PFG 2015 / 1, 0007–0020
Stuttgart, February 2015

Article

Impact of Categorical and Spatial Scale on Supervised Crop Classification using Remote Sensing

Fabian Löw, Würzburg, Grégory Duveiller, Ispra, Italy, Christopher Conrad, Würzburg & Ulrich Michel, Heidelberg

Keywords: agricultural monitoring, crop classification, categorical scale, spatial scale, pixel purity

Summary: High temporal revisit frequency over vast geographic areas is necessary to properly use satellite earth observation for monitoring agricultural production. However, this often limits the spatial resolution that can be used. The challenge of discriminating pixels that correspond to a particular crop type, a prerequisite for crop specific monitoring remains daunting when the signal encoded in pixels stems from several land uses (mixed pixels). Naturally, the concept of spatial scale arises but the issue of selecting a proper class legend (the categorical scale) should not be neglected. A framework is presented that addresses these issues and that can be used to quantitatively define pixel size requirements for crop identification and to assess the effect of categorical scale. The framework was applied over two agricultural landscapes. It was demonstrated that there was no unique spatial resolution that provided the best classification result for all classes at once at a given categorical scale. The suitability of pixel populations characterized by pixel size and purity differed for identifying specific crops within tested landscapes, and for one crop there were large differences among the landscapes. In the context of agricultural crop growth monitoring the framework described above can be used to draw guidelines for selecting appropriate imagery, e.g. suitable pixel sizes, and for selecting class legends suitable for accurate crop classification when the interest is only on pixels covering arable land as a prerequisite for crop specific monitoring. The framework could be used to plot the suitability (or accuracy) of pixels as a function of their purity to provide a spatial assessment of classification performance.

Zusammenfassung: *Einfluss der thematischen und räumlichen Auflösung auf die überwachte, fernerkundungsbasierte Feldfrucht-Klassifizierung.* Häufige und regelmäßige Aufnahmen über großen Gebieten sind wichtige Voraussetzungen für das Monitoring von Agrarproduktion basierend auf Erdbeobachtungsdaten. Jedoch schränken diese Voraussetzungen oftmals die räumliche Auflösung (Pixelgröße) ein, welche von bestehenden Sensorsystemen genutzt werden kann. Die Unterscheidung unterschiedlicher Landnutzungstypen, eine Voraussetzung für ein Feldfrucht spezifisches Monitoring, mittels Klassifizierung wird erschwert, wenn das in einem Pixel kodierte Signal von mehreren Landnutzungstypen stammt (Mischpixel-Problematik). Dies wirft Fragen bezüglich der Wahl der optimalen Pixelgröße, aber auch der thematischen Auflösung, also eines geeigneten Klassenschlüssels auf. Um diese Fragen zu beantworten, wird eine Methode vorgestellt, um quantitativ geeignete Charakteristika von Pixelpopulationen hinsichtlich deren Größe und Reinheit in Bezug auf die zu klassifizierende Klasse zu bestimmen. Zudem wurde der Einfluss von verschiedenen Klassenschlüsseln auf das Klassifizierungsergebnis untersucht. Die Methode wurde in zwei landwirtschaftlich genutzten Gebieten getestet. Es wurde gezeigt, dass es keine spezifische „optimale" Pixelgröße gibt, welche für alle Klassen und bei einem bestimmten Klassenschlüssel gleichermaßen das beste Klassifizierungsergebnis liefert. Die Eignung von Pixelpopulationen charakterisiert durch Pixelgröße und -reinheit unterschied sich innerhalb einer bestimmten Landschaft für verschiedene Landnutzungsklassen bzw. für eine spezifische Klasse in verschiedenen Landschaften deutlich. Die vorgestellte Methode kann im Kontext von satellitengestütztem Agrar-Monitoring genutzt werden, um Empfehlungen für die Wahl von geeigneten Pixelgrößen und Sensoren sowie auch geeigneter Klassenschlüssel zu formulieren.

www.schweizerbart.de
DOI: 10.1127/pfg/2015/0252
1432-8364/15/0252 $ 3.50

1 Introduction

Crop type identification and discrimination are essential for subsequent crop-specific agricultural production monitoring using satellite earth observation (EO), e.g. when crop maps are used as input for agricultural modelling. However, the high temporal revisit frequency and the large geographic swath that are required to do a proper monitoring often limit the spatial resolution that can be used. An instrument that satisfies the criteria of swath and revisit frequency is MODIS, but its spatial resolution of 250/500 m is often coarser than desired for many agricultural landscapes. When such data is used as input for crop classification, its coarse observation supports can lead to non-detection of certain land use fragments, e.g. when individual fields of a certain crop type are smaller than individual pixels and the signal encoded in coarser pixels stems from several land uses (mixed pixels). The question of determining the optimal pixel size for an application such as crop identification is therefore inclined towards finding the coarsest acceptable pixel sizes.

When discussing spatial scale, the issue of the choice of a categorical scale, e.g. the number and type of classes used in classification, or class legend, naturally arises and should not be neglected (Ju et al. 2005). In general it would be desirable to have all land use types in a landscape included in the class legend. In reality, however, not all classes will be present in all regions of an image at all scales and there is indication in the literature that there exist different ranges of "optimal" pixel sizes for different classes (Aplin 2006). When selecting coarser pixels the spatial scale increases relative to the patch sizes, e.g. size of agricultural fields, in the underlying landscape and forces to use labels of coarser categorical scale, e.g. "arable land" instead of land use types like "rice" or "wheat" (Ju et al. 2005). In such a situation classification quality can deteriorate when selecting pixel sizes that are too coarse since this can result in excessive mixed pixels when the heterogeneity of the land cover class in one pixel increases (Hsieh et al. 2001, Smith et al. 2003). Selecting too small pixels can result in increased within class variability. Such variation can lead to errors in the class identification (Atkinson & Aplin 2004, Cushnie 1987, Hsieh et al. 2001), and better classification accuracies may sometimes be attained using coarser pixel sizes (McCloy & Bøcher 2007). The issue of pixel size and its implications for image classification have long been noted (Hsieh et al. 2001, Marceau et al. 1994a, McCloy & Bøcher 2007, Woodcock & Strahler 1987) and the selection of one single optimal scale has been questioned (Aplin 2006, Levin 1992). Yet, the effect of categorical scale in combination with spatial scale has been analysed to a lesser extent. Marceau et al. (1994b) demonstrated for different pixel sizes that the definition of the categorical scale can considerably influence classification accuracy. It was shown that the range of pixel sizes for which maximum accuracy can be achieved is specific to certain land cover categories (McCloy & Bøcher 2007). Ju et al. (2005) suggested a method to jointly adapt spatial, e.g. pixel size, and categorical scale in an object-based context. But the dominant standard in remote sensing classification studies often remained being a classification at a single spatial and a fixed categorical scale, i.e. with category labels from only one level of a potential categorical hierarchy (Ju et al. 2005).

But what type of remote sensing data with respect to spatial resolution should be used as classification input and what is the impact of categorical scale on the definition of suitable pixel sizes? To answer these questions, this study builds upon a previous framework by Löw & Duveiller (2014), which was used to quantitatively define pixel size requirements for crop identification via image classification. In this study, this framework will be extended to assess the impact of categorical scale and to explore trade-offs between pixel size and pixel purity when addressing the question of crop classification via image classification. The analysis is restricted to pixels covering arable land using a dedicated masking procedure, based on high-resolution data and simulating how much crop specific signal is encoded in coarser pixels. This procedure restricts the analysis to a subset of the region's pixels to better explore the effect of pixel purity on crop classification, thereby allowing to draw guidelines for selecting suitable class legends and

pixel sizes, respectively. More specifically, this study assesses the impact of the categorical scale on crop classification accuracy in different agricultural landscapes, using satellite images with different pixel sizes as classification input. Second, the impact of the categorical scale on the definition of suitable pixel sizes for crop classification, e.g. maximum and minimum tolerable pixel sizes, is analysed.

2 Study Area

This study is based on two contrasting agro-ecological landscapes in Central Asia. They are located between the Amu-Darya and Syr-Darya Rivers. They are characterized by vast agricultural systems, which were extensively developed under the aegis of the former Soviet Union during the second half of the 20th century. Each test site is 30 km × 30 km.

The first site is located in the Khorezm region (KHO) in the north-western part of Uzbekistan. The agricultural landscape appears heterogeneous due to a comparatively high diversity of crops, e.g. cotton, rice, sorghum, maize, winter wheat, and fruit trees. Cover fraction (C_f, the fraction of the sites covered by agricultural fields) is high (Tab. 1). Multiple cropping is sometimes practiced, e.g. growing sequentially two or more crops in the same field within a single growing season, typically starting with wheat and following with another crop. In this study, such land use type will be labelled: "wheat-other".

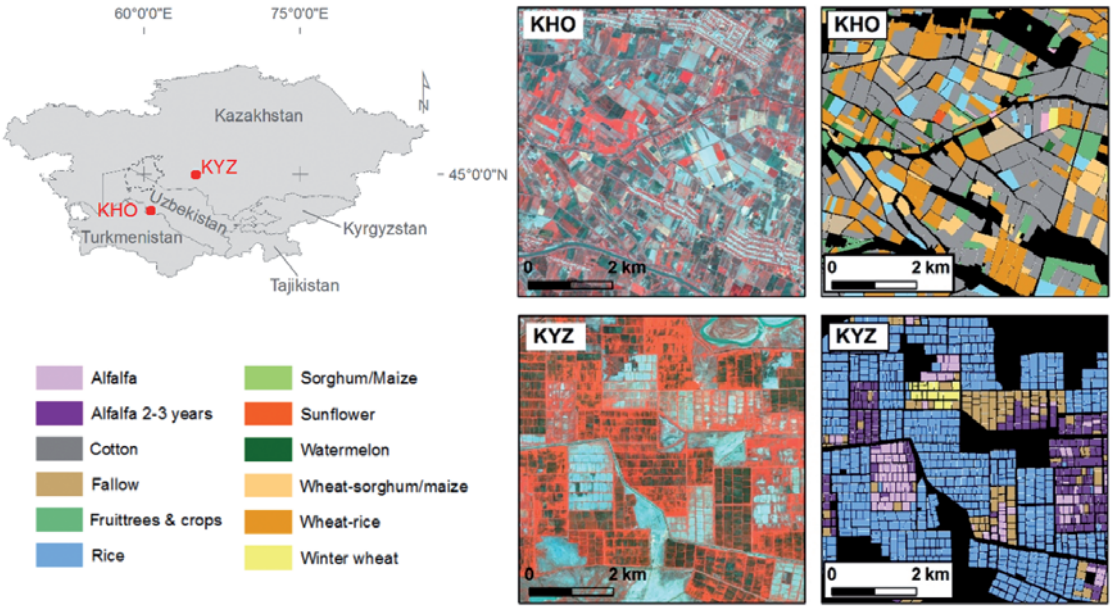

Fig. 1: Subsets (6.5 km × 6.5 km) of the satellite imagery and crop masks illustrating the typical cropping patterns within the two test sites Khorezm (KHO) in Uzbekistan and Kyzyl-Orda (KYZ) in Kazakhstan. The imagery is displayed using a near-infrared-green-blue band combination of the RapidEye sensor recorded in May.

Tab. 1: Characteristics of the four study sites. Total number of fields, field sizes and cover fractions C_f are based on own calculations based on segmented image objects.

Study site	Scene centre [Lat/Lon]	Total number of fields	Mean, median, and standard deviation of field size (ha)	C_f
KHO	60.69°E, 41.53°N	22,247	4.31, 3.21, 2.07	0.59
KYZ	64.55°E, 44.58°N	14,561	2.45, 2.14, 1.62	0.25

The second site is located in Kyzl-Orda (KYZ) in southern Kazakhstan, and was chosen to have an example with more regularly shaped field structures. Only few crops are dominating the agricultural landscape: rice and alfalfa. Large and regular shaped agricultural fields of approx. 2 ha – 3 ha each characterize this landscape, where the same crop is often grown on adjacent fields, that are aggregated to blocks which together exceed the area of between 500 m × 500 m and 1,000 m × 1,000 m (25 ha – 100 ha). Due to this pattern, the agricultural landscape in KYZ with smaller fields (2.45 ha) appears more homogeneous than in KHO with larger mean field sizes (4.31 ha).

3 Data and Methods

Images from the RapidEye mission with a ground sampling distance (GSD) of 6.5 m, were available for each site in 2011 (KYZ) and 2010 (KHO) (Fig. 2). These images have five spectral bands: blue (440 nm – 510 nm), green (520 nm – 590 nm), red (630 nm – 685 nm), red edge (690 nm – 730 nm), and near infrared (NIR, 760 nm – 850 nm). Images were atmospherically corrected using the ATCOR-2 module (Richter 2011), and geometrically corrected and co-registered with ground control points, resulting in RMSEs of < 6.5 m.

Crop specific masks are necessary to identify the target objects, i.e. agricultural fields cultivated with a certain crop, in the scene, and later for calculating the purity of coarser pixels with regard to specific crops. For the study sites access to vector databases of the agricultural fields including information on crops was either non-existent or restricted. Crop masks for the two sites were created according to the methodology described in Löw et al. (2012). These masks were created using supervised object-based image classification applied to a set of high-resolution time series of RapidEye images acquired over the growing seasons. The overall accuracies of the crop masks were more than reasonable (> 93%) and assumed to have a negligible error for the purpose of this study. Sorghum and maize in KHO were merged into the class "sorghum/maize" because they could not be distinguished from each other (Fig. 1).

3.1 *Simulating Coarser Images*

The methodology employed here is based on the same conceptual framework designed in a previous study for determining pixel size requirements for crop growth monitoring (Duveiller & Defourny 2010) that was extended by Löw & Duveiller (2014) for an application to crop classification. It relies on using high spatial resolution images and corresponding crop specific masks to generate various sets of pixel populations over which a classification step can be applied. The pixel populations are characterized by increasingly coarser pixel sizes and with a range of different crop specific purity thresholds. To simulate coarser pixel sizes, a spatial response model is convolved over the original RapidEye images. It consists of a point spread function (PSF) that characterizes both optical (PSF_{opt}) and detector (PSF_{det}) components of a generic sensor:

$$PSF_{net} = PSF_{opt} * PSF_{det}\,, \tag{1}$$

$$PSF_{opt}(x,y) = exp\left(-\frac{x^2+y^2}{2*(w*\sigma)^2}\right), \tag{2}$$

$$PSF_{det(x,y)} = rect(x-w) * rect(y-w). \tag{3}$$

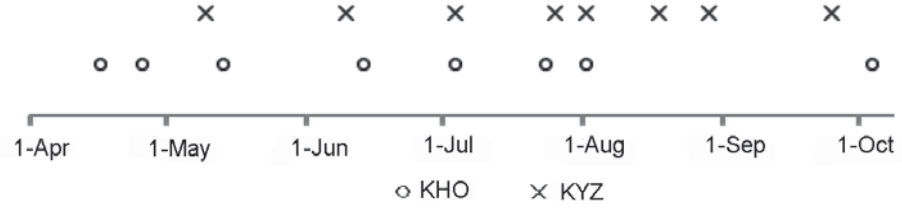

Fig. 2: Acquisition dates of the datasets from the RapidEye instrument utilised in this study.

where x and y are the cross-track and in-track coordinates, respectively, in the image space with their origin at the centroid of the ground instantaneous field of view (GIFOV), w the width of the optics, and σ the standard deviation of the Gaussian curve. Note that the width of the detector in both in-track and cross-track directions, respectively, is assumed to be equal. *rect* is the rectangular function, a uniform square pulse function with amplitude one and width w. The simulation was done in increments of 6.5 m, in order to simulate a continuum of coarser images, with pixel sizes ranging from 6.5 m – 747.5 m. More information on the simulation of coarser data in this framework can be found in Duveiller & Defourny (2010).

As mentioned in the introduction this study focuses on detecting and discriminating agricultural fields of specific crops within successively coarser pixels. In order to select pixels covering arable land and to further assess the effect of pixel purity on crop classification accuracy, the convolution of the same spatial response model over the high-resolution crop masks was performed. This produced crop specific "purity maps" at each scale, which map the pixel purity with respect to the spatial structures represented in the high resolution crop masks (Duveiller & Defourny 2010). This allows controlling the degree at which the footprints of coarser pixels coincide with the target structures, e.g. fields belonging to certain crops. At each spatial resolution, pixel populations can be selected based on thresholds on the pixel purity, here denoted π. A threshold can be chosen to separate the aggregated binary crop masks into two sets: target pixels and non-target pixels. The threshold can vary from 0, where all pixels in the images are selected as target, to 1, where only completely pure pixels, e.g. pixels lying completely within agricultural fields are selected. The results are the sets of selected target pixels, or "pixel populations", defined by their pixel size (v) and by the minimum acceptable purity threshold that defines them (π).

3.2 *Image Classification*

The second step consists in applying supervised classification procedures to the pixel populations selected in the previous step. The implementation of Breiman's RF (Breiman 2001) within the randomForest package (Liaw 2013) in the R programming environment was used for the classification in this study. The number of trees in the ensemble was set to a relatively high value of 500 so that the out-of-bag error (OOB) converges. OOB is calculated based on roughly 1/3 of the reference data, which are withheld from tree construction and used to calculate an error matrix and unbiased estimate of accuracy (Liaw & Wiener 2002). The remainder 2/3 of the reference data is used to build each tree. The number of features at each split node was set to the square root of the total number of input features $\sqrt{f}$ where f is the number of predictor variables within the corresponding input dataset. The NDVI and EVI plus RapidEye bands were the input to the classification, calculated for eight acquisition dates (f = 56). At each spatial scale and for each pixel purity threshold (here from 0 to 1 in increments of 0.05), independent training and testing datasets were generated following an equalized random sampling design to obtain approximately the same number of pixels for each class. The target size of both the training and testing sets was initially set to 400 randomly selected pixels per class, the minimum number per class required was set to 20, e.g. when coarser pixels were selected. RF was trained and applied to the entire time series data at each spatial scale, and all classes present in the corresponding study sites were included in the legend. To enhance the reliability of the experiments the random draws of training and validation data were repeated 10 times, and the classification performance estimates (see next section) were averaged over the 10 independent model runs.

3.3 *Characterizing Crop Identification Performance*

Pixel size and pixel purity can be considered as two dimensions of a $v - \pi$ space. For each selected pixel population in this $v - \pi$ space,

information regarding the classification performance, e.g. overall accuracy, can be calculated. Similar to the method proposed by Löw & Duveiller (2014) variables describing crop classification performance are calculated for each pixel population:

3.3.1 Quantifying classification performance

A set of confusion matrices (Congalton 1991) was computed on the hard result of the test sets defined along the $\pi - v$ dimensions. The overall accuracy parameter (ACC) is defined as the total proportion of correctly classified test pixels per total number of test pixels:

$$ACC = \frac{n_c}{n} \tag{4}$$

where n is the number of test samples, and n_c the number of correctly allocated test samples. As class-wise accuracy metric the F_β-measure of van Rijsbergen (1979) was employed. This measure combines the precision pr_i (which gives the proportion of samples, which truly have class i among all samples that were classified as class i) and the recall tp_i (the TPR which gives the proportion of samples classified into class i among all samples which truly have class i). The former determines the error of omission (false exclusion), the latter the error of commission (false inclusion). The traditional F_β-measure equally weights precision and recall ($\beta = 1$) and is sometimes referred to as F_1 measure:

$$CA_i = (1+\beta^2)\frac{pr_i * tp_i}{\beta^2 * pr_i + tp_i} \tag{5}$$

Measures of classification uncertainty like entropy assess the spatial variation of the classification quality on a per-case, e.g. per-pixel, basis, and can be used to supplement the global summary provided by standard accuracy statements like overall accuracy (Foody 2002). It can be characterized as a quantitative measure of doubt when a final classification decision is made. Beneath the final ("hard") class label, non-parametric algorithms such as support vector machines or RF can generate for each classified case x (agricultural field or pixel) a "soft" output in form of a vector $p(x) = (p_1, ..., p_i, ..., p_n)$ that contains the probabilities that a pixel is classified into a class i, n being the total number of classes (Löw et al. 2013). Each of the elements in $p(x)$ can be interpreted as a degree of belief or posterior probability that a pixel actually belongs to i. From this vector, the α-quadratic entropy (Pal & Bezdek 1994) for a given pixel (x) can be calculated as a measure of uncertainty, which is defined as:

$$\overline{\mathrm{AQE}}(x) = \frac{1}{n * (2^{-2\propto})} * \sum_{i=1}^{n} p_i^{\propto}(1-p_i)^{\propto} \tag{6}$$

Where p_i is one element in $p(x)$, n the number of classes, and α an exponent that determines the behaviour of AQE(x). The entropy of the total classified pixel population can be quantified with the median of all classified pixels $\overline{\mathrm{AQE}}(x)$, denoted AQE. This can also be done at the per-class basis, by calculating the median entropy of all pixels classified into a class i, denoted AQE_i.

The number of available reference pixels N_i of a given class i represents the total available size of pixel populations in the $v - \pi$ dimensions that can be used for training and testing the classifier. In supervised crop classification a minimum number of pixels per crop class can be desirable to assure the generalizability of the classifier model to the unseen dataset, and to reduce the influence of (random) variability in the training data on the classification result.

3.3.2 Determining suitable pixel populations

The final step to determine the suitable pixel populations for crop classification is to isolate the (v, π) combinations for which the classification performance fulfils certain criteria. This is accomplished by defining acceptable thresholds for the variables defined above. Such thresholds will be used to define a frontier in this $v - \pi$ space dividing pixel populations that are above or below the acceptable threshold for a given surface. As an example, if an application requires a minimum class-wise accuracy of 80%, the surface CA_i is sliced by a plane passing by the value $CA_i = 0.80$ (see Fig. 3, left images). When the inter-

section of CA_i and the plane is projected onto the 2-D space $v - \pi$, it separates this domain into the region where selected pixel populations have classification accuracy higher than 75% and the region where the accuracy of the remaining population will be lower than 75% (Fig. 3, right image). By drawing limits on the different parameters, the parameter surfaces were sliced and the intersection points of these slices in $v - \pi$ space were used to identify the position of the coarsest acceptable pixel sizes (v_{max}) and the corresponding minimum required pixel purities π respectively (Fig. 3, right image). Accordingly, pixel population suitability is defined by the number of fulfilled criteria, e.g. highest suitability means that all criteria were fulfilled. In this study, thresholds were defined as follows: $ACC/CA_i > 0.75$, $AQE/AQE_i < 0.50$, and $N_i > 100$.

3.3.3 Selecting categorical scale

Three levels of aggregation were created by grouping or discarding certain crop classes, hereafter called level I, II, and III (Fig. 4). The rationale for defining the class legends in the aggregation levels was (i) to group crop classes with similar spectral and temporal NDVI

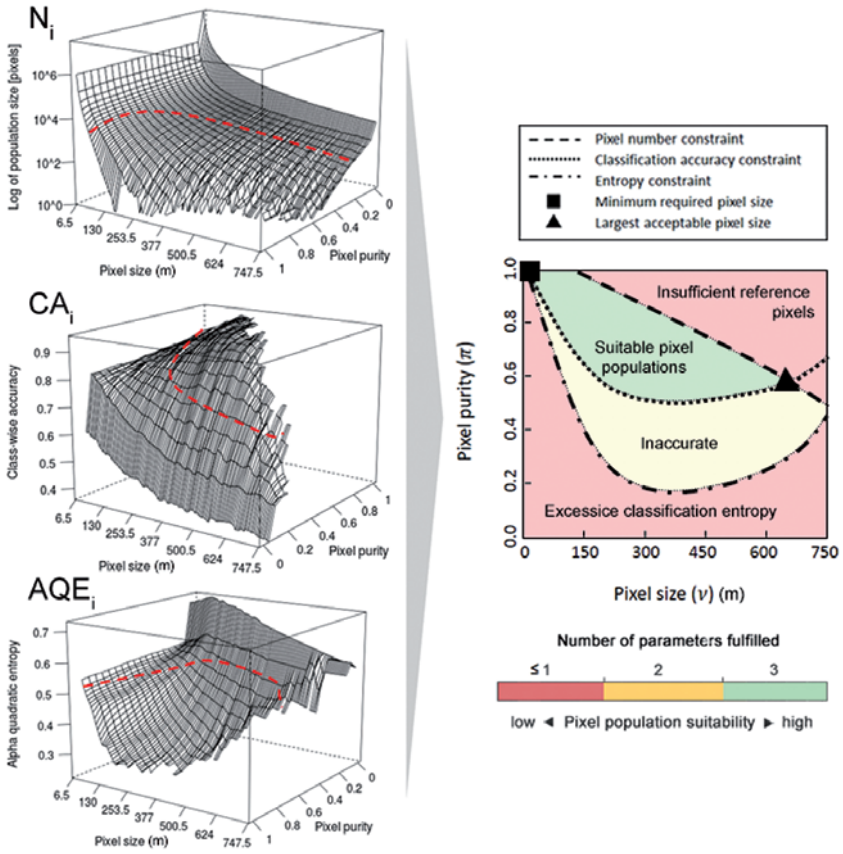

Fig. 3: Left column: Schematic examples of parameters chosen for crop identification for the pixel populations along the pixel size – pixel purity dimensions. Red dashed line indicates were predefined thresholds slice the parameter surfaces and separate $v - \pi$ space into two domains: one that fulfils a certain threshold and a second one that does not. Note that the pixel purity axis is inverted for N_i and AQE. Right column: Suitability map with theoretical boundaries in $v - \pi$ space used to define the requirements for pixel populations to be used for supervised classification. Circle indicates the position of maximum tolerable pixel size v_{max}, black filled square the minimum required pixel size v_{min}.

signatures and (ii) to discard classes that cover only a small fraction of the landscape (C_f < 0.01) and that are spectrally too distinct from other classes to be merged. Crop rotation classes in KHO ("Wheat-Sorghum / Maize" and "Wheat-rice") were merged in level II due to the similarity of the temporal NDVI profiles. Further, minor classes (C_f < 0.01) were merged to one class "mixed crops", due to their spectral similarity. Likewise, "Alfalfa 1y" and "Alfalfa 3y" were merged in level II in KYZ due to their spectral similarity, reducing the number of classes from five to three. Winter wheat was discarded in level II due to its marginal cover fraction in both sites (C_f < 0.01). Fallow fields and mixed crops in KHO, which together cover less than 2% of the landscape were completely removed from the class legend in level III, reducing the number of classes from eleven to six. In this level a binary class legend was established in KYZ, i.e. active, e.g. rice, vs. unused, e.g. fallow, alfalfa, fields. The pixel population suitability was then determined separately at each aggregation level, according to the methodology described in the previous section. The pixel population suitability was then determined separately at each aggregation level, according to the methodology described in the previous section.

4 Results and Discussion

As expected, the suitability of pixel populations for classifying crops varied in the $v - \pi$ space and the suitability was enhanced when selecting higher aggregation levels (Fig. 5) as was demonstrated in previous studies (e.g. Marceau et al. 1994b). Suitable pixel populations were restricted to rather small pixel sizes at level I, with v_{max} = 162.5 m in KYZ, similar to KHO (v_{max} = 149.5 m). In KYZ aggregating crop classes in level II resulted in enhancing the suitability of pixel populations in the $v - \pi$ space, in particular coarser pixels could be tolerated (v_{max} = 747.5 m at level III) and the minimum purity requirements could be relaxed for v_{max} (the corresponding π at level I was 0.60, and decreased to 0.45 for v_{max} at level III). Likewise, in KHO aggregating classes enhanced the suitability of pixel populations for crop identification, but compared with KYZ the use of coarser pixels was limited, most obvious because of the spatial pattern of fields in the KHO landscape, e.g. more crop classes in a more heterogeneous landscape where the use of coarser pixels resulted in a higher degree of pixel mixing, which resulted in having generally higher purity requirements than in KYZ. This could be explained by the limited availability of purer reference pixels in this landscape. Further, tolerating some signal contamination may be beneficial

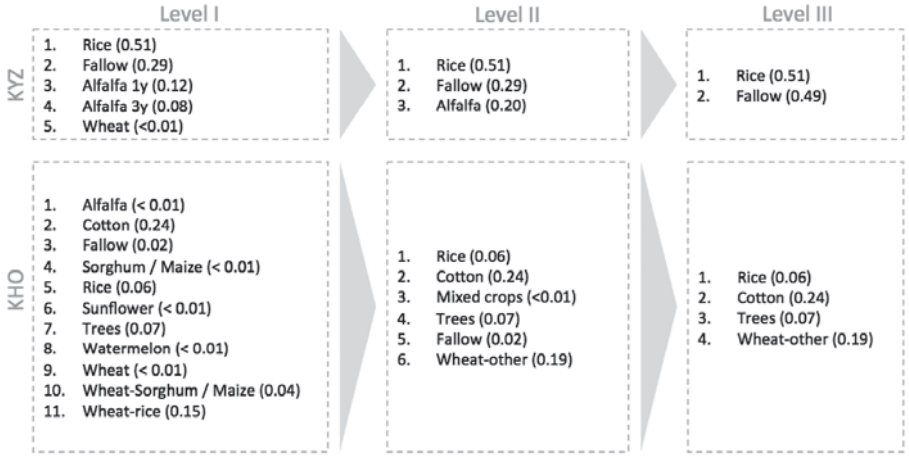

Fig. 4: Class legends according to three levels of aggregation in the two study sites. Cover fractions (C_f), i.e. the fraction of the sites covered by agricultural fields of a certain crop class, are given in brackets.

in the case of crop identification in KHO, e.g. a larger sample size for classification training including mixed pixels may better represent the diversity of the spectral response of the target class within this landscape (Foody 1996).

For specific classes the required conditions to achieve highest CA_i varied among the three aggregation levels (Fig. 6). The corresponding pixel sizes tended to become coarser when increasing the aggregation level. An explanation why coarser pixel sizes generally achieve higher accuracy could be the interplay of increasing error-rates of smaller but purer pixels, which become more abundant when pixels become smaller, caused by increasing within-class variability (Hsieh et al. 2001) and decreasing error of mixed pixels, which become less abundant when pixels become smaller. The within-class variability might in particular become an important issue when such heterogeneous crop classes like alfalfa and other fallow fields are merged, e.g. in KYZ. In such a situation it might be better to have coarser pixels (McCloy & Bøcher 2007), thereby reducing this variance and counterbalancing the effect of pure-pixel heterogeneity within smaller pixels. After merging quite diverse crop classes in KYZ in level III (two types of alfalfa fields and other fallow land to one class "fallow") the condition for best CA_i was pushed to 604.5 m, compared to 195.0 m at level I. In this context, image segmentation should be considered (Blaschke 2010, Yan & Roy 2014). Image segmentation of high spatial resolution images or time series results in image-objects that minimize the variance but that are not constrained by the rectangular nature of the pixels. Analysing the optimal size of multi-date image objects for crop identification could be an interesting extension of the proposed conceptual framework. Another advantage of object-based image analysis (OBIA) is that different segment sizes for specific land use types can be used and analyzed. Image-objects could also be analysed with respect to their homogeneity at different spatial scales using a concept analogous to pixel purity. However, such questions are beyond the scope of this current paper.

For fallow fields in KYZ the pixel sizes for which maximum CA_i was archived were in the same order at level I and II (195.0 m and 208.0 m), but coarser pixel sizes were required at level III (604.5 m), which could be attributed to the spatial aggregation pattern of fields

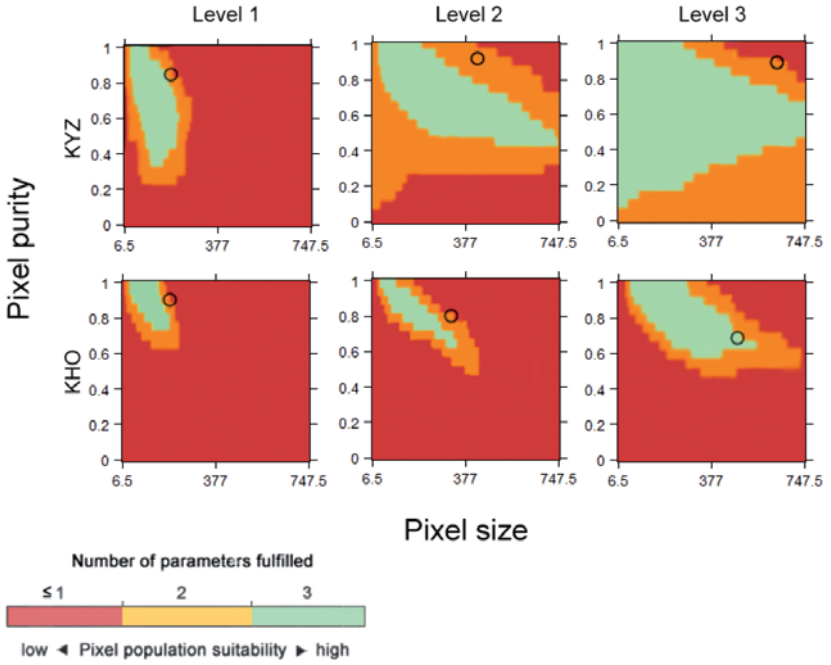

Fig. 5: Suitability maps for three different class aggregation levels for all classes. Circles indicate position of maximum achievable ACC in $v - \pi$ space.

in this landscape, i.e. large blocks of fields with the same crop class "fallow" in level III. In KHO pixel populations suitable for classifying rice and fallow fields were characterized by higher pixel purity requirements than in KYZ. This can be explained by the spatial pattern or field fields which are more spatially dispersed than in KYZ (Fig. 1), resulting in mixed signal once the purity of the (coarser) pixels becomes too low, e.g. $\pi < 0.7$. Likewise, the maximum possible CA_i for rice fields in KHO was achieved with smaller pixels than in KYZ.

The high ACC in KYZ when selecting level II or III (Fig. 7) offers positive prospects for using images from existing satellite missions (Landsat 30 m, MODIS 250 m / 500 m, NPP-VIIRS 747.5 m) used for crop mapping. Yet, results also highlight the need to consider that the class legend must be selected properly and separately adapted in different agricultural landscapes. The lowest ACC and CA_i of rice

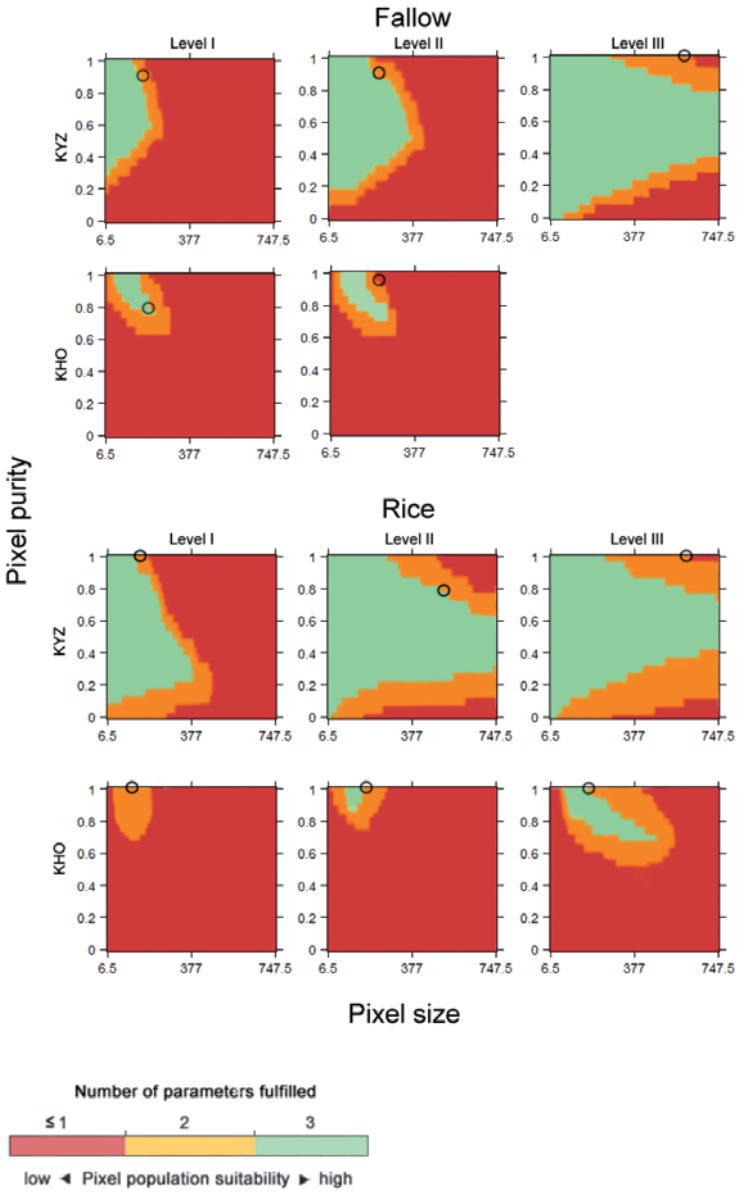

Fig. 6: Suitability maps for three different class aggregation levels and for class fallow (top rows) and fallow (bottom rows). Ovals indicate position of maximum achievable CA_i in $\nu - \pi$ space.

and fallow fields were achieved when selecting coarser pixels (745.5 m) at level I. The smaller the pixel sizes the higher was ACC at this level. Further, the higher the aggregation level the smaller was the absolute differences between ACC, e.g. at 0.043 at level III but 0.791 and 0.103 in level I and II, respectively. CA_i of fallow fields in KYZ increased when shifting from level II to III, whilst the difference in CA_i for rice was within 0.01, i.e. for 32.5 m and 247.5 m. In KHO using coarser pixels (247.5 m and 747.5 m) resulted in a clear drop in CA_i for rice fields, for which smaller pixels gave better results due to the spatial pattern of fields (see above).

Overall, the results indicate that there was no unique spatial resolution for identifying and discriminating all classes at once at a given aggregation level, confirming previous studies (Marceau et al. 1994b, Löw & Duveiller 2014). Within a particular aggregation level, some classes are better classified at fine spatial resolutions, while others require coarser spatial resolutions. Further, the suitability of pixel populations for identifying specific crops differed within the landscapes, and for a given crop there were large differences among the landscapes. Of course, the results are only valid within the parameterization chosen for this study, and the user might select other metrics more appropriate to the targeted application which is not limited to crop discrimination. The parameters in this study were purposefully selected because different metrics evaluate different components of accuracy as they are based on different statistical assumptions on the input data. Consequently, seeking to optimize classifier algorithm performance or defining suitable pixel sizes with only one metric may have led to a non-optimal result when viewed from another point of view or quantified with a different metric that is sensitive to different features concerning accuracy (Foody 2002, Provost & Fawcett 1997).

Landscape heterogeneity, e.g. the size of agricultural fields and the properties of their neighbourhood, were shown to be important factors determining classification accuracy. When the crops were grown on larger fields, or when the cover fraction was high, coarser pixel sizes could be tolerated for crop identification. Crops grown on fields dispersed over

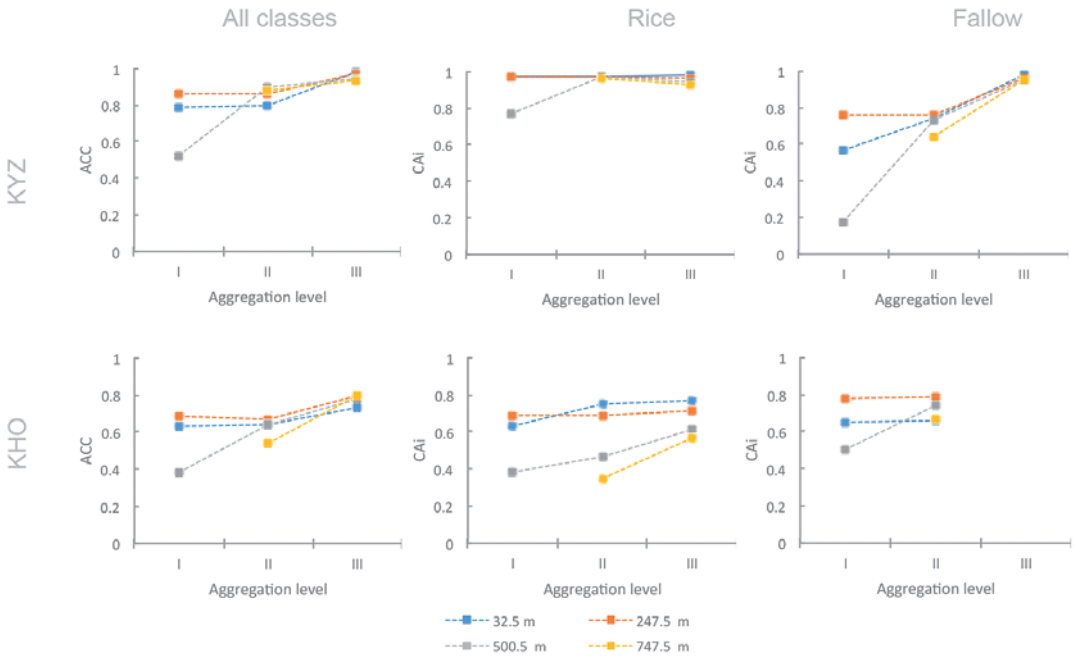

Fig. 7: Maximum achievable overall and class-wise accuracies (ACC and CA_i) for selected pixel sizes at different aggregation levels. Note that in KHO the class "fallow" was omitted in aggregation level III. At aggregation level I the number of reference pixels in KHO and KYZ dropped below 20 for at least one class at 747.5 m ($N_i < 20$), which forced the experiments to stop.

the landscape like rice in KHO could only be detected using smaller pixel sizes, and only using relatively pure pixels. However, the crop classes displayed differences in the specific nature of these relationships, and the landscape heterogeneity with respect to the spatial pattern also influenced the choice of pixel sizes. For instance, while the median field sizes in KYZ and KHO are comparable, the farmer's fields are more regular in shape, less variable in size, and the same crops are found on blocks of fields that together can aggregate to more than 100 ha in size. Due to this spatial aggregation pattern, it is easier to have coarser pixels fall within target fields and thus conferring higher acceptable pixel sizes for crop identification, resulting in notably higher values for v_{max} in KYZ than in KHO.

The results confirm previous studies that found large differences in accuracy depending on the degree of pixel mixing, e.g. Foody (1996). Therefore, this study suggests for a more spatially explicit assessment of accuracy. The framework could be used to plot the suitability (or accuracy) of each pixel covering arable land as a function of its purity. This is similar to spatial assessments of classification uncertainty (Löw et al. 2013) and could be of interest when coarser satellite sensors like MODIS or Sentinel-3 are to be used for crop mapping. For instance, one practical utility of this framework could be "masking out" unsuitable pixel populations, according to the user's specific needs, before applying them in agricultural crop-specific growth modelling. Further, knowledge on the spatial distribution of map quality, e.g. defined as pixel suitability, allows for a better interpretation of the results of agricultural model outputs.

5 Conclusions

The overall methodology presented in this study was used to assess the impact of categorical and spatial scale (pixel size) on crop classification accuracy. Coarser satellite images were simulated, based on RapidEye data, and classified using the RF algorithm. Different class legends (aggregation levels) were tested, which were created based on merging spectrally similar classes or by discarding classes that only covered small fractions of the landscapes. The results show that there was no unique spatial resolution that provided the best classification result for all classes at once at a given aggregation level. Classification accuracy could be improved by aggregating certain crop classes. Further, the suitability of pixel populations for crop discrimination differed within the landscapes, and for a given crop there were differences among the landscapes. The results imply that classifications based on purer pixels were generally the most suitable (and most accurate) for crop type discrimination, but the number of such pure pixels might be limited in heterogeneous landscapes, where only a few pixels fall into larger fields. Although crops within mixed pixels could be more accurately distinguished when aggregating crop classes, a meaningful evaluation of classification accuracy still could only be achieved when accommodating the effect of pixel purity.

The results of this study suggest that it is important to consider the field size distribution and pattern when shifting between regions, and not every sensor might be equally suitable for a given application like crop mapping in a large region with different spatial pattern like in Central Asia or other regions worldwide. Neglecting the effect of pixel purity and aggregation level might produce haphazard results, which could negatively impact spatial modelling when crop maps are taken as input. In the context of agricultural crop growth monitoring the framework described above can be used to draw guidelines for selecting appropriate imagery, e.g. suitable pixel sizes, and for selecting appropriate class legends for accurate crop identification and crop type discrimination over a given agricultural landscape when the interest is only on a subset of the landscape, e.g. pixels covering arable land. In the context of the CAWa-project this framework is recently being implemented to define pixel size requirements and appropriate class legends for crop identification in the irrigated landscapes in Fergana Valley and Karakalpakstan (Uzbekistan). This is of relevance because agricultural production monitoring (yield, evapotranspiration) in this project is based on MODIS data, which delivers the required swath and high revisit frequen-

cy. Due to the need for crop specific masks at the MODIS scale, knowledge about the constraints of such coarse image data for crop identification is essential. The results of this study provide an opportunity to discuss the effects of pixel size and purity and the classification algorithm independent factors such as parcel size, spatial distribution of crop types and crop patterns on agricultural monitoring related applications. In a world with increasingly diverse geospatial data sources in terms of combinations of spatial and temporal resolutions, the tool can also help users to choose the different data sources that meet the requirements imposed by their applications.

Acknowledgements

This study was undertaken at the Department of Remote Sensing (University of Würzburg) within the CAWa project funded by the German Federal Foreign Office, and within the German-Uzbek Khorezm project. We thank the German Aerospace Center (DLR) for providing data from the RapidEye Science Archive (RESA). The German National Academic Foundation (Studienstiftung des deutschen Volkes) funded this research by way of a PhD grant to the first author. The authors also acknowledge the support of the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) during the field studies in Uzbekistan and Kazakhstan.

References

Aplin, P., 2006: On scales and dynamics in observing the environment. – International Journal of Remote Sensing **27** (11): 2123–2140.

Atkinson, P.M. & Aplin, P., 2004: Spatial variation in land cover and choice of spatial resolution for remote sensing. – International Journal of Remote Sensing **25** (18): 3687–3702.

Blaschke, T., 2010: Object based image analysis for remote sensing. – ISPRS Journal of Photogrammetry and Remote Sensing **65** (1): 2–16.

Breiman, L., 2001: Random forests. – Machine Learning **45** (1): 5–32.

Congalton, R.G., 1991: A review of assessing the accuracy of classifications of remotely sensed data. – Remote Sensing of Environment **37** (1): 35–46.

Cushnie, J.L., 1987: The interactive effect of spatial resolution and degree of internal variability within land-cover types on classification accuracies. – International Journal of Remote Sensing **8** (1): 15–29.

Duveiller, G. & Defourny, P., 2010: A conceptual framework to define the spatial resolution requirements for agricultural monitoring using remote sensing. – Remote Sensing of Environment **114** (11): 2637–2650.

Foody, G.M., 1996: Incorporating mixed pixels in the training, allocation and testing stages of supervised classifications. – Pattern Recognition Letters **17** (13): 1389–1398.

Foody, G.M., 2002: Status of land cover classification accuracy assessment. – Remote Sensing of Environment **80** (1): 185–201.

Hsieh, P., Lee, L.C. & Chen, N., 2001: Effect of spatial resolution on classification errors of pure and mixed pixels in remote sensing. – IEEE Transactions on Geoscience and Remote Sensing **39** (12): 2657–2663.

Ju, J., Gopal, S. & Kolaczyk, E.D., 2005: On the choice of spatial and categorical scale in remote sensing land cover classification. – Remote Sensing of Environment **96** (1): 62–77.

Levin, S.A., 1992: The problem of pattern and scale in ecology. – Ecology **73** (6): 1943–1967.

Liaw, A. & Wiener, M., 2002: Classification and Regression by randomForest. – R News **2** (3): 18–22

Löw, F., Schorcht, G., Michel, U., Dech, S. & Conrad, C., 2012: Per-field crop classification in irrigated agricultural regions in Middle Asia using random forest and support vector machine ensemble. – Proceedings of SPIE 8538, Edinburgh, United Kingdom.

Löw, F. & Duveiller, G., 2014: Defining the spatial resolution requirements for crop identification using optical remote sensing. – Remote Sensing **6** (9): 9034–9063.

Löw, F., Michel, U., Dech, S. & Conrad, C., 2013: Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using Support Vector Machines. – ISPRS Journal of Photogrammetry and Remote Sensing **85:** 102–119.

Marceau, D.J., Howarth, P.J. & Gratton, D.J., 1994a: Remote sensing and the measurement of geographical entities in a forested environment. 1: The scale and spatial aggregation problem. – Remote Sensing of Environment **49** (2): 93–104.

Marceau, D.J., Gratton, D., Fournier, R. & Fortin, J., 1994b: Remote sensing and the measurement of geographical entities in a forested environment. 2. – The optimal spatial resolution. Remote Sensing of Environment **49** (2): 105–117.

McCloy, K. & Bøcher, P., 2007: Optimizing image resolution to maximize the accuracy of hard classification. – Photogrammetric Engineering and Remote Sensing **73** (8): 893–903.

Pal, N.R. & Bezdek, J.C., 1994: Measuring fuzzy uncertainty. – IEEE Transactions on Fuzzy Systems **2** (2): 107–118.

Provost, F. & Fawcett, T., 1997: Analysis and visualization of classifier performance: Comparison under imprecise class and cost distributions. – Third international conference on knowledge discovery and data mining. – AAAI Press, Menlo Park, CA, USA.

Richter, R., 2011: Atmospheric/Topographic Correction for Satellite Imagery. – ATCOR-2/3 User Guide 7.1.

Smith, J.H., Stehman, S.V., Wickham, J.D. & Yang, L., 2003: Effects of landscape characteristics on land-cover class accuracy. – Remote Sensing of Environment **84** (3): 342–349.

van Rijsbergen, C.J., 1979: Information Retrieval, 2nd ed., Butterworths. – Butterworth & Co Publishers Ltd., United Kingdom.

Woodcock, C.E. & Strahler, A.H., 1987: The factor of scale in remote sensing. – Remote Sensing of Environment **21** (3): 311–332.

Yan, L. & Roy, D.P., 2014: Automated crop field extraction from multi-temporal Web Enabled Landsat Data. – Remote Sensing of Environment **144:** 42–64.

Addresses of the Authors

Dipl. Geogr. Fabian Löw & Prof. Dr. Christopher Conrad, Department of Remote Sensing, University of Würzburg, Oßwald-Külpe-Weg 86, D-97074 Würzburg, e-mail: fabian.loew@uni-wuerzburg.de

Dr. Grégory Duveiller, European Commission Joint Research Centre (JRC), Via E. Fermi, 2749, I-21027 Ispra (VA), Italy

Prof. Dr. Ulrich Michel, Department of Geography, University of Education, Czernyring 22/11-12, D-69115 Heidelberg

Manuskript eingereicht: Juni 2014
Angenommen: September 2014