für die Viermastbark „Peking“ fand eine differenzierte Auseinandersetzung mit den erzielten Resultaten statt. Hierfür sind verschiedene Kriterien zu berücksichtigen. So werden qualitative Merkmale, wie die implementierten interaktiven Elemente, näher betrachtet, bevor die benötigte Rechenleistung unter quantitativen Gesichtspunkten reflektiert wird.

#### 3.1 Interaktive Wissensvermittlung am Beispiel der „Peking“

Dieses Kapitel gliedert sich in zwei Abschnitte. Zunächst werden einige zentrale Theorien der Wissensvermittlung aufgeführt und hinsichtlich der erstellten VR-Applikation untersucht. Im zweiten Abschnitt wird eine stufenweise Steigerung von Interaktivität eingeführt, welche mit den implementierten interaktiven Elementen der VR-Applikation „Peking“ verglichen wird.

Nachfolgend werden **zentrale Thesen der Wissensvermittlung** (BENKERT 2001; ISSING & KLISMA 2002; DÖRNER 2013) mitsamt kurzer Erläuterung aufgeführt und inwieweit *diese umgesetzt werden konnten*:

- **Situiertheit und Authentizität** - Dieser Punkt beschreibt, dass der Lernprozess gefördert wird, wenn zum einen der Lerninhalt in Zusammenhang mit vorhandenem Wissen gesetzt wird und zum anderen in einer bestimmten bzw. authentischen Umgebung stattfindet.
  - *In der VR-Anwendung bietet die Viermastbark ein authentisches Umfeld und die Wissensaufnahme findet mithilfe bekannter Strukturen statt. So besitzen alle Interaktions-Menüs ein gleichbleibendes Schema (durch verschiedene Tasten werden immer Zusatzinformationen generiert und eine Veränderung der virtuellen Welt hervorgerufen), wodurch der Nutzer beim Erlernen des Handlungsprinzips unterstützt wird.*

- **Multiple Kontexte** - Ein weiterer Faktor ist, dass die Wissensaufnahme durch die gelegentliche Veränderung der Perspektive sowie der Variation von Situationen und Aufgaben begünstigt wird.
  - *Durch die verschiedenen Ansätze der virtuellen Umgebung kann der Nutzer seine Perspektive beliebig variieren.*
- **Sozialer Kontext** - Außerdem wird der Wissenstransfer vorteilhaft beeinflusst, wenn ein Austausch mit weiteren Personen, unabhängig der Fachkenntnis, stattfindet.
  - *Die soziale Interaktion mit weiteren Personen ist momentan nur eingeschränkt möglich.*
- **Selbststeuerung und Kooperation** - Weiterhin steigert eine selbständige Auswahl der Lernmethode die Motivation des Lernenden. Auch eröffnet die Kooperation mit zusätzlichen Personen neue Sichtweisen.
  - *Mit der entwickelten VR-Applikation ist dem Betrachter die anzuwendende Lernstrategie freigestellt.*

Eine weitere Thematik ist die Interaktivität, für die aufgrund vieler betroffener Fachdisziplinen keine einheitliche Klassifikation existiert. ISSING & KLISMA (2002) stellen folgende Steigerung auf, welche die schrittweise **Zunahme von Interaktivität** repräsentiert und abschnittsweise der erstellten VR-Applikation gegenübergestellt wird.

- **Zugriff auf bestimmte Informationen**
  - *In der VR-Anwendung sind mehrere Interaktionsmöglichkeiten implementiert, die weiterführende Informationen anzeigen können.*
- **Ja/Nein- und Multiple-Choice-Antwortmöglichkeiten**
  - *Alle interaktiven Menüs gestatten die Auswahl einzelner Optionen, ein interaktives Element ermöglicht auch die kombinierte Selektion mehrerer Checkboxes.*
- **Aktivierung bestimmter Zusatzinformationen**
  - *Die interaktiven Komponenten sind so gestaltet, dass zusätzliche Erklärungen und Informationen angeboten werden.*
- **Komplexe Fragestellung mit intelligentem Feedback**
  - *Der Nutzer in dieser VR-Applikation kann nur zwischen bestimmten Antwortmöglichkeiten wählen und keine selbstständigen Fragestellungen definieren.*
- **Freier ungebundener Dialog mit dem System**
  - *Diese Funktion ist nicht implementiert worden.*

### 3.2 Betrachtung der Rechenleistung

Die Darstellung der digitalen Elemente in Virtual-Reality-Systemen bedarf sehr geringer Latenzen, um ein einwandfreies, schwindelfreies VR-Erlebnis zu generieren. Um eine ideale VR-Erfahrung für den Anwender zu erschaffen, wird eine „motion-to-photon time“ (Zeitspanne zwischen der Sensordetektion der Bewegung und der Reaktion auf dem Bildschirm) von weniger als 20 Millisekunden (ms), dies entspricht 50 FPS (frames per second), angestrebt (MCCAFFREY 2017). Für die nachfolgende Untersuchung wurde ein im geodätischen Labor der HafenCity Universität befindlicher Computer verwendet. Wie anhand von Tab. 1 zu erkennen ist, liegt die Recherausstattung über den empfohlenen Systemanforderungen.

Tab. 1: Minimale, empfohlene und verwendete Systemanforderungen für die genutzte HTV Vive Pro (HTC CORPORATION 2020)

Hardware	Minimale Systemanforderungen	Empfohlene Systemanforderungen	Verwendete Systemanforderungen
Prozessor	Intel® Core™ i5-4590 3,3 GHz (3,7 GHz)	Intel® Core™ i5-4590 3,3 GHz (3,7 GHz)	Intel® Core™ i7-6700K 4,0 GHz (4,2 GHz)
Grafikkarte	NVIDIA GeForce™ GTX 970	NVIDIA GeForce™ GTX 1070	NVIDIA GeForce™ GTX 1080
Arbeitsspeicher	4 GB	4 GB	16 GB
Betriebssystem	Windows 7 SP1, Windows 8.1, Windows 10	Windows 10	Windows 10

Bei einem gleichbleibenden Blick auf eine Szene mit wenigen Elementen und einem geringen Detaillierungsgrad werden in der VR-Anwendung Spitzenwerte von bis zu 54 FPS/19 ms (Abb. 8, links) erreicht. Mit zunehmenden Details nimmt die Leistung, wie bei der Ansicht des gesamten Schiffes ( $\varnothing$ : 21 FPS/48 ms - Abb. 8, Mitte), immer mehr ab. Durch die physische Simulation von Prozessen nimmt die benötigte Rechenleistung rapide zu und dadurch verzögert sich die Visualisierung, so dass bei der Betrachtung der Segel ( $\varnothing$ : 15 FPS/67 ms - Abb. 8, rechts) aus nächster Nähe der Nutzer die Latenzen deutlich wahrnehmen kann.



Abb. 8: Performance verschiedener Szenarien: Blick über die Hafenszenerie (links), Ansicht des gesamten Schiffes (Mitte) und physische Simulation eines Segels (rechts)

## 4 Fazit & Ausblick

Die Zielsetzung dieses an der HafenCity Universität Hamburg durchgeführten Projektes bestand darin, die Viermastbark „Peking“ in einem Zustand um das Jahr 1927/1928 in einer Virtual-Reality-Applikation zu visualisieren, da das Schiff zu der Zeit noch als Frachtsegler auf der Hamburg-Chile-Route verkehrte. Das Schiff stellt als ein bedeutender Teil der Hamburger Geschichte in virtuell aufbereiteter Form ein einzigartiges immersives Erlebnis dar (Abb. 9 und Abb. 10). Im Zuge dieser VR-Realisierung konnten die formulierten theoretischen Anforderungen der Wissensvermittlung durch Interaktivität untersucht und größtenteils erfolgreich umgesetzt werden. Es konnte dadurch gezeigt werden, dass VR-Systeme in der Lage sind, eine neue Form der Wissensvermittlung zu ermöglichen. Optimierungspotential besteht bei der Wissensvermittlung im Hinblick auf den Austausch mit weiteren Personen, der in dieser 1. Version der VR-Applikation nicht vorgesehen war. Jedoch existieren Technologien, die eine Kooperation und Kommunikation von mehreren Nutzern an verschiedenen Orten in VR erlauben und in anderen Projekten an der HafenCity Universität Hamburg bereits erfolgreich erprobt wurden (KERSTEN et al. 2018; TSCHIRSCHWITZ et al. 2019). In Bezug auf die Interaktivität kann geschlussfolgert

werden, dass VR-Systeme nur bis zu einem gewissen Grad interaktiv sind und vor allem der Umgang mit eigenständigen Fragestellungen noch nicht umfassend implementiert werden kann. Die Performance des VR-Systems ist stark abhängig von der Anzahl der implementierten Elemente, Interaktionen und dynamischen Komponenten. Maximal werden Framerates von 60 FPS erreicht, während in einigen Sequenzen lediglich im Durchschnitt 15 Bilder pro Sekunde angezeigt werden, vor allem Punktwolken und die physikalisch korrekt simulierten Segel des Schiffes beanspruchen große Mengen an Rechenleistung. Die Optimierung der Performance war allerdings kein vordergründiges Ziel in diesem Projekt.



Abb. 9: Überblick über die in der VR-Applikation visualisierte Viermastbark "Peking"



Abb. 10: Die visualisierte "Peking" bei gesetzten Segeln in der VR-Applikation

Des Weiteren sind noch viele Ansätze bzw. Optionen vorhanden, welche zusätzliches Entwicklungspotential besitzen und eine Erweiterung der VR-Applikation erlauben. Beispielsweise könnten zukünftig mobile Endgeräte mit der AR-Technologie (Augmented Reality) eingesetzt werden. Bei dieser Variante würde ein Protagonist in VR das Geschehen steuern, während weitere Personen zugleich die Szenerie individuell betrachten, jedoch nicht eingreifen können (z.B. anlehnend an GROOVE JONES 2017). Allein diese VR-Applikation zeigt beispielhaft, dass durch den Einsatz von Virtual Reality interaktive Wissensvermittlung mit Unterstützung von Animationen in einer neuen Form realisierbar ist und in Zukunft eine immer größere Rolle in der Mediennutzung bei Bildungseinrichtungen einnehmen kann.

## 5 Dank

Die Autoren danken Ursula Richenberger, der Direktorin des Hafenumuseums, für die allgemeine Unterstützung des Projekts. Wir danken für die Informationen, technischen Zeichnungen und das 3D-CAD-Modell der „Peking“, welche von Helmut Radebold (TECHNOLOG Services GmbH, Hamburg) zur Verfügung gestellt wurden, das 3D-Modell der Viermastbark, das von Lars Spieckermann (Detlev Löll Ingenieurbüro GmbH, Peenemünde) bereit gestellt wurde, die terrestrischen Laserscan-Daten der „Peking“, die vom Vermessungsbüro Overath & Sand, Rendsburg, geliefert wurden, und für die Unterstützung bei der photogrammetrischen Datenerfassung des Hilfslenkrades durch Dr. Ralf Wiechmann vom Museum für Hamburgische Geschichte.

## 6 Literaturverzeichnis

- BENKERT, S., 2001: Wissensvermittlung mit neuen Medien – Untersuchungen am Beispiel Niedrigenergie- und Solartechnik. Dissertation, Universität-Gesamtschule Siegen, <http://webdoc.sub.gwdg.de/ebook/dissts/Siegen/Benkert2001.pdf>.
- DEGGIM, S., KERSTEN, T., TSCHIRSCHWITZ, F. & HINRICHSSEN, N., 2017: Segeberg 1600 – Reconstructing a Historic Town for Virtual Reality Visualisation as an Immersive Experience. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, **42**(2W8), 87-94.
- DÖRNER, R., 2013: Virtual und Augmented Reality (AR/VR) Grundlagen und Methoden der Virtuellen und Augmentierten Realität. Springer-Verlag Berlin Heidelberg.
- EPIC GAMES, 2020: Unreal Engine 4 Documentation. <https://docs.unrealengine.com/en-US/Resources/Showcases/MatineeFightScene/index.html>, letzter Zugriff 10.01.2020,
- FREUNDE DER VIERMASTBARK PEKING E.V., 2020: Die Viermastbark Peking. <https://peking-freunde.de/index.php>, letzter Zugriff 10.01.2020
- GROOVE JONES, 2017: AR + VR Mixed Reality: Look Into VR Worlds With an iPad. [https://www.groovejones.com/mixed\\_reality\\_with\\_an\\_ipad\\_and\\_htc\\_vive/](https://www.groovejones.com/mixed_reality_with_an_ipad_and_htc_vive/), letzter Zugriff 10.01.2020
- HTC CORPORATION, 2020: What are the system requirements [https://www.vive.com/ca/support/vive-pro-hmd/category\\_howto/what-are-the-system-requirements.html](https://www.vive.com/ca/support/vive-pro-hmd/category_howto/what-are-the-system-requirements.html), letzter Zugriff 10.01.2020.

- ISSING, L. & KLISMA, P., 2002: Information und Lernen mit Multimedia und Internet: Lehrbuch für Studium und Praxis. Beltz Weinheim.
- JAKOBSEN, T., 2001: Advanced Character Physics. Game developers conference, (3), IO Interactive, Copenhagen, Denmark, 383-401.
- KERSTEN, T., HINRICHSEN, N., LINDSTAEDT, M., WEBER, C., SCHREYER, K. & TSCHIRSCHWITZ, F., 2013: Baugeschichtliche 3D-Dokumentation des Alt-Segeberger Bürgerhauses durch Photogrammetrie und Terrestrisches Laserscanning. Denkmäler3.de – Von low-cost bis high-tech: 3D-Dokumentation in Archäologie und Denkmalpflege, Przybilla, H.-J. et al. (Hrsg.), LWL Industriemuseum Zeche Zollern Dortmund, 30-37.
- KERSTEN, T. & LINDSTAEDT, M., 2012a: Image-Based Low-Cost Systems for Automatic 3D Recording and Modelling of Archaeological Finds and Objects. Progress in Cultural Heritage Preservation, Lecture Notes in Computer Science, Vol. **7616**, Ioannides, M. et al. (eds.), Springer, Cham, 1-10.
- KERSTEN, T. & LINDSTAEDT, M., 2012b: Virtual Architectural 3D Model of the Imperial Cathedral (Kaiserdom) of Königsutter, Germany through Terrestrial Laser Scanning. EuroMed 2012 - Int. Conference on Cultural Heritage, Ioannides, M.; Fritsch, D.; Leissner, J.; Davies, R.; Remondino, F.; Caffo, R. (Eds.), Lecture Notes in Computer Science (LNCS), Volume 7616, Springer-Verlag Berlin Heidelberg, 201-210.
- KERSTEN, T., TSCHIRSCHWITZ, F., DEGGIM, S. & LINDSTAEDT, M., 2018: Virtual Reality for Cultural Heritage Monuments – from 3D Data Recording to Immersive Visualisation. Digital Heritage - Progress in Cultural Heritage: Documentation, Preservation, and Protection, Part II, Lecture Notes in Computer Science, Vol. **11197**, Ioannides, M. et al. (eds.), Springer, Cham, 74-83.
- MCCAFFREY, M., 2017: Unreal Engine VR Cookbook: Developing Virtual Reality with UE4. Addison-Wesley Professional, 1. Auflage, 288 S.
- TSCHIRSCHWITZ, F., RICHERZHAGEN, C., PRZYBILLA, H.-J. & KERSTEN, T., 2019: Duisburg 1566: Transferring a Historic 3D City Model from Google Earth into a Virtual Reality Application. Photogrammetrie, Fernerkundung, Geoinformation (PFG), **87**(1-2), 47-56.
- ULRICH, F., 2020: Die „Peking“ wird wieder zur Grande Dame. Hamburger Abendblatt, 4. Januar 2020, <https://www.abendblatt.de/hamburg/article228054579/peking-restaurierung-schiff-frachtsegler-hafen.html>, letzter Zugriff 12.01.2020.
- WALMSLEY, A. P. & KERSTEN, T., 2019: Low-cost Development of an Interactive, Immersive Virtual Reality Experience of the Historic City Model Stade 1620. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, **42**(2/W17), 405-411.

# Semantic labelling of building types. A comparison of two approaches using Random Forest and Deep Learning

ARIANE DROIN<sup>1,2</sup>, MICHAEL WURM<sup>2</sup> & WOLFGANG SULZER<sup>1</sup>

*Abstract: In the context of sustainable planning, knowledge about building type is crucial. Yet, this information is scarce and mostly inhomogeneous. In regard to the big-data era, two approaches for building type classification are presented based on different data basis. The first approach shows semantic classification of building footprints using a set of features (simple geometric, morphological and topological features) and the machine learning algorithm Random Forest. Very high accuracies for the federal states of Germany could be achieved with Kappa Coefficients between 0.87 and 0.98. The second framework presents the possibility to conduct semantic labelling of aerial images using Fully Convolutional Neural Networks. The gained accuracy in this case is a Kappa of 0.73 for the federal state of Berlin.*

## 1 Introduction

In the general discourse of energy consumption, buildings make up for about 30-40% of the global final energy use consumption (INTERNATIONAL ENERGY AGENCY AND THE UNITED NATIONS ENVIRONMENT PROGRAMME 2018). However, the building sector is also known to have huge potentials in regard to energy savings and hence to officiate as key holder to meet the defined energy saving goals (STEEMERS & YUN 2009). Consequently, knowledge about building types is crucial in the context of heat demand calculations and predictions. In spite of the importance of this data, the availability of this information is scarce, not up-to-date and heterogeneous (HECHT et al. 2015).

Derivation and classification of building types based on remote sensing and Geographic Information System (GIS) data has been researched and conducted by several studies (e.g. MEINEL et al. 2009; BELGIU et al. 2014; WURM et al. 2016). The researched approaches however vary broadly depending on data basis, on local conditions, on the desired output and focus mostly on city applications.

The development and evolution of artificial intelligence makes it possible to emerge further into the field of big data and to solve tasks which were not feasible up to now. Hence, the presented study shows two approaches for semantic labelling of building types based on the assumption of the availability of different data basis. The first approach uses the machine learning algorithm Random Forest (RF) to derivate building types on two semantic stages (see Figure 1) based on Level of Detail 1 (LoD1) and census data. The second approach relies solely on remote sensing data and hence on the classification of different building types based on spectral data when

---

<sup>1</sup> Karl-Franzens Universität Graz, Department of Geography and Regional Sciences, Heinrichstraße 36, A-8010 Graz, Austria, E-Mail: wolfgang.sulzer@uni-graz.at

<sup>2</sup> German Aerospace Center (DLR), German Remote Sensing Data Center (DFD), Department of Geo-Risks and Civil Security, Münchener Straße 20, D-82234 Weßling, Germany  
E-Mail: [Ariane.Droin, Michael.Wurm]@dlr.de

building footprint data is not available. Up to now this task was not feasible, based on the broad intra-spectral variability of buildings, but with the emergence and evolution of Neural Networks, which can detect non-linear and complex relationships in the data, new fields arise (ZHANG et al. 2016; ZHU et al. 2017). Consequently, the second approach focuses on the semantic segmentation of building types based on aerial images.

The nomenclature for the semantic stages (see Fig. 1) is based on the nomenclature of the Institute of Housing and Environment (IWU), as these related types are used as basis for energy consumption modelling.

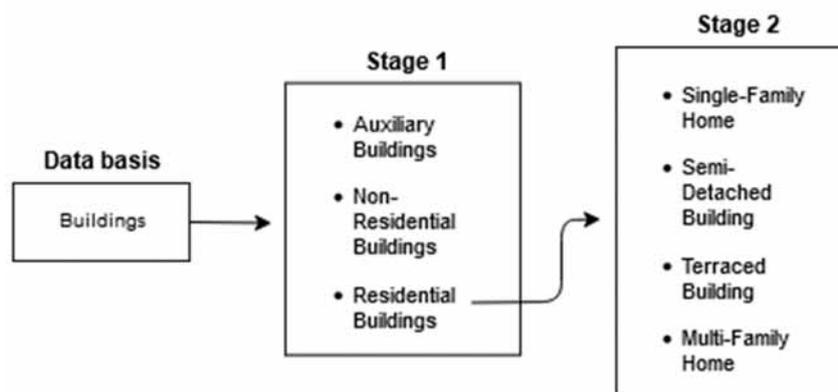


Fig. 1: Two-stage approach for semantic labelling of building types

## 2 Methods and data

### 2.1 Semantic labelling of building types using Random Forest

The building footprint dataset (LoD1) is acquired from the Federal Agency of Cartography and Geodesy (BKG) of Germany and contains ~51 million buildings. Additionally, the census data from 2011, provided in 100 m grid cells, is used for the generation of training and reference datasets as building type information is contained within the census (see Fig. 2).

The RF algorithm is based on the majority vote of a multitude of decision trees which are built using bootstrap aggregating (BREIMAN 2001). The efficiency of RF, and hence making it a state-of-the-art machine learning algorithm, lies in the fact that it takes random subsamples of data and features for building each single tree (bootstrap aggregating) and thus making it prone to overfitting. The remaining data that was not used in the building process of the trees is used to estimate the performance of each single tree (Out-Of-Bag (OOB) estimate) as well as the ensemble of trees. Furthermore, the importance of each feature can be estimated, which is useful information for restricting the calculation performance (RODRIGUEZ-GALLIANO 2012; LIAW & WIENER 2002).

For the classification task, different sets of features are chosen (see Tab. 1). It is assumed that the different building types have different morphological and topological properties based on their type (see Tab. 2).

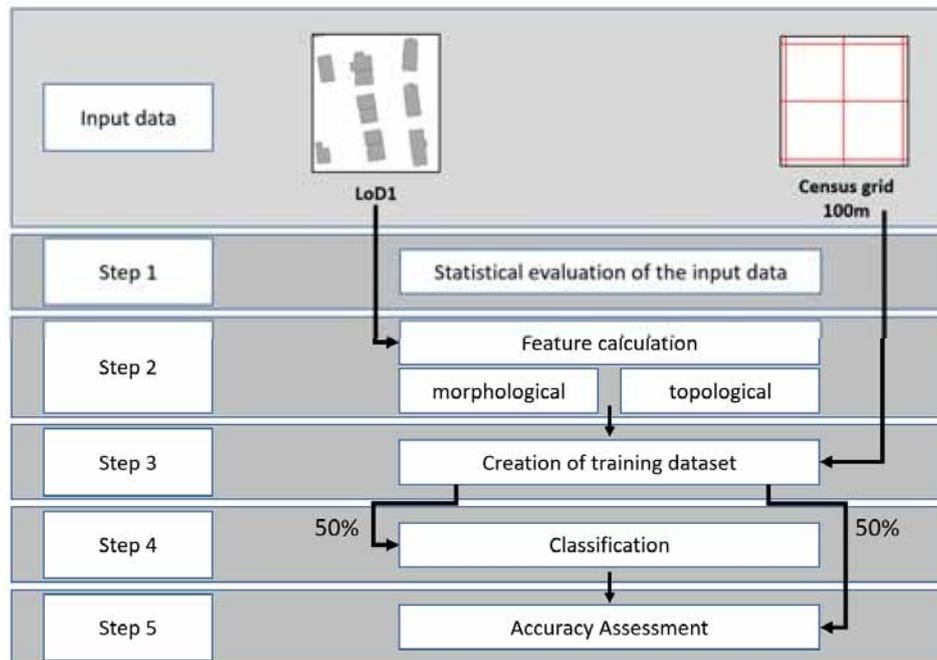


Fig. 2: General layout of the developed methodology for semantic labelling of LoD1 data

The feature sets are calculated for every single building and used for the classification task. The information of the building types of the first semantic stage is already included officially for almost every federal state of Germany (apart from Bremen, Saxony and Thuringia) and is hence used for prediction of the latter three. Using the defined features and RF, the information of the first semantic stage is predicted for those three states.

Tab. 1: Feature set used for Random Forest classification of building types

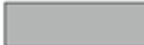
<b>Simple Geometric Features:</b>	Perimeter [m], Area [m <sup>2</sup> ], Height [m], Proportion between Height and Area, Volume [m <sup>3</sup> ]
<b>Morphological Features</b>	Detour, Range, Exchange, Cohesion, Proximity, Spin, Shape Index, Fractal Dimension
<b>Topological Features</b>	Consecutive neighbors, Dissolved area of neighboring buildings [m <sup>2</sup> ], Dissolved perimeter [m], Relative area [%] of the building area compared to the dissolved area

For the prediction of the second semantic stage (see Fig. 1) training data is generated using the census grid cells. Buildings which are located in cells with only one building type are selected. However, as it is assumed that certain errors can occur, additional thresholds are defined, to ensure that only the desired building type is included (e.g. Single-Family Homes have no consecutive neighbors or Semi-Detached buildings must have exactly one neighbor and at least a share of 25% on the dissolved area).

For the classification procedure the default values of RF classifiers are used (500 trees and  $\sqrt{n}$  features) and to reduce calculation time each federal state is classified separately. A random subsample of 50% is taken from the training data for the classification process while the other 50% are used for Accuracy Assessment. Furthermore, several set-ups are constructed to assess

the quality, applicability and conformity of the developed approach. The first set-up evaluated the performance of census data as training dataset. Therefore, a manually labelled and randomly selected training sample of 3,000 buildings footprints for the federal state of Berlin is generated. The second set-up assesses the influence and importance of the different feature sets on the classification result.

Tab. 2: Examples of different morphological properties for different residential building types (Image Source: Google Earth 2018)

	Single-Family Home	Semi-Detached Building	Terraced Building	Multi-Family Home
Schema	  	  	  	     
Perimeter Index	0.85	0.87	0.84	0.69
Detour Index	0.89	0.87	0.84	0.69
Range Index	0.80	0.76	0.72	0.53

## 2.2 Semantic segmentation of building types using Deep Learning

Semantic segmentation of aerial images is the task of assigning each pixel a semantic meaning. Deep learning architectures have proven to be far more successful in this sense than conventional machine learning algorithms (LONG et al. 2015). The advantage of Deep Learning is that it can automatically learn abstract and discriminative features from the input data. Hence, over-specification or incompleteness of features is hindered (ZHU et al. 2017).

LONG et al. (2015) brought up Fully Convolutional Networks (FCN) where no fully connected layers are needed for semantic segmentation of images. For semantic segmentation of building types based on aerial images the FCN-vgg19 from the Visual Group of Oxford University (SIMONYAN & ZISSERMAN 2014) is used and for the architecture of the network, we refer to WURM et al. (2019) who deployed the network for semantic segmentation of informal settlements. In the proposed approach, we use orthophotos with a resolution of 40 cm in the visible spectrum.

For semantic segmentation, rasterization of the LoD1 data (using the first and the second semantic stage respectively) is carried out for the generation of training data. Additionally, training data using only the building footprints with no further semantic differentiation is also generated. This is done to be able to assess if the Random-Forest approach can also be applied on footprint data derived using Deep Learning algorithms. The general layout of the semantic segmentation approach using Deep Learning is shown Fig. 3.

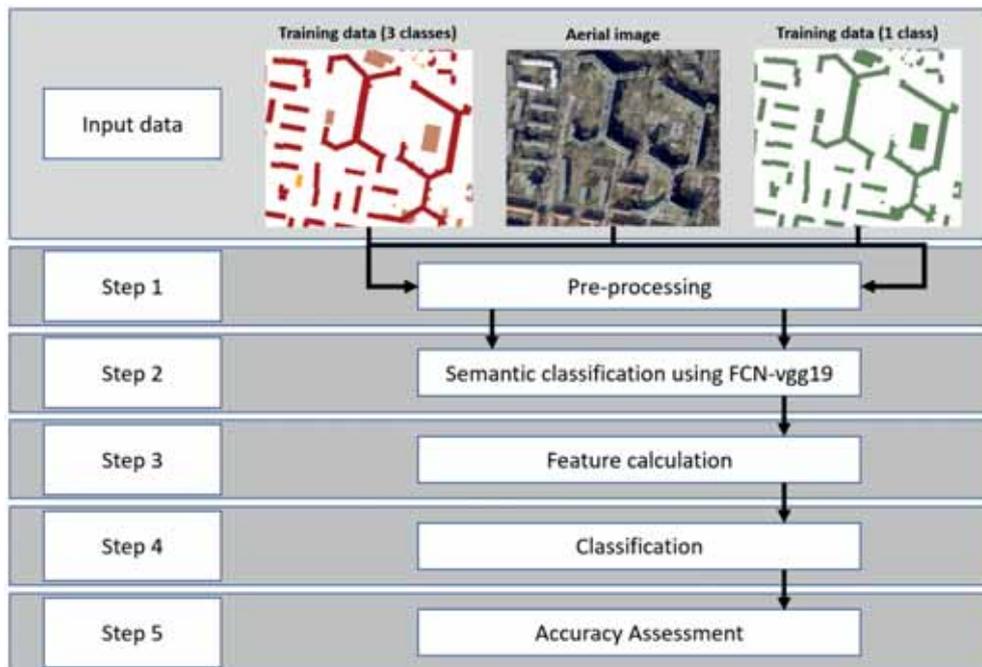


Fig. 3: General layout of the developed methodology for semantic segmentation of aerial images

The FCN algorithm works with image tiles with a size of  $224 \times 224$  pixels, resulting in a training data set with 55,843 tiles without overlap. For assessing the accuracy, a 4-fold cross validation is carried out using the columns of the tiling process.

While the results of the semantic segmentation process on the first stage only need to be post-processed due to tiling-effects, the semantic segmented building footprints need to be further processed to gain a higher semantic meaning. Therefore, the morphological features from the RF-approach are calculated for every derived footprint. Based on the fact, that the resulting footprints are always one polygon and that the housing units cannot be separated, no topological features can be calculated. Furthermore, due to missing elevation data no features including height information could be included in the feature set. The accuracy of the results from semantic classification on the first stage on the Deep Learning derived footprints is assessed in two ways. On the one hand, 4,700 building footprints are annotated manually for the generation of a reference dataset. On the other hand, the LoD1 data with the included semantic information which intersect the derived building footprints are dissolved into one building and assigned the class with the biggest share on the dissolved area. Hence, 81,000 reference buildings for Berlin could be generated (see Fig. 4).

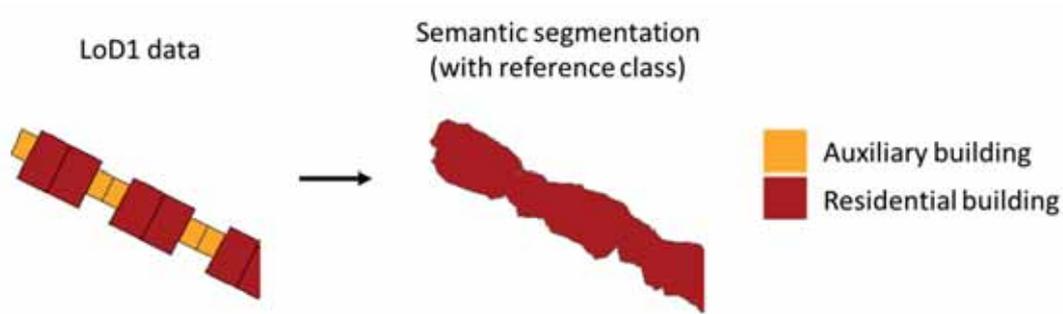


Fig. 4: Generation of reference data using the class with the largest share on the intersecting building

### 3 Results

For the RF approach, Kappa accuracies of 0.86 (Bremen) and 0.91 (Saxony and Thuringia) could be reached for the first semantic stage. The other federal states already had the first semantic stage information included. For the second semantic stage Overall Accuracy (OA) for every singly federal state is above 95% and Kappa is around 0.9 for the states of Baden-Württemberg and Saxony and above 0.9 for the other federal states (see Fig. 5).

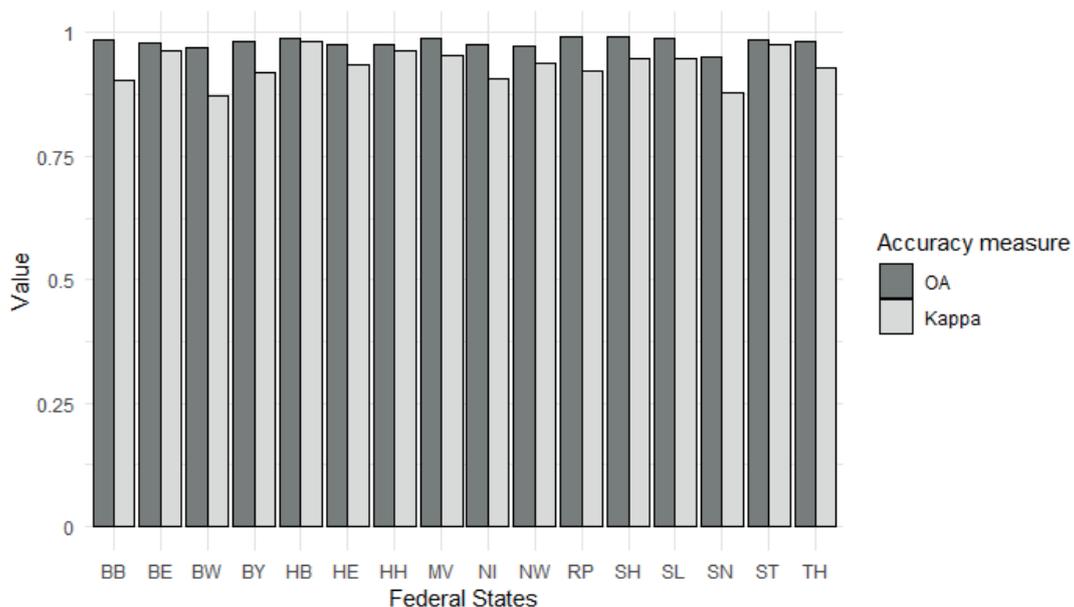


Fig. 5: Accuracy results for each federal state

Regarding the feature importance (see Fig. 6) differences between the different sets can be deduced. Simple geometric features, especially those containing height information, and topological features score the highest importance between the federal states. The indexed morphological features show big variations in their importance. The importance of the feature “Relative area” has the lowest variation and has therefore approximately the same importance in the classification process of each federal state.

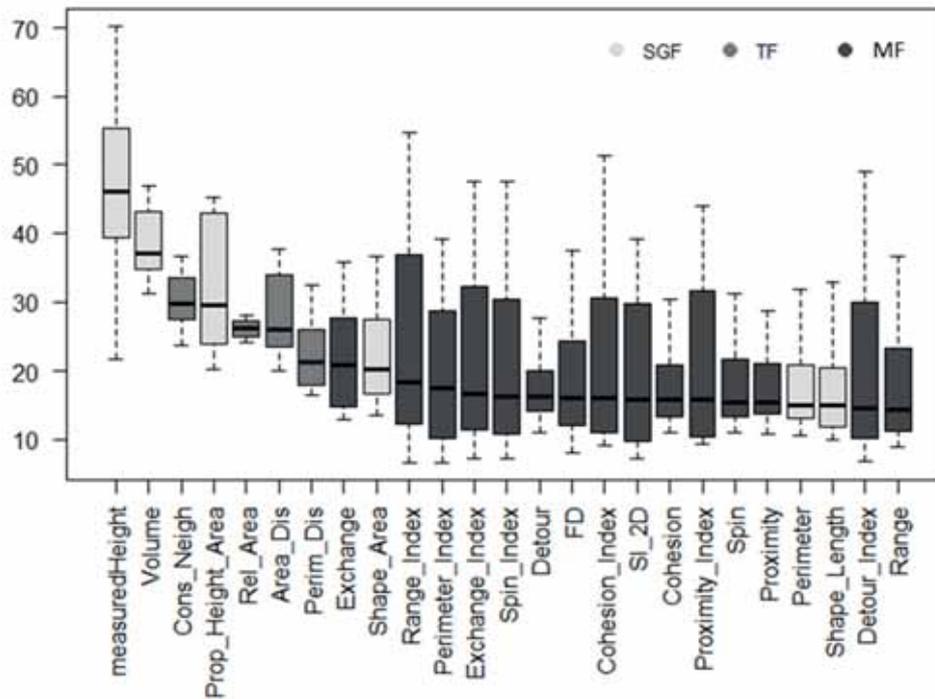


Fig. 6: Feature importance for the RF approach for residential building type classification (SGF = Simple geometric features, TF = Topological features and MF = morphological features)

The accuracy results for the evaluation of the usability of census data using a manually labelled reference data set for comparison yields quite resembling results, with a slightly lower accuracy for the manual data set (see Fig. 7).

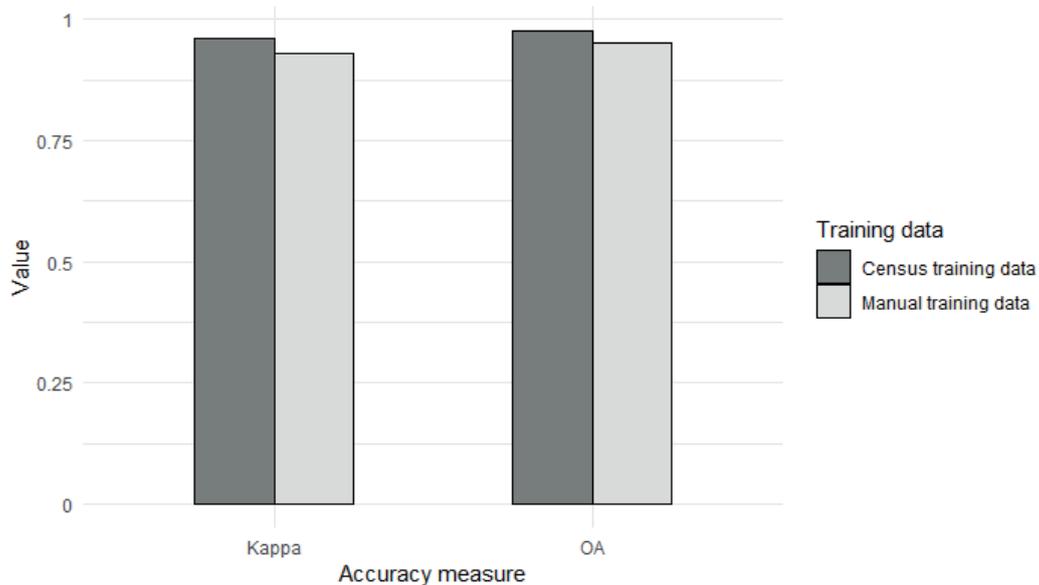


Fig. 7: Comparison of accuracy measures for reference data using different sets of training data

The impact on the accuracy when different sets of features are used (first set is with all features, second set is only with simple geometric and topological features combined and third set is only morphological features) show clear tendencies. The accuracy results show that there is a clear decrease in accuracy for every federal state when only morphological features are used, and only a slight decrease in accuracy when no morphological features are included (see Fig. 8). Furthermore, it can be stated, that the lowering of accuracy is considerably less in urban federal states than in the rural ones.

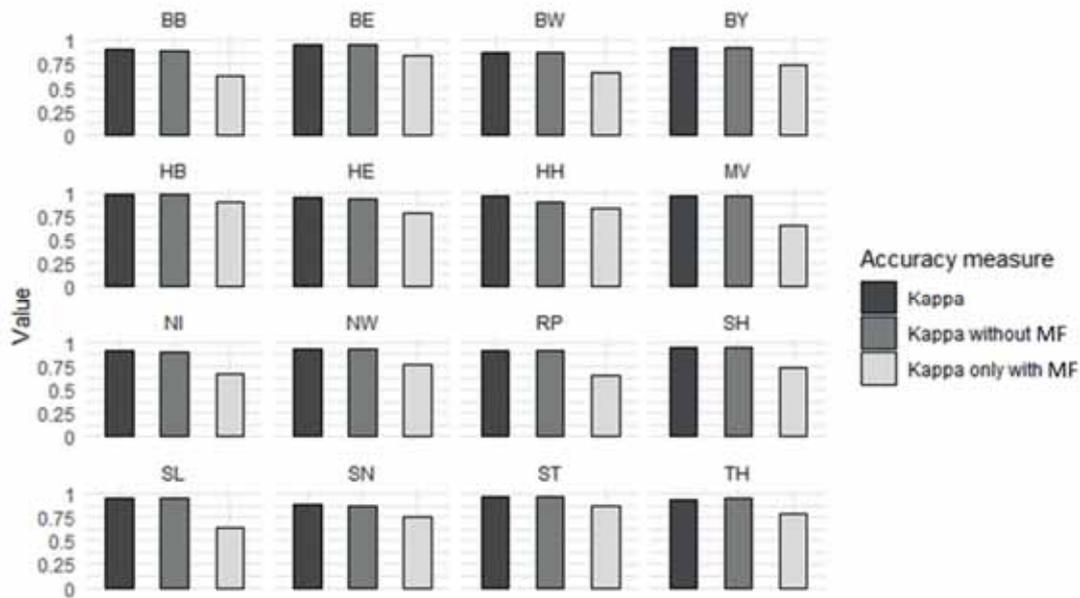


Fig. 8: Accuracy result for the different federal states using different sets of features (MF = morphological features)

The approach using Fully Convolutional networks results in an OA of 93% and a Kappa of 0.73 for the first semantic stage. If no semantic differentiation is conducted but only the detection of building footprint data is regarded a Kappa of 0.8 is achieved (see Fig. 9).

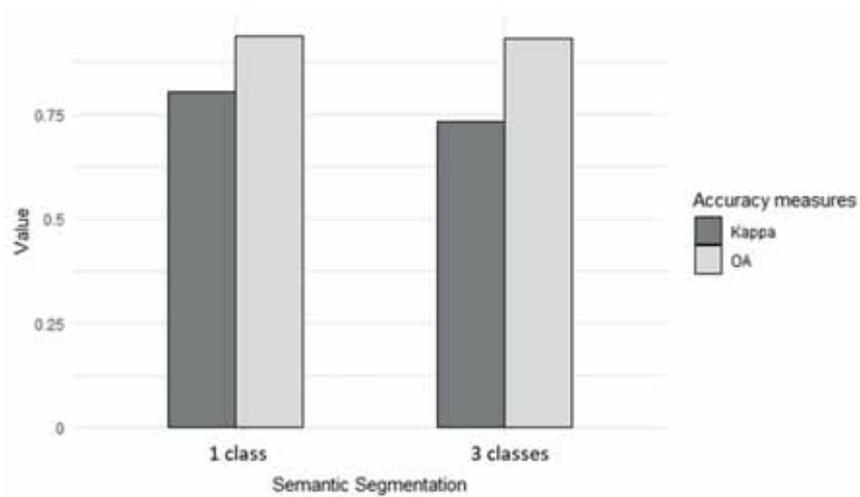


Fig. 9: Accuracy results for semantic segmentation of building footprints and the first semantic stage

Applying the Random Forest approach on the derived footprints with no semantic specification using the LoD1 data yielded low accuracies with a Kappa value of 0.32 and using the manually generated training and reference data yielded results with a Kappa value of 0.46 (see Fig. 10).

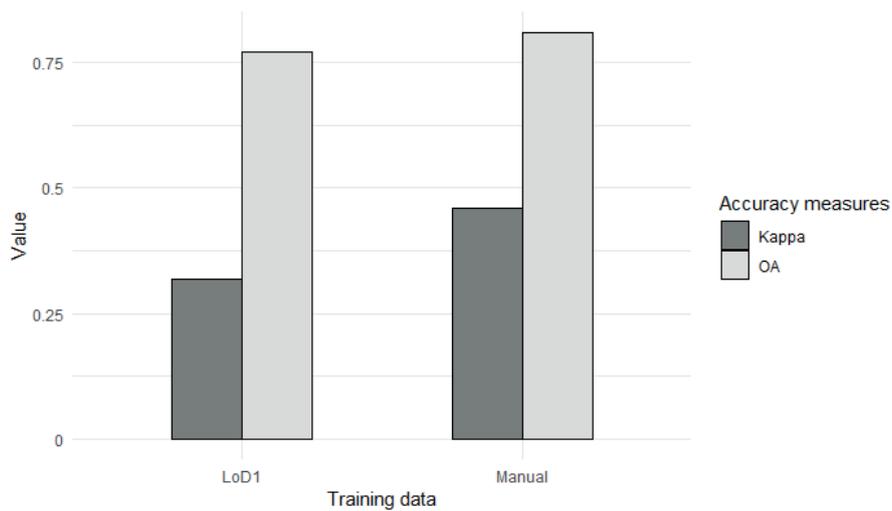


Fig. 10: Accuracy measures of Random Forest classification based on Deep Learning derived footprints using Random Forest

A visualisation of the results of the first and second semantic stage based on the different data basis can be taken from Fig. 11.

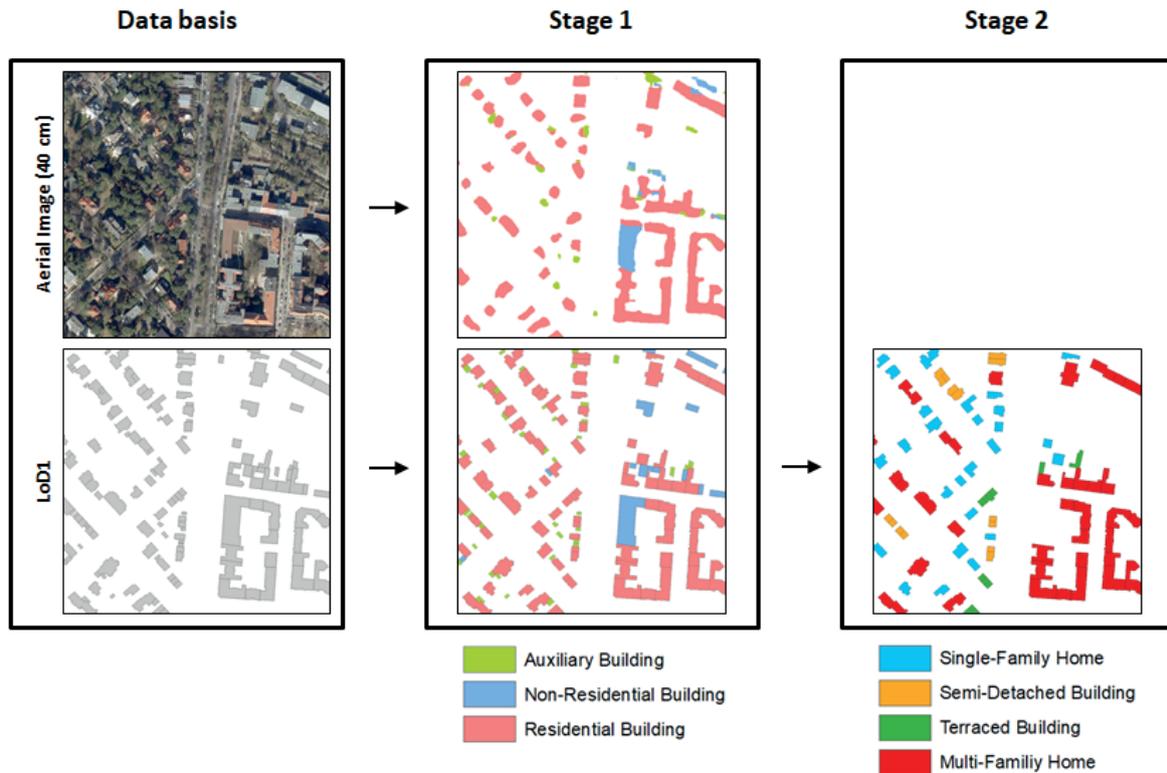


Fig. 11: Exemplary results for the two approaches

## 4 Discussion and Outlook

All accuracies of the RF are very high and thus, show that this approach can be applied for the derivation of building type information. Using features with topological and height information is important in this regard, as they are crucial in the process of distinguishing different building types. This is not only shown by the feature importance but also by the results of the set-up where different feature sets have been used. Even if morphological features do not seem to be as important as the ones mentioned before and no high accuracy can be gained when only morphological features are used, a stabilization of the classification result can be concluded (Kappa is lowered by  $\sim 0.2$  without them). Hence, it is suggested to use these features as well. The higher Kappa value for the urban federal states, when only morphological features are used, are based on the fact that the forms of different building types in cities are more homogeneous than in the bigger rural states. The higher variation of importance for the indexed morphological features underlines this assumption. However, the exact reasons and implications for that need to be further researched.

When the results of the accuracy using census reference data is compared to the results of the manually generated reference data it can be deduced that the census data is indeed a viable source for training and reference data. The slight lowering in accuracy for the manual training data can be concluded on the one hand on the highly imbalanced amount of class samples, based on the random selection, and on the lower amounts of reference data.

The results of the first semantic stage for the Neural Network approach are promising, regarding the fact that building types are only differentiated based on spectral data. However, the variation of roof material and hence the broad spectral variability for the further differentiation of residential building types using Neural Networks resulted in comparatively inaccurate results. Further research is to be conducted in this context, especially in regard to the inclusion of height information for example. The lower classification accuracy using only the derived footprints for the RF approach from the Deep Learning approach can be traced back to the fact that differentiation based only on morphological aspects is quite difficult. The inclusion of height information in this case could increase the accuracy considerably (cf. WURM et al. 2016). Hence, this aspect needs to be further researched, especially since in some cases spectral data is cheaper to acquire than elevation data, making it a valuable data source for semantic labelling of building footprints for areas where no official footprint data is available.

## 5 References

- BELGIU, M., TOMLJENOVIC, I., LAMPOLTSHAMMER, T.J., BLASCHKE, T. & HÖFLE, B., 2014: Ontology-based Classification of Building Types Detected from Airborne Laser Scanning Data. *Remote Sensing* **6**, 1347-1366.
- BREIMAN, L., 2001: Random Forests. *Machine Learning* **45**(1), 5-32.
- HECHT, R., MEINEL, G. & BUCHROITHNER, M., 2015: Automatic Identification of Building Types Based on Topographic Databases – a Comparison of Different Data Sources. *International Journal of Cartography* **1**(1), 18-31.
- INTERNATIONAL ENERGY AGENCY AND THE UNITED NATIONS ENVIRONMENT PROGRAMME, 2018: 2018 Global Status Report: Towards a Zero-Emission, Efficient and Resilient Buildings and Construction Sector.
- LIAW, A. & WIENER, M., 2002: Classification and Regression by randomForest. *R News* **2**(3), 18-22.
- LONG, J., SHELHAMER, E. & DARRELL, T., 2015: Fully Convolutional Networks for Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431-3440.
- MEINEL, G., HECHT, R. & HEROLD, H., 2009: Analyzing Building Stock Using Topographic Maps and GIS. *Building Research and Information* **37**(5-6), 468-482.
- RODRIGUEZ-GALIANO, V.F., GHIMIRE, B., ROGAN, J., CHICA-OLMO, M. & RIGOL-SANCHEZ, J.P., 2012: An Assessment of the Effectiveness of a Random Forest Classifier for Land-Cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing* **67**(1), 93-104.
- SIMONYAN, K. & ZISSERMAN, A., 2014: Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint arXiv:1409.1556*.
- STEEMERS, K. & YUN, G.Y., 2009: Household Energy Consumption: A Study of the Role Occupants. *Building Research & Information* **37**(5-6), 625-637.
- WURM, M., SCHMITT, A. & TAUBENBÖCK, H., 2016: Building Types' Classification Using Shape-Based Features and Linear Discriminant Functions. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **9**(5), 1901-1912.

- WURM, M., STARK, T., ZHU, X.X., WEIGAND, M. & TAUBENBÖCK, H., 2019: Semantic Segmentation of Slums in Satellite Images Using Transfer Learning on Fully Convolutional Networks. *ISPRS Journal of Photogrammetry and Remote Sensing* **150**, 59-69.
- ZHANG, L., ZHANG, L. & DU, B., 2016: Deep Learning for Remote Sensing Data. A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine* **4**(2), 22-40.
- ZHU, X.X., TUIA, D., MOU, L., XIA, G.S., ZHANG, L., XU, F. & FRAUNDORFER, F., 2017: Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine* **5**(4), 8-36.

# AUTORENVERZEICHNIS

## A

ADAM, NICO .....	87
ADELWÖHRER, MANUEL .....	184
ANDERS, KATHARINA.....	312, 373
ASTOR, THOMAS.....	172

## B

BÄUMKER, MANFRED.....	47
BAUER-MARSCHALLINGER, BERNHARD .....	505
BLASER, STEFAN .....	254
BOLZ, TOBIAS .....	457
BONERATH, ANNIKA .....	488
BRAATZ-NEBEL, JULIAN-HELGE.....	462
BRECHTKEN, RAINER .....	346
BRUNN, ANSGAR.....	411
BUCHER, TILMAN .....	13
BUDDE, LINA EMILIE .....	280
BUTTINGER-KREUZHUBER, ANDREAS .....	242

## C

CERMAK, JAN .....	400
CRAMER, MICHAEL.....	33
CROCETTI, LAURA.....	469

## D

DINKEL, AGNES .....	381
DÖLLNER, JÜRGEN.....	219, 437
DORIGO, WOUTER .....	469
DROIN, ARIANE .....	528
DUBOVYK, OLENA .....	1

## E

EGGERT, DANIEL .....	317
EHMER, FELIX.....	422
EMMERT, ADRIAN.....	381
ESCHELBACH, CORNELIA .....	346

## F

FRICKER, LEA .....	254
--------------------	-----

## G

GERIGHAUSEN, HEIKE .....	462
GESSNER, MATTHIAS .....	13
GREIWE, ANSGAR .....	346
GRENZDÖRFFER, GÖRRES.....	230
GÜLCH, EBERHARD .....	358

<b>H</b>	
HAALA, NORBERT .....	205
HAAS, RÜDIGER.....	346
HAMMER, HORST .....	73
HAUNERT, JAN-HENRIK .....	488
HEIN, DANIEL.....	13
HEß, KATRIN.....	400
HINZ, STEFAN .....	457
HODAM, HENRYK .....	126
HOEGNER, LUDWIG .....	381
HÖFLE, BERNHARD .....	290, 312, 373
HOLLAUS, MARKUS .....	242
HRUBY, FLORIAN .....	114
<b>I</b>	
IWASZCZUK, DOROTA .....	270
<b>J</b>	
JACOBSEN, KARSTEN .....	22
JAVANMARD-GH, ARASH.....	270
JÜRGENS, CARSTEN.....	126
<b>K</b>	
KERSTEN, THOMAS .....	99, 515
KLEIN, REINHARD.....	391
KÖLLE, MICHAEL.....	156, 299, 478
KOMMA, JÜRGEN .....	242
KÖNIG, SIMON .....	1
KRAFT, THOMAS .....	13
KRESSE, WOLFGANG .....	197
KRUMPEN, STEFAN .....	391
KRZYSZEK, PETER .....	497
<b>L</b>	
LAUPHEIMER, DOMINIK .....	205
LAVRENTIEV, VALERY .....	317
LILIENTHAL, HOLGER.....	462
LINDENBERGH, RODERIK .....	312
LINDNER, CLAUDIA .....	126
LOREI, HEINRICH.....	290
LÖSLER, MICHAEL.....	346
LUDWIG, ROLF RÜDIGER .....	22
<b>M</b>	
MAAS, HANS-GERD.....	143
MADER, DAVID.....	143
MANDLBURGER, GOTTFRIED .....	62, 156, 242, 321, 335
MARA, HUBERT.....	312
MAYER, MICHAEL.....	400
MEIßNER, HENRY .....	13, 33
MEYER, JONAS .....	254

<b>N</b>	
NAVACCHI, CLAUDIO .....	505
NEBIKER, STEPHAN .....	254
NEELMEIJER, JULIA .....	317
NIEDERMANN, BENJAMIN .....	488
NÜBEL, HANNES .....	156
<b>O</b>	
OBROCK, LARS SÖREN.....	358
ORTWEIN, ANNETTE .....	126
<b>P</b>	
PFEIFER, NORBERT .....	242
PFENNIGBAUER, MARTIN .....	321
PLESS, SEBASTIAN.....	13
PRZYBILLA, HEINZ-JÜRGEN.....	47
<b>Q</b>	
<b>R</b>	
RAFFL, LUKAS .....	381
RESSL, RAINER.....	114
REULKE, RALF .....	33, 87
RICHTER, KATJA .....	143
RICHTER, RICO .....	219, 437
RIENOW, ANDREAS .....	126
<b>S</b>	
SCHAUMBERGER, ANDREAS.....	184
SCHERER, DANIEL.....	497
SCHMITZ, SYLVIA .....	73, 457
SCHMOHL, STEFAN .....	280, 299
SCHNEIDER, PHILIPP .....	448
SCHULTZ, JOHANNES .....	126
SCHWARZ, ROLAND.....	321
SCHWATKE, CHRISTIAN .....	497
SELENSKI, JÜRGEN .....	411
SELG, FABIAN .....	126
SHAMS EDDIN, MOHAMAD HAKAM .....	205
SIPS, MIKE .....	317
SÖRGEL, UWE .....	156, 280, 299, 448, 457
SOSZYŃSKA, AGNIESZKA .....	13
SPRECKELS, VOLKER .....	87
STAAR, KILIAN .....	126
STILLA, UWE.....	381
STOJANOVIC, VLADETA.....	437
STOTKO, CHRISTOPH .....	391
SULZER, WOLFGANG.....	528

<b>T</b>	
THIELE, ANTJE .....	73, 457
THONFELD, FRANK .....	1
TOTH, CHARLES.....	270
TRAPP, MATTHIAS.....	437
TRAU, DOMINIK .....	515
TSCHIRSCHWITZ, FELIX .....	515
<b>U</b>	
<b>V</b>	
VOS, SANDER E.....	312
<b>W</b>	
WACHENDORF, MICHAEL .....	172
WAGNER, WOLFGANG .....	505
WALMSLEY, ALEXANDER P. ....	99
WALTER, VOLKER .....	299, 478
WASER, JÜRGEN.....	242
WEBER, DANIEL .....	254
WEIDNER, UWE .....	73, 422
WEINMANN, MARTIN.....	73, 422
WEINMANN, MICHAEL .....	391
WEIß, ROBERT.....	62
WESTERHOLT, RENE .....	290
WESTFELD, PATRICK.....	143
WICH, VALENTIN .....	411
WIJESINGHA, JAYAN .....	172
WIMMER, MICHAEL.....	242
WINIWARTER, LUKAS.....	312, 373
WOLF, JOHANNES .....	219
WURM, MICHAEL.....	528
<b>X</b>	
<b>Y</b>	
YIN, YIFEI .....	478
<b>Z</b>	
ZHANG, SHUHANG.....	33

# PROGRAMMKOMITEE

Abmayr, Thomas

Bareth, Georg

Brunn, Ansgar

Cramer, Michael

Edler, Dennis

Grenzdörffer, Görres

Gülch, Eberhard

Haala, Norbert

Höfle, Bernhard

Jacobsen, Karsten

Jung, András

Kaufmann, Viktor

Kersten, Thomas P.

Kresse, Wolfgang

Krzystek, Peter

Mandlbürger, Gottfried

Motagh, Mahdi

Petzold, Bettina

Reulke, Ralf

Sörgel, Uwe

Stilla, Uwe M.

Weinmann, Martin

Wolff, Kirsten

