



Comparison of SVM and Boosted Regression Trees for the Delineation of Lacustrine Sediments using Multispectral ASTER Data and Topographic Indices in the Lake Manyara Basin

FELIX BACHOFER, GERALDINE QUÉNÉHERVÉ, MICHAEL MÄRKER & VOLKER HOCHSCHILD, Tübingen

Keywords: SVM, boosted regression trees, ASTER, multispectral, topographic indices

Summary: The lower member of the so called Manyara Beds is a distinct lacustrine sedimentary layer which indicates, with an elevation of more than 140 m above today's lake level, a high stand of the paleolake Manyara in the Monduli District in northern Tanzania. The Manyara Beds are rich in Pleistocene vertebrate fossils. In this study we focus on the delineation of this specific stratigraphic layer in order to yield new insights into paleontological settings, landscape evolution and to plan paleontological fieldwork. We compare the performance of a support vector classifier with a linear as well as a Gaussian kernel, with boosted regression tree approaches to identify the lithostratigraphic layers of the Manyara Beds. For the identification of the lacustrine sediments, multispectral information of ASTER satellite imagery and topographic indices derived from a digital elevation model were utilized as input feature sets. Acceptable classification accuracies were obtained with all methods. Thus, the Manyara Beds can be delineated and new sites with paleolake sediments were detected. The highest overall accuracy with 92% was provided by the support vector machine approach with a linear kernel for a binary classification problem. For a multi-class classification problem with three target classes the support vector classifier achieved 80% accuracy with a linear, as well as a Gaussian kernel.

Zusammenfassung: Vergleich von SVM und Boosted Regression Trees zur Abgrenzung von lakustrinen Sedimenten anhand von multispektralen ASTER Daten und topographischen Parametern im Einzugsgebiet des Manyara Sees. Die aus vornehmlich lakustrinen Sedimenten bestehende ältere Gruppe der stratigraphischen Einheit der Manyara Beds beschreibt mit einer Höhe von mehr als 140 m über dem heutigen Seespiegel einen Hochstand des Paläosees Manyara im Monduli Distrikt im nördlichen Tansania. Die Manyara Beds sind reich an pleistozänen Wirbeltierfossilien. Die vorliegende Arbeit beschäftigt sich mit der räumlichen Abgrenzung dieser stratigraphischen Einheit um mehr über die paläontologischen Ablagerungsbedingungen und die Landschaftsgeschichte zu erfahren, sowie die Planung von paläontologischen Geländearbeiten zu unterstützen. Wir vergleichen anhand der lithostratigraphischen Einheit der Manyara Beds die Leistungsfähigkeit eines Support Vector (Stützvektoren) Klassifizierungsansatzes, mit einem linearen und einem Gaußschen Kernel, und mit Klassifizierungsbäumen (Boosted Regression Trees). Um die lakustrinen Sedimente zu unterscheiden, wurden multispektrale Informationen einer ASTER Satellitenaufnahme und topographische Parameter von einem digitalen Höhenmodell als Eingangsvariablen genutzt. Mit allen Klassifizierungsmethoden wurden zufriedenstellende Genauigkeiten erzielt. Somit konnte das Auftreten der Manyara Beds räumlich abgegrenzt und bisher nicht dokumentierte Flächen mit lakustrinen Sedimenten erfasst werden. Die höchste Klassifizierungsgenauigkeit von 92% wurde von der Support Vector Machine Klassifizierung mit einem linearen Kernel für eine binäre Klassifizierung erreicht. Für eine Aufgabenstellung mit Support Vector Machines für drei Zielklassen wurde eine Genauigkeit von 80% sowohl mit einem linearen, als auch mit einem Gaußschen Kernel erreicht.

1 Introduction

Lacustrine sediments and paleo-shorelines of different Quaternary lake-level high stands can be observed in the north, south and east of the Lake Manyara basin of northern Tanzania. The study area is located in the Gregory Rift in Central North Tanzania. The basin is of paleontological and archeological interest documented by several investigations in recent years (e.g. SCHLÜTER et al. 1992, KAISER et al. 2010, PRENDERGAST et al. 2013). One of the richest stratigraphic units in vertebrate fossils and artifacts in the region are the Manyara Beds, which indicate a high level of the paleolake Manyara at more than 140 m above today's lake level. The identification of the Manyara Beds contributes to the understanding of landscape evolution and the spatial distribution of potential paleontological sites. Thus, the study also serves for the planning of future fieldwork in the study area.

Remote sensing images are used in different studies to derive information on the extent of paleolakes and other paleo-landscape forms. EL-SHEIKH et al. (2011) and ELMAHDY (2012) used remote sensing, GIS and geophysical methods to delineate a paleolake in northern Darfur. GHONEIM et al. (2012) used an integrated approach with optical and microwave data to map a paleo-drainage system. The use of remote sensing in combination with topographic analysis for the delineation of paleolakes has been successfully applied by GABER et al. (2009) on the Sinai Peninsula.

MOUNTRAKIS et al. (2011) review the application of support vector machines (SVM) in the classification of remotely sensed images. SVM and boosted regression tree analysis (BRT) were more and more used in the last decade and yield high accuracies (FOODY & AJAY 2004, ESCH et al. 2009, WANG et al. 2011, GÓMEZ et al. 2012, GESSNER et al. 2013). SVM and BRT analyses are capable of handling multiple input features, outliers, non linear tasks and redundant data (FOODY & AJAY 2004, ELITH et al. 2008). HAHN & GLOAGUEN (2008) used SVM to classify soil types and soil texture from ASTER multispectral data and topographic parameters in the Erzgebirge in Germany. BRT methods have been assessed with remote sensing data for land use issues (PAL &

MATHER 2003, CHAN & PAELINCKX 2008). The mapping of lithological units and the distribution of soil with multispectral data and terrain attributes as well as classification methods were reviewed by MULDER et al. (2011). In this study we compare the accuracy of SVM and BRT classifier in identifying the Manyara Beds in a small scaled, heterogeneous environment. PAL & FOODY (2010) showed that an increase of input features may lead to a decline of classification accuracy. Therefore, we apply a feature selection to choose a subset of different ASTER spectral bands, multispectral indices and topographical indices. From this selected set of features, we expect also improved model interpretability, as well as an enhanced generalisation of the resulting models.

2 Regional setting and data preparation

2.1 Study Area

The Lake Manyara catchment in northern Tanzania is an endorheic basin and part of the eastern branch of the East African Rift System (Fig. 1). Today Lake Manyara is a shallow soda lake (954 m a.s.l.) with a maximum depth of 1.18 m (DEUS et al. 2013). The basin is an asymmetrically shaped half graben, with a 200 m to 600 m high escarpment along the western shoulder. The eastern shoulder of the Rift is lower in elevation and consists of tectonic blocks that are dipping towards the west. The North-eastern parts of the catchment area are dominated by the volcano Essimngor. Paleo-shorelines can be found especially on the Eastern part of the rift tracing different paleolake levels up to 80 m above today's lake level. The latter forms also the lowest possible outlet into the Engaruka and Lake Natron basin (KELLER et al. 1975, BACHOFER et al. 2014). The maximum age of the paleo-shorelines was established with radiocarbon dating (Th/U series) to 140,000 a BP (CASANOVA & HILLAIRE-MARCEL 1992). Besides the springs at the base of the escarpment, Lake Manyara is mainly fed with seasonal drainages of the Tarangire and Makuyuni rivers. Today a bimodal precipitation pattern with an average annual rainfall of about 700 mm can be observed for

the study area. The resulting semiarid vegetation cover is sparse and dominated by bushed grassland (BACHOFER et al. 2014).

2.2 The Manyara Beds

The lacustrine and fluvial stratigraphic units, known as the Manyara Beds in the east of Lake Manyara, reach up to approximately 140 m above today's lake surface. They describe the maximum extent of the lake (SCHWARTZ et al. 2012). The Manyara Beds are rich in Pleistocene vertebrate fossils. In the Lake Manyara area, especially close to the village of Makuyuni, two hominin-bearing sites (0.63 and 0.78 Ma), vertebrate fossils and handaxes from different periods were found (KAISER et

al. 2010, FROST et al. 2012). The Manyara Beds consist of a lacustrine grayish lower member (mudstones, siltstones, diatomites, marls and tuff) which was deposited between 1.03 and 0.633 Ma, and a fluvial and terrestrial upper member which is composed of up to 13 m thick reddish brown upper member (siltstones, mudstones, conglomerates and breccias) deposited between 0.633 and 0.44 (0.27) Ma. The transition between both members is in most sections marked by a distinct tephra layer (FROST et al. 2012). Sections of the Manyara Beds are best exposed in the surroundings of the town Makuyuni, where Holocene soils and caliche overlay the sediments and where various gully systems erode into the savanna landscape. Laboratory analysis of representative samples of alluvial material of the Low-

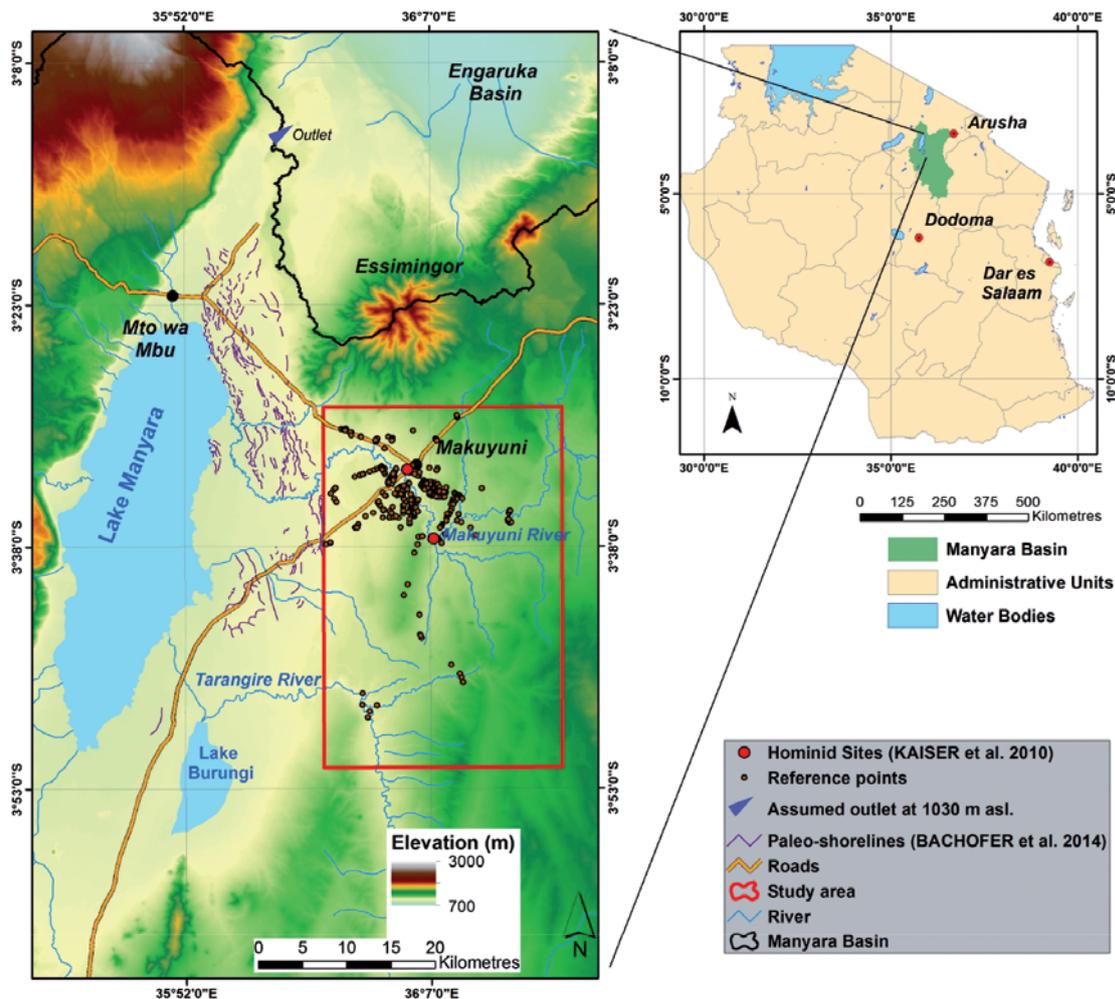


Fig. 1: The study area and the regional setting of Lake Manyara.

er Manyara Beds (LMB) and Upper Manyara Beds (UMB) were taken from the slopes of the Makuyuni river valley. They show heterogeneous texture, but distinct higher carbonate content in LMB and Fe^{2+} content in UMB. Organic carbon is dependent on the topographic position and adjacent soils and vegetation cover. Hence, the Manyara Beds show specific spectral and positional properties. Consequently, a successful delineation of the Manyara Beds with spectral bands, multispectral indices and topographical indices seem possible.

2.3 ASTER Multispectral Data

The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) was launched with NASA's TERRA spacecraft in December 1999 (YAMAGUCHI et al. 1998). Its subsystems cover three bands in the visible-near infrared (VNIR), six bands in the short-wave infrared (SWIR) and five bands in the thermal infrared (TIR) wavelength regions (Tab. 1). The ground resolution is 15 m, 30 m and 90 m respectively (FUJISADA 1995). YAMAGUCHI et al. (1998) stated that the VNIR spectral information was designed for use in mapping vegetation and iron oxides in soil and rocks, while the SWIR wavelengths were designed for soil and mineral mapping.

A cloud free ASTER L1B scene was obtained at August 23, 2006, 8:07 UTC during dry season. Because the SWIR bands of the L1B data is not corrected for a cross-detector leakage, crosstalk correction was applied following IWASAKI et al. (2002) using the correction software product from Earth Remote Sensing Data Applications Centre (ERSDAC).

The average geometric accuracy of the ASTER scene was validated by own ground

control points (GCP) measurements and high resolution WorldView-2 imagery. A mean locational residual error of 61.4 m could be estimated which is close to the residual error calculated by HEWSON et al. (2005). The GCPs and a Landsat ETM+ (L1T) panchromatic scene with 15 m ground resolution were used to improve the geometric accuracy using an automatic point matching algorithm. For our GCPs the Landsat scene showed a total RMSE of 16.5 m. The ASTER scene could be aligned with an RMSE of 0.9. However, the TIR bands were excluded from the analysis because of their low spatial resolution and some artifacts which were visible in the L1B and also in the surface emissivity product (AST05). To preserve the spectral information of the VNIR bands, the SWIR bands were resampled to the respective ground resolution of 15 m.

2.4 Spectral Indices

Multispectral indices derived from ASTER spectral bands are used in a broad range of studies with a main emphasis on vegetation, soil and lithology (MULDER et al. 2011, POUR & HASHIM 2011). The spectral rationing of selective band absorption features of different materials at distinct wavelengths is utilized to emphasize the presence or absence of distinct mineral compositions or vegetation. From an extensive literature review a broad range of indices were collected and processed for this analysis (Tab. 2). Not all minerals, for which the indices were developed, are abundant in the study area. In addition, many of the indices carry redundant information because of the use of similar input bands and band combinations.

Tab. 1: ASTER spectral bands with the minimum lower and maximum upper band edges.

VNIR Green	0.52–0.60 μm	SWIR 3	2.185–2.225 μm
VNIR Red	0.63–0.69 μm	SWIR 4	2.235–2.285 μm
VNIR Near Infrared	0.76–0.86 μm	SWIR 5	2.295–2.365 μm
SWIR 1	1.600–1.700 μm	SWIR 6	2.360–2.430 μm
SWIR 2	2.145–2.185 μm	TIR 1 - 5	8.125–11.65 μm

Tab. 2: Spectral indices of ASTER VNIR and SWIR bands.

Index and literature reference	Formula	Index and literature reference	Formula
AlOH Group (CUDAHY 2012)	$(5/7)$	AKP (ROWAN & MARS 2003)	$(4+6)/5$
Alteration/Laterite (BIERWIRTH 2002)	$(4/5)$	Amphibole (BIERWIRTH 2002)	$(6/8)$
Alunite (POUR & HASHIM 2011)	$(7/5)*(7/8)$	Calcite (POUR & HASHIM 2011)	$(6/8)*(9/8)$
CCE (ROWAN & MARS 2003)	$(7+9)/8$	Dolomite (ROWAN & MARS 2003)	$(6+8)/7$
Clay 1 (ROWAN & MARS 2003)	$(5+7)/6$	MgOH Group (CUDAHY 2012)	$(6+9)/(7+8)$
Clay 2 (BIERWIRTH 2002)	$(5*7)/(6*6)$	MgOH 1 (HEWSON et al. 2005)	$(6+9)/8$
Kaolinitic (HEWSON et al. 2005)	$(7/5)$	MgOH 2 (CUDAHY 2012)	$(7/8)$
Kaolin Group (CUDAHY 2012)	$(6/5)$	Ferric Iron ³ (ROWAN & MARS 2003)	$(2/1)$
Kaolinite (POUR & HASHIM 2011)	$(4/5)*(8/6)$	Ferrous Iron 1 (ROWAN et al. 2005)	$(1/2)$
Muscovite (HEWSON et al. 2005)	$(7/6)$	Ferrous Iron 2 (ROWAN & MARS 2003)	$(5/3)+(1/2)$
OH 1 (POUR & HASHIM 2011)	$(7/6)*(4/6)$	Ferric Oxide (CUDAHY 2012)	$(4/3)$
OH 2 (NINOMIYA et al. 2005)	$(4*7/6)/6$	Gossan (VOLESKY et al. 2003)	$(4/2)$
OH 3 (NINOMIYA et al. 2005)	$(4*7/5)/5$	Opaque Index (CUDAHY 2012)	$(1/4)$
PHI (HEWSON et al. 2005)	$(5/6)$	Ferrous Iron/Silicates (CUDAHY 2012)	$(5/4)$
RBD6 (ROWAN et al. 2005)	$(4+7)/(6*2)$	Burn Index (HUDAK et al. 2004)	$(3-5)/(3+6)$
RBD8 (ROWAN et al. 2005)	$(7+9)/(8*2)$	VI (TUCKER 1979)	$(3/2)$
NDVI (ROUSE et al. 1974)	$(3-2)/(3+2)$	Salinity (AL-KHAIER 2003)	$(4-5)/(4+5)$
STVI (POUR & HASHIM 2011)	$(3/2)*(1/2)$		

Tab. 3: Topographic indices.

Slope (TRAVIS et al. 1975)	Aspect (TRAVIS et al. 1975)
Slope height (BOEHNER & CONRAD 2008)	Valley Depth (BOEHNER & CONRAD 2008)
Standardized Height (BOEHNER & CONRAD 2008)	Normalized Height (BOEHNER & CONRAD 2008)
Mid Slope Position (BOEHNER & CONRAD 2008)	Downslope Distance Gradient (HJERDT et al. 2004)
Plan Curvature (ZEVENBERGEN & THORNE 1987, DIKAU 1988)	Profile Curvature (ZEVENBERGEN & THORNE 1987, DIKAU 1988)
Negative Openness (YOKOYAMA et al. 2002)	Positive Openness (YOKOYAMA et al. 2002)
Morphometric Protection I. (YOKOYAMA et al. 2002)	Terrain Ruggedness Index (RILEY et al. 1999)
Multiresolution Index of Valley Bottom Flatness (GALLANT & DOWLING 2003)	Multiresolution Index of Ridge Top Flatness (GALLANT & DOWLING 2003)
Relative Slope Position (CONRAD 2005)	Geomorphones (JASIEWICZ & STEPINSKI 2013)
Stream Power Index (MOORE et al. 1991)	LS Factor (MOORE et al. 1991)
Terrain Classification Index for Lowlands (BOCK et al. 2007)	Topographic Position Index (GUISAN et al. 1999, JENNESS 2006)
Topographic Wetness Index (BEVEN & KIRKBY 1979)	Vertical Distance to Channel Network (CONRAD 2005)
Elevation (height above sea level; a.s.l) (DLR 2012)	

2.5 Topographic Indices

A track of the shuttle radar topography mission X-band (SRTM-X) digital elevation model (DEM) with 25 m ground resolution covers the study area. To eliminate the noise in the SRTM-X DEM, a multidirectional Lee filter was applied to preserve topographic features (LEE 1980). Different topographic indices (Tab. 3) were derived from the DEM to serve as independent features in the classification. The indices are used to characterize the topographic conditions of the Manyara Bed's location. The selected indices are listed in Tab. 3.

2.6 Training and Reference Data

As reference for this study, 498 ground reference points were collected during field campaigns between 2010 and 2014. Because of the time gap between the acquisition of the ASTER scene and the reference point selection, all points were taken with care even though the landscape is considered as relatively stable in relation to the ground resolution of the ASTER and SRTM-X data. Moreover, we assume that the mineral components of the Manyara Beds are conservative, means that in the study area environment they will change insignificantly over such a time period.

Some parts in the south and southeast of the study area are remote and partly inaccessible. Therefore, we applied a random clustered sampling strategy. The reference points are imbalanced with 40 points describing the UMBs, 139 points describing the LMBs, and 320 points with dissimilar landcover. To the latter class we refer to as "other landcover", which involves a rather complete reference selection of soils, minerals and vegetation within the study area, which were merged to take into account a potential landcover change. UMBs are not as abundant in the field as LMBs, which are the more important sediments for the reconstruction of the paleolake history. 20% of points from each class were randomly selected to serve exclusively as test datasets. Soils which are adjacent to or developed from the Manyara Beds were not classified as Manyara Beds. Reference points were collected describing the relative spatial distri-

bution of the Manyara Beds with a minimum size of at least the VNIR resolution (15 m²) for the UMBs and at least 30 m² for LMBs and "other landcover". The relatively small area for the UMBs was defined because they generally appeared as small sections or outcrops of red tuffs on the valley slopes.

3 Methodology

3.1 Support Vector Machines

The concept of support vector machines (SVM) based supervised classification originates from machine learning methodology and was introduced by VAPNIK (1995, 1999). Due to different characteristics, the SVM algorithm has become very popular for pattern recognition and classification (FOODY & MATHUR 2004). While most remote sensing classification methods are mainly based on statistical properties of pixel and objects, SVMs maximizes the boundaries between intended classes. The problem of linear separating classes in an n-dimensional feature space, resulting from multiple independent input features is solved by applying kernel functions. By maximizing the margin between classes, an optimal separating hyperplane is strived for (BURGES 1998, HEARST 1998). Only a small selection of feature values in the training data, which are close to the margin, are needed to define the hyperplane. These features are referred to as support vectors. Too many outliers within the training dataset would result in an over-fitting of the hyperplane. The cost parameter C determines a penalty for the support vectors which excludes outliers and results in a so called "soft margin". C controls thereby the balance between over-fitting (high values) and generalization (low values) of the maximum margin and must be selected carefully (VAPNIK 1995 & 1999, SCHÖLKOPF & SMOLA 2002, FOODY & MATHUR 2004).

For this analysis, support vector classifier (C-SVC) from the *Library for Support Vector Machines* (LIBSVM) developed by CHANG & LIN (2011) was utilized. It implements a "one-against-one" approach, which builds a classifier for each target value pair. The classification was conducted with a linear kernel and

the radial basis kernel function (RBF) which is widely used when a nonlinear relation is expected (FOODY & MATHUR 2004). The width of the RBF or Gaussian kernel is controlled by the constant γ , with high values describing a far reaching influence of the training sample and a low value for influencing the adjacent feature space. A grid search was applied by iteratively cross-validating the accuracy of test data classification, while optimizing the constants C and γ . All input feature sets were scaled to the range $[-1, +1]$. For the selection of features we applied the recursive feature elimination (RFE) technique which is widely used with SVM approaches following GUYON & ELISSEFF (2003). In an iterative process the features are weighted according to their ability of discriminating the target classes. At each step the most insignificant features are eliminated recursively.

3.2 Boosted Regression Trees

Boosted regression trees (BRT), also known as stochastic gradient boosting (ELITH et al. 2006), combine classification and regression trees with the gradient boosting algorithm (FRIEDMAN 2001). This method employs a learning algorithm to identify a model that fits best the relationship between a feature set and the class label of the target classes. We ran the model using the free statistical programming language R (R DEVELOPMENT CORE TEAM 2008) with the package *adabag* (ALFARO et al. 2013). The boosting algorithm used in *adabag* is the AdaBoost (adaptive boosting) algorithm based on FREUND & SCHAPIRE (1996). The goal of the algorithm is to improve the accuracy of a tree by combining single predictor variables into classifiers. The points along the tree where the features are split are called nodes. Bagging reduces the variance and hence increases the prediction accuracy by taking repeated samples from the training dataset to build a prediction model and then averages the resulting predictions. Boosting constructs each tree on the original dataset but each tree is grown using information from previously grown trees. When a binary classification problem is extended to a multi-class classification problem, most boosting algorithms have to reduce the

multi-class classification problem to multiple binary-class problems. However, the AdaBoost.M1 and the SAMME algorithms extend the AdaBoost algorithm to the multi-class case (ZHU et al. 2009). The difference between the algorithms is the calculation of the α constant, which estimates the error of the classifier for each tree iteration. For binary classifications SAMME is equivalent to AdaBoost.M1. The measure of the relative importance of the input features uses the gain of the Gini index (ALFARO et al. 2013), which measures the divergences between the probability distribution of the values of a feature. Best results for both approaches were achieved with the building of 500 trees and 5 nodes for each tree.

4 Results

The classification of the Manyara Beds for two classes (LMB, “other landcover”) and three classes (LMB, UMB, “other landcover”) with SVM and BRT was conducted with different sets of input feature combinations. As shown in Tabs. 4 and 5, the nine “Spectral Bands” of ASTER, the derived “Spectral Indices” and “Topographic Indices” derived from the DEM were considered for classification separately. Additionally, we show also the model performance considering “All Features” as well as a “Selection of Features” which were identified for SVM and BRT as the features with the highest importance for solving the classification problem (Fig. 2). The number of relevant features varies between the different methods and parameterizations. The smallest number of features (20) was identified for the SVM with an RBF kernel and the binary classification problem, the highest number (66) for the BRT approaches (Tabs. 4 and 5). The BRT methods stabilized with about 80 trees for the binary classification and for the multi-class problem between 130 and 500, depending on the feature set (Fig. 3).

The highest accuracy with 92% was achieved with SVM (linear kernel) and all “Spectral Bands” as well as “Spectral Indices” and “Topographic Indices” for the binary classification of LMB against “other landcover” (Tab. 4). Both linear and RBF kernels perform for the two classes with similar ac-

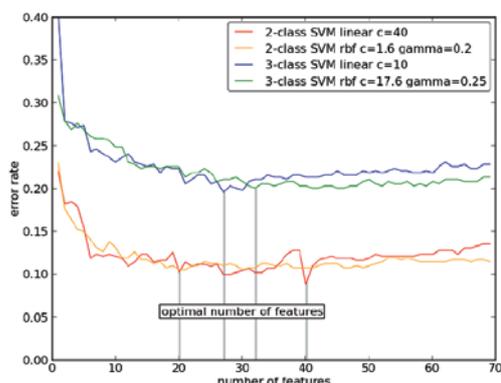


Fig. 2: Relationship between error rate and number of features selected by SVM-RFE.

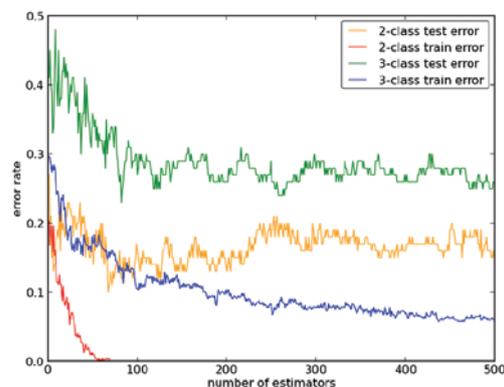


Fig. 3: AdaBoost SAMME error rate in relation to the number of trees for selected features.

Tab. 4: Overall accuracy and applied parameters for SVM and BRT. Binary classification scheme: LMB & “other landcover”. Highest overall accuracy displayed in boldface (no = number of features).

	Spectral Bands (no: 9)	Spectral Indices (no: 35)	Topographic Indices (no: 25)	All Features (no: 69)	Selected Features
SVM linear	85% C: 75	86% C: 8	86% C: 8	92% C: 50	91% (no: 40) C: 40
SVM RBF	84% C: 29 γ : 1.5	87% C: 39.5 γ : 0.135	88% C: 50 γ : 1	89% C: 1.5 γ : 0.16	89% (no: 20) C: 1.6 γ : 0.2
BRT AdaBoost.M1	82%	85%	86%	89%	90% (no: 66)

Tab. 5: Overall accuracy and applied parameters for SVM and BRT. Three-class classification scheme: LMB, UMB & “other landcover”. Highest overall accuracy displayed in boldface (no = number of features).

	Spectral Bands (no: 9)	Spectral Indices (no: 35)	Topographic Indices (no: 25)	All Features (no: 69)	Selected Features
SVM linear	74% C: 2	73% C: 4	72% C: 8	79% C: 1.5	80% (no: 27) C: 10
SVM RBF	75% C: 20 γ : 1.6	78% C: 12.5 γ : 0.105	73% C: 3 γ : 0.15	80% C: 3 γ : 0.235	80% (no: 32) C: 17.6 γ : 0.25
BRT AdaBoost.M1	75%	76%	72%	78%	78% (no: 66)
BRT SAMME	72%	75%	73%	79%	79% (no: 66)

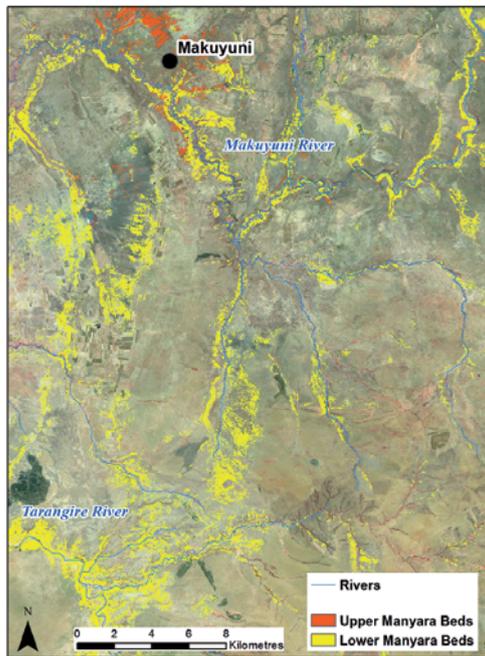


Fig. 4: Three-class SVM RBF result with “All Features”, ASTER false colour infrared image (near-infrared, red, green) as background.

curacies. Only for the classification with “All Features” and the “Selected Features” the linear kernel can achieve a higher accuracy. The accuracies show a slightly lower performance for BRT than for both SVM methods.

When expanding the classification problem with the LMBs as the third class the overall accuracy drops with all possible combinations. The RBF kernel and the linear kernel perform similarly well and achieve 80% accuracy with “All Features” (Tab. 5). Fig. 4 shows the associated spatial distribution of the LMBs and UMBs. Stratigraphic units with similar topographic and spectral properties compared to the LMBs were identified in the south of the study area in the Tarangire River valley and further east in the Makuyuni River valley.

5 Discussion

In comparing the SVM classifier with linear and RBF kernels, as well as with the BRT Adaboost.M1, results show that the binary classification problem can be solved with high accuracies of up to 92%. Regarding the binary

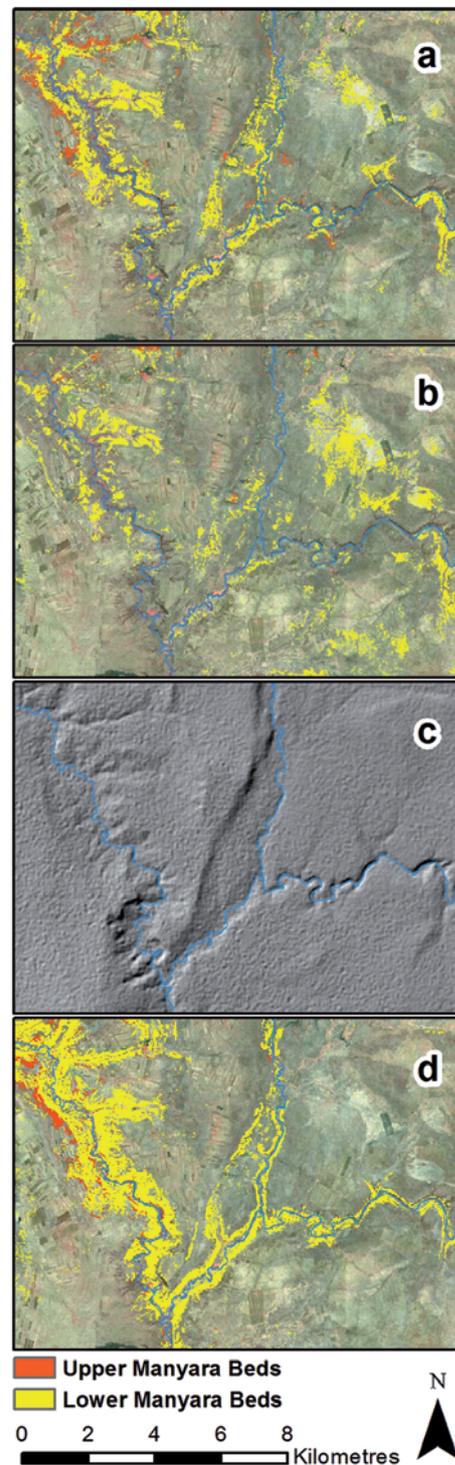


Fig. 5: Three-class SVM RBF comparison of different input features. ASTER false colour infrared image as background; a) SVM RBF with “All Features”; b) SVM RBF with 9 “Spectral Bands” of ASTER; c) hillshade of DEM; d) SVM RBF with “Topographic Indices”.

classification problem, the linear and the RBF kernels yielded a very good performance. The BRT model performs marginally lower, but also at a high level. By enlarging the classification problem to three classes the overall accuracy drops by nearly 12% on average (Tabs. 4 and 5). The small number of UMB training features and their irregular spatial distribution result in an imbalanced training set and cause a lower overall accuracy. As for the three-class classification, the SVM with the RBF kernel as well as the SVM paired with the linear kernel perform slightly better than the BRT methods. The SVM three-class problem requires a higher generalization, leading to a wider hyperplane margin. The SVM binary problem however displays no such requirement (lower C values).

“Spectral Bands”, “Spectral Indices” and “Topographic Indices” may be used as input features to explain the location of LMBs and UMBs. Both SVM methods perform similarly when using “All Features” and the RFE-“Selected Features”. The same is true for the BRT methods. When using only “Topographic Indices” the binary classification approach detects the distribution of the Lower Manyara Beds better than with spectral information. Though for the three-class approach, the Lower and the Upper Manyara Beds seem to be separated better by spectral input features. “Spectral Bands” and “Spectral Indices” identify areas where the spectral information of the target classes is not (or only marginally) disturbed by heterogeneous land cover (Fig. 5b). The use of solely “Topographic Indices” results in a separation of the different topographic positions of the Manyara Beds within the study area (Fig. 5c, d). Consequently, the “Topographic Indices” with the multi-class scheme achieved the lowest accuracies because the topographic characteristics of the LMBs and the UMBs partly overlap. Accordingly, the combination of both spectral and topographic features best explains the distribution of the Manyara Beds (Fig. 5a).

Several optimized feature sets were identified which explain the distribution of the target classes with the training data (Fig. 2). Eight input features are common in all optimized feature sets (Tab. 6). The spectral feature VNIR Green (ASTER band 1) correlates with the

Tab. 6: Features which are common in all feature selection results.

VNIR Green	SWIR 6
Opaque Index	Stream Power Index
Multiresolution Index of Valley Bottom Flatness	Positive Openness
Plan Curvature	Elevation (height a.s.l.)

Manyara Beds. The spectral “Opaque Index” (CUDAHY 2012) is sensitive to magnetite-bearing rocks, maghemite gravels, and manganese oxides. Whether the value distribution of the “Opaque Index” for this ASTER scene results from those specific absorption features can only be determined through further laboratory analysis or field spectroscopy. The spectral reflection characteristics may also result from other materials. The topographic information “Elevation (height a.s.l.)” describes the deposition on a distinct paleolake level. The topographic index “Positive Openness” expresses the degree of geometric dominance of one or several convex relief features. It therefore accurately highlights elevated areas (YOKOYAMA et al. 2002). “Plan Curvature” differentiates between ridges and valleys and may describe the incision of streams and gully systems into the lacustrine sediments. The “Multiresolution Index of Valley Bottom Flatness” (GALLANT & DOWLING 2003) may describe the depositional areas of the Manyara Beds.

Lacustrine sediments similar to the LMB are identified in the eastern part of the study area. After the first classification results we conducted a field check proving the predicted lacustrine sediments in the eastern part of the study area. In addition, vertebrate fossils are abundant at this location. Since the elevation of these sediments is higher than the LMBs, their elevation can be explained by a tectonic downshift of the block with the LMB as is proposed by SCHWARTZ et al. (2012). This must have happened after or during the sedimentation of the Manyara Beds. The second explanation would propose a lake or swamp situation parallel to the paleolake Manyara. A gneissic ridge that is incised today by the Makuyuni River would have functioned as a barrier. The drainage of this lake was directed into the Tarangire River

in the south. Knick points of tributary river valleys indicate a change of the drainage direction. The third possibility would be a combination of both scenarios. For a further deciphering of the lake history a better dating of volcanic ridges and tuffs is needed.

6 Conclusions

This paper focused on a comparison between SVM and BRT methods, as well as the use of different sets of input features. All methods had performed with similar overall accuracies for the multi-class and the binary problem respectively. The best results were obtained using all spectral and topographic features to explain the distribution of the Manyara Beds. The highest accuracy for the binary classification of LMBs and “other landcover” were achieved with the SVM method and a linear kernel. For the classification of three classes (LMBs, UMBs, “other landcover”) the SVM method with a RBF, as well as with a linear kernel performed best. In the case of the Lake Manyara area, small scale lithostratigraphic units could be delineated in a challenging environment, which entails a heterogeneous landcover with spectral similarity between different soils and a patchy vegetation cover. Besides the already known finding localities, new sites with paleolake sediments east of Makuyuni could be identified and may contribute to assessing the paleolake history of Lake Manyara. Lacustrine sediments in the south of the study area have to be validated in the field and their paleontological importance must be evaluated.

Acknowledgements

We would like to thank the editor and the reviewers for their time and valuable remarks. This study was financed by the Heidelberg Academy of Sciences and Humanities research center: “The Role of Culture in Early Expansions of Humans” (ROCEEH). The ASTER LIB data were obtained through the online Data Pool at the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Sci-

ence (EROS) Center, Sioux Falls, South Dakota, USA. We would like to thank the DLR and the German Remote Sensing Data Center (DFS) for providing the SRTM/X-SAR data.

References

- ALFARO, E., GAMEZ, M. & GARCÍA, N., 2013: adabag: An R Package for Classification with Boosting and Bagging. – *Journal of Statistical Software* **54** (2): 1–35.
- AL-KHAIER, F., 2003: Soil Salinity Detection Using Satellite Remote Sensing. – Thesis, ITC Enschede, Netherlands.
- BACHOFER, F., QUÉNÉHERVÉ, G. & MÄRKER, M., 2014: The Delineation of Paleo-Shorelines in the Lake Manyara Basin Using TerraSAR-X Data. – *Remote Sensing* **6** (3): 2195–2212.
- BEVEN, K. & KIRKBY, M.J., 1979: A physically based, variable contributing area model of basin hydrology. – *Hydrological Sciences Journal* **24** (1): 43–69.
- BIERWIRTH, P., 2002: Evaluation of ASTER Satellite Data for Geological Applications. – Consultancy Report to Geoscience Australia.
- BOCK, M., BÖHNER, J., CONRAD, O., KÖTHE, R. & RINGELER, A., 2007: XV. Methods for creating Functional Soil Databases and applying Digital Soil Mapping with SAGA GIS. – JRC Scientific and technical Reports, Office for Official Publications of the European Communities, Luxembourg.
- BOEHNER, J. & CONRAD, O., 2008: Terrain Parameters described in the SAGA-GIS Software, v.2.1.0. <http://sourceforge.net/projects/saga-gis/files/latest/download?source=files> (16.6.2014).
- BURGES, C., 1998: A Tutorial on Support Vector Machines for Pattern Recognition. – *Data Mining and Knowledge Discovery* **2** (2): 121–167.
- CASANOVA, J. & HILLAIRE-MARCEL, C., 1992: Chronology and paleohydrology of late Quaternary high lake levels in the Manyara basin (Tanzania) from isotopic data (18O, 13C, 14C, ThU) on fossil stromatolites. – *Quaternary Research* **38** (2): 205–226.
- CHAN, J.C.-W. & PAELINCKX, D., 2008: Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. – *Remote Sensing of Environment* **112** (6): 2999–3011.
- CHANG, C.C. & LIN, C.J., 2011: LIBSVM: A library for support vector machines. – *ACM Transac-*

- tions on Intelligent Systems and Technology **2** (3): 1–27.
- CONRAD, O., 2005: Terrain Parameters described in the SAGA-GIS Software, v.2.1.0. <http://sourceforge.net/projects/saga-gis/files/latest/download?source=files> (16.6.2014).
- CUDAHY, T., 2012: Satellite ASTER Geoscience Product. – Notes for Australia; CSIRO: http://c3dmm.csiro.au/WA_ASTER/WA%20ASTER%20Geoscience%20Product%20Notes%2015112011.pdf (16.6.2014).
- DEUS, D., GLOAGUEN, R. & KRAUSE, P., 2013: Water Balance Modeling in a Semi-Arid Environment with Limited in situ Data Using Remote Sensing in Lake Manyara, East African Rift, Tanzania. – *Remote Sensing* **5** (4): 1651–1680.
- DIKAU, R., 1988: Entwurf einer geomorphographisch-analytischen Systematik von Reliefeinheiten. – Heidelberg Geographische Bausteine, Heidelberg.
- DLR, 2012: SRTM X-SAR Digital Elevation Models. Status: 2012-09-28. http://eoweb.dlr.de:8080/eoweb-ng/licenseAgreements/DLR_SRTM_Readme.pdf (23.12.2013).
- ELITH, J., GRAHAM, C.H., ANDERSON, R.P., DUDÍK, M., FERRIER, S., GUISAN, A., HIJMANS, R.J., HUETTMANN, F., LEATHWICK, J.R., LEHMANN, A., LI, J., LOHMANN, L.G., LOISELLE, B.A., MANION, G., MORITZ, C., NAKAMURA, M., NAKAZAWA, Y., OVERTON, J.M.M., TOWNSEND PETERSON, A., PHILLIPS, S.J., RICHARDSON, K., SCACHETTI-PEREIRA, R., SCHAPIRE, R.E., SOBERÓN, J., WILLIAMS, S., WISZ, M.S. & ZIMMERMANN, N.E., 2006: Novel methods improve prediction of species' distributions from occurrence data. – *Ecography* **29** (2): 129–151.
- ELITH, J., LEATHWICK, J.R. & HASTIE, T., 2008: A working guide to boosted regression trees. – *Journal of Animal Ecology* **77** (4): 802–813.
- ELMAHDY, S.I., 2012: Hydromorphological Mapping and Analysis for Characterizing Darfur Paleolake, NW Sudan Using Remote Sensing and GIS. – *International Journal of Geosciences* **2012** (3): 25–36.
- EL-SHEIKH, A., ABDELSALAM, M.G. & MICKUS, K., 2011: Geology and geophysics of the West Nubian Paleolake and the Northern Darfur Megalake (WNPL-NDML): Implication for groundwater resources in Darfur, northwestern Sudan. – *Journal of African Earth Sciences* **61** (1): 82–93.
- ESCH, T., HIMMLER, V., SCHORCHT, G., THIEL, M., WEHRMANN, T., BACHOFER, F., CONRAD, C., SCHMIDT, M. & DECH, S., 2009: Large-area assessment of impervious surface based on integrated analysis of single-date Landsat-7 images and geospatial vector data. – *Remote Sensing of Environment* **113** (8): 1678–1690.
- FOODY, G.M. & AJAY, M., 2004: A relative evaluation of multiclass image classification by support vector machines. – *IEEE Transactions on Geoscience and Remote Sensing* **42** (6): 1335–1343.
- FOODY, G.M. & MATHUR, A., 2004: Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. – *Remote Sensing of Environment* **93** (1–2): 107–117.
- FREUND, Y. & SCHAPIRE, R.E., 1996: Experiments with a New Boosting Algorithm. – Thirteenth International Conference on Machine Learning, 148–156, Bari, Italy.
- FRIEDMAN, J.H., 2001: Greedy function approximation: a gradient boosting machine. – *Annals of Statistics*, 1189–1232.
- FROST, S.R., SCHWARTZ, H.L., GIEMSCH, L., MORGAN, L.E., RENNE, P.R., WILDGOOSE, M., SAANANE, C., SCHRENK, F. & HARVATI, K., 2012: Refined age estimates and Paleoanthropological investigation of the Manyara Beds, Tanzania. – *Journal of Anthropological Sciences* **90**: 1–12.
- FUJISADA, H., 1995: Design and performance of ASTER instrument. – SPIE, the Advanced and Next generation Satellites **2583**: 16–25, Paris, France.
- GABER, A., GHONEIM, E., KHALAF, F. & EL-BAZ, F., 2009: Delineation of paleolakes in the Sinai Peninsula, Egypt, using remote sensing and GIS. – *Journal of Arid Environments* **73** (1): 127–134.
- GALLANT, J.C. & DOWLING, T.I., 2003: A multiresolution index of valley bottom flatness for mapping depositional areas. – *Water Resources Research* **39** (12): 1347.
- GESSNER, U., MACHWITZ, M., CONRAD, C. & DECH, S., 2013: Estimating the fractional cover of growth forms and bare surface in savannas. A multi-resolution approach based on regression tree ensembles. – *Remote Sensing of Environment* **129**: 90–102.
- GHONEIM, E., BENEDETTI, M. & EL-BAZ, F., 2012: An integrated remote sensing and GIS analysis of the Kufrah Paleoriver, Eastern Sahara. – *Geomorphology* **139–140**: 242–257.
- GÓMEZ, C., WULDER, M.A., MONTES, F. & DELGADO, J.A., 2012: Modeling Forest Structural Parameters in the Mediterranean Pines of Central Spain using QuickBird-2 Imagery and Classification and Regression Tree Analysis (CART). – *Remote Sensing* **4** (1): 135–159.
- GUISAN, A., WEISS, S. & WEISS, A., 1999: GLM versus CCA spatial modeling of plant species distribution. – *Plant Ecology* **143** (1): 107–122.
- GUYON, I. & ELISSEEFF, A., 2003: An Introduction to Variable and Feature Selection. – *Journal of Machine Learning Research* **3**: 1157–1182.
- HAHN, C. & GLOAGUEN, R., 2008: Estimation of soil types by non linear analysis of remote sensing

- data. – *Nonlinear Processes in Geophysics* **15** (1): 115–126.
- HEARST, M.A., 1998: Support Vector Machines. – *IEEE Intelligent Systems* **13** (4): 18–28.
- HEWSON, R.D., CUDAHY, T.J., MIZUHIKO, S., UEDA, K. & MAUGER, A.J., 2005: Seamless geological map generation using ASTER in the Broken Hill-Curnamona province of Australia. – *Remote Sensing of Environment* **99** (1–2): 159–172.
- HJERDT, K.N., MCDONNELL, J.J., SEIBERT, J. & RODHE, A., 2004: A new topographic index to quantify downslope controls on local drainage. – *Water Resources Research* **40** (5): W05602.
- HUDAK, A.T., ROBICHAUD, P., EVANS, J.S., CLARK, J., LANNOM, K., MORGAN, P. & STONE, C., 2004: Field validation of Burned Area Reflectance Classification (BARC) products for post fire assessment. – University of Nebraska, Lincoln, NE, USA.
- IWASAKI, A., FUJISADA, H., AKAO, H., SHINDOU, O. & AKAGI, S., 2002: Enhancement of spectral separation performance for ASTER/SWIR. – *SPIE, Infrared Spaceborne Remote Sensing* **IX 4486**: 42–50, San Diego, CA, USA.
- JASIEWICZ, J. & STEPINSKI, T.F., 2013: Geomorphons – a pattern recognition approach to classification and mapping of landforms. – *Geomorphology* **182**: 147–156.
- JENNESS, J., 2006: Topographic Position Index (TPI) v.1.3a. – TPI_Documentation.pdf. <http://www.jennessent.com/arcview/tpi.htm> (17.6.2013).
- KAISER, T.M., SEIFFERT, C., HERTLER, C., FIEDLER, L., SCHWARTZ, H.L., FROST, S.R., GIEMSCH, L., BERNOR, R.L., WOLF, D., SEMPREBON, G., NELSON, S.V., SCHRENK, F., HARVATI, K., BROMAGE, T.G. & SAANANE, C., 2010: Makuyuni, a new Lower Palaeolithic Hominid Site in Tanzania. – *Mitteilungen Hamburgisches Zoologisches Museum und Institut* **106**: 69–110.
- KELLER, C.M., HANSEN, C. & ALEXANDER, C.S., 1975: Archaeology and Paleoenvironments in the Manyara and Engaruka Basins, Northern Tanzania. – *Geographical Review* **65** (3): 364–376.
- LEE, J.S., 1980: Digital image enhancement and noise filtering by use of local statistics. – *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2** (2): 165–168.
- MOORE, I.D., GRAYSON, R.B. & LADSON, A.R., 1991: Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. – *Hydrological Processes* **5** (1): 3–30.
- MOUNTRAKIS, G., IM, J. & OGOLE, C., 2011: Support vector machines in remote sensing: A review. – *ISPRS Journal of Photogrammetry and Remote Sensing* **66** (3): 247–259.
- MULDER, V.L., DE BRUIN, S., SCHAEPMAN, M.E. & MAYR, T.R., 2011: The use of remote sensing in soil and terrain mapping – A review. – *Geoderma* **162** (1–2): 1–19.
- NINOMIYA, Y., FU, B. & CUDAHY, T.J., 2005: Detecting lithology with Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) multispectral thermal infrared “radiance-at-sensor” data. – *Remote Sensing of Environment* **99** (1–2): 127–139.
- PAL, M. & MATHER, P.M., 2003: An assessment of the effectiveness of decision tree methods for land cover classification. – *Remote Sensing of Environment* **86** (4): 554–565.
- PAL, M. & FOODY, G.M., 2010: Feature Selection for Classification of Hyperspectral Data by SVM. – *IEEE Transactions on Geoscience and Remote Sensing* **48** (5): 2297–2307.
- POUR, A.B. & HASHIM, M., 2011: Application of advanced spaceborne thermal emission and reflection radiometer (ASTER) data in geological mapping. – *International Journal of the Physical Sciences* **6** (33): 7657–7668.
- PRENDERGAST, M.E., MABULLA, A.Z.P., GRILLO, K.M., BRODERICK, L.G., SEITSONEN, O., GIDNA, A.O. & GIFFORD-GONZALEZ, D., 2013: Pastoral Neolithic sites on the southern Mbulu Plateau, Tanzania. – *Azania: Archaeological Research in Africa* **48** (4): 498–520.
- R DEVELOPMENT CORE TEAM, 2008: R: A Language and Environment for Statistical Computing. – R Foundation for Statistical Computing, Vienna, Austria.
- RILEY, S.J., DEGLORIA, S.D. & ELLIOT, R., 1999: A Terrain Ruggedness Index that quantifies topographic heterogeneity. – *Intermountain Journal of Sciences* **5** (1–4): 23–27.
- ROUSE, J.W., HAAS, R.H., SHELL, J.A., DEERING, D.W. & HARLAN, J.C., 1974: Monitoring the vernal advancement of retrogradation of natural vegetation. – Final Report, Type III, NASA/GSFC, Greenbelt, MD, USA.
- ROWAN, L.C. & MARS, J.C., 2003: Lithologic mapping in the Mountain Pass, California area using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data. – *Remote Sensing of Environment* **84** (3): 350–366.
- ROWAN, L.C., MARS, J.C. & SIMPSON, C.J., 2005: Lithologic mapping of the Mordor, NT, Australia ultramafic complex by using the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER). – *Remote Sensing of Environment* **99** (1–2): 105–126.
- SCHLÜTER, T., KOHRING, R. & MEHL, J., 1992: Hyperostotic fish bones (“Tilly bones”) from presumably Pliocene phosphorites of the Lake Man-

- yara area, northern Tanzania. – *Paläontologische Zeitschrift* **66** (1–2): 129–136.
- SCHÖLKOPF, B. & SMOLA, A.J., 2002: Learning with kernels: support vector machines, regularization, optimization, and beyond. – MIT Press, Cambridge, MA, USA.
- SCHWARTZ, H., RENNE, P.R., MORGAN, L.E., WILDGOOSE, M., LIPPERT, P.C., FROST, S.R., HARVATI, K., SCHRENK, F. & SAANANE, C., 2012: Geochronology of the Manyara Beds, northern Tanzania: New tephrostratigraphy, magnetostratigraphy and $^{40}\text{Ar}/^{39}\text{Ar}$ ages. – *Quaternary Geochronology* **7**: 48–66.
- TRAVIS, M.R., ELSNER, G.H., IVERSON, W.D. & JOHNSON, C.G., 1975: VIEWIT: computation of seen areas, slope, and aspect for land-use planning. – Report PSW-11, Berkeley, CA, USA.
- TUCKER, C.J., 1979: Red and photographic infrared linear combinations for monitoring vegetation. – *Remote Sensing of Environment* **8** (2): 127–150.
- VAPNIK, V.N., 1995: The nature of statistical learning theory. – Springer, New York, NY, USA.
- VAPNIK, V.N., 1999: An overview of statistical learning theory. – *Neural Networks, IEEE Transactions* **10** (5): 988–999.
- VOLESKY, J.C., STERN, R.J. & JOHNSON, P.R., 2003: Geological control of massive sulfide mineralization in the Neoproterozoic Wadi Bidah shear zone, southwestern Saudi Arabia, inferences from orbital remote sensing and field studies. – *Precambrian Research* **123** (2–4): 235–247.
- WANG, X., NIU, R. & WU, K., 2011: Lithology intelligent identification using support vector machine and adaptive cellular automata in multi-spectral remote sensing image. – *Optical Engineering* **50** (7): 076201, 1–12.
- YAMAGUCHI, Y., KAHLE, A.B., TSU, H., KAWAKAMI, T. & PNIEL, M., 1998: Overview of Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER). – *IEEE Transactions on Geoscience and Remote Sensing* **36** (4): 1062–1071.
- YOKOYAMA, R., SHIRASAWA, M. & PIKE, R.J., 2002: Visualizing topography by openness: a new application of image processing to digital elevation models. – *Photogrammetric engineering and remote sensing* **68** (3): 257–266.
- ZEVENBERGEN, L.W. & THORNE, C.R., 1987: Quantitative analysis of land surface topography. – *Earth Surface Processes and Landforms* **12** (1): 47–56.
- ZHU, J., ZOU, H., ROSSET, S. & HASTIE, T., 2009: Multi-class AdaBoost. – *Statistics and its Interface* **2**: 349–360.

Address of the Authors:

Dipl.-Geogr. FELIX BACHOFER, Dipl.-Geogr. GERALDINE QUÉNÉHERVÉ, Prof. Dr. VOLKER HOCHSCHILD, Eberhard Karls Universität Tübingen, Geographisches Institut, D-72070 Tübingen, Tel.: +49-7071-29-77528, Fax: +49-7071-29-5378, e-mail: felix.bachofer@uni-tuebingen.de

Dr. MICHAEL MÄRKER, Heidelberger Akademie der Wissenschaften, ROCEEH, D-72070 Tübingen, Tel.: +49-7071-29-72135, e-mail: michael.maerker@geographie.uni-tuebingen.de

Manuskript eingereicht: Juni 2014

Angenommen: September 2014