

Large-area settlement pattern recognition from Landsat-8 data

Marc Wieland^{1*}, Massimiliano Pittore¹

¹ *Centre for Early Warning Systems, GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany*

^{*} *Corresponding author: marc.wieland@zoo.ox.ac.uk
Present address of corresponding author: Department of Zoology, University of Oxford, South Parks Road, OX1 3PS Oxford, United Kingdom*

Abstract

The study presents an image processing and analysis pipeline that combines object-based image analysis with a Support Vector Machine to derive a multi-layered settlement product from Landsat-8 data over large areas. 43 image scenes are processed over large parts of Central Asia (Southern Kazakhstan, Kyrgyzstan, Tajikistan and Eastern Uzbekistan). The main tasks tackled by this work include built-up area identification, settlement type classification and urban structure types pattern recognition. Besides commonly used accuracy assessments of the resulting map products, thorough performance evaluations are carried out under varying conditions to tune algorithm parameters and assess their applicability for the given tasks. As part of this, several research questions are being addressed. In particular the influence of the improved spatial and spectral resolution of Landsat-8 on the SVM performance to identify built-up areas and urban structure types are evaluated. Also the influence of an extended feature space including digital elevation model features is tested for mountainous regions. Moreover, the spatial distribution of classification uncertainties is analyzed and compared to the heterogeneity of the building stock within the computational unit of the segments. The study concludes that the information content of Landsat-8 images is sufficient for the tested classification tasks and even detailed urban structures could be extracted with satisfying accuracy. Freely available ancillary settlement point location data could further improve the built-up area classification. Digital elevation features and pan-sharpening could, however, not significantly improve the classification results. The study highlights the importance of dynamically tuned classifier parameters, and underlines the use of Shannon entropy computed from the soft answers of the SVM as a valid measure of the spatial distribution of classification uncertainties.

Key words: Landsat-8; object-based image analysis; machine learning; settlements; Central Asia.

1 Introduction

According to the United Nations' Vancouver Declaration on Human Settlements (1976), a settlement is defined as the totality of the human community with all the structural, organizational, social and cultural elements that sustain it [1]. Commonly accepted definitions for the categorization of settlements are not existing, even though distinguishing different settlement types is recognized as being important [2]. Especially differencing between “urban” and “rural” settlements is a widely discussed topic and common approaches to categorize settlements include thresholding of the population size, density or the geographic area, the political, administrative and economic importance, the availability of services or the sphere of influence. Information on human settlements is crucial for a large range of applications including disaster risk reduction [3], [4] and rapid emergency response [5]. A central issue in this context is the availability of up-to-date information on the extent, type and composition of human settlements. In particular in less developed countries such information is largely unavailable, due to rapid urbanization that often cannot be adequately mapped by local or regional authorities.

It is largely recognized that satellite remote sensing can provide the needed information in a time and cost-efficient manner over large areas. Despite the apparent potentials of remote sensing, there exist only few global and some regional settlement mapping approaches that include information beyond a simple binary classification of built-up / not built-up areas [6]. A majority of the global approaches are based on low-resolution earth-observation data [7]. Due to their ground sampling distance (>100 m) they have the tendency to under-represent small, scattered rural settlements. Efforts towards global high-resolution settlement products such as the Global Human Settlement Layer [8] or the Global Urban Footprint [9] have been started recently to overcome these limitations. However, given the type of input data (very high resolution optical and SAR images), their time and cost-efficiency in producing regularly up-to-date settlement information may be rather limited. To this regard, the medium-resolution Landsat-8 satellite sensors show great potential for timely and cost-efficient settlement studies at regional scale [10], [11], [12]. Previous Landsat sensors have already been exploited by several authors to conduct regional and global land-cover / land-use studies [13], [14], [15], but to the authors' knowledge, no study exists so far that uses Landsat-8 data for settlement pattern recognition at the regional scale.

A wide range of studies exists on built-up area extraction from medium resolution multi-spectral satellite images [16]. Most of the studies use pixel-based approaches with unsupervised or supervised classification or regression. When further analyzing the built-environment from Landsat images, different built-up land-cover / land-use classes are mainly defined by the alignment of buildings, streets and open spaces and therefore cannot sufficiently be described by just the spectral values of a single pixel. In this context, object-based image analysis showed improvements over pixel-based approaches [17]. In a previous study of the authors, the potential of combining object-based image analysis with machine learning for urban pattern recognition could be highlighted [18].

For the stratification of a settlement into meaningful spatially defined entities of similar building types and morphological structure, “urban structure types” can provide a valuable concept. Banzhaf and Hofer (2008) [19] give a comprehensive overview of urban structure type studies, which shows their multiple applications in a variety of research fields. Classification schemes and analysis scales of urban structure types studies mainly depend on the particular area and structures of interest. Urban structure types mapping is usually carried out by means of a visual aerial or satellite image interpretation and field surveys. Taubenböck *et al.* (2008) [20] use the urban structure types concept within the context of

mega city mapping from satellite imagery. Herold *et al.* (2003) [21] use a combination of remote sensing and landscape metrics to describe structures in urban land-use.

The aim of this study is to derive a multi-layered settlement product from Landsat-8 data covering a large area of Central Asia (Southern Kazakhstan, Kyrgyzstan, Tajikistan and Eastern Uzbekistan). The main tasks tackled by this work include:

- Built-up area identification;
- Settlement type classification;
- Urban structure types pattern recognition.

The study follows an object-based approach to image analysis and combines it with machine learning techniques. Besides commonly used accuracy assessments of the resulting map products, thorough performance evaluations are carried out under varying conditions to tune algorithm parameters and assess their applicability for the given tasks. As part of this, several research questions are being addressed in particular:

- Influence of the feature space on the classification performance: Is the Landsat-8 multi-spectral feature space sufficient for the classification tasks at hand? Which influence plays texture in the context of human settlement pattern recognition? Can features derived from digital elevation models improve the identification of built-up areas in mountainous regions?
- Influence of image pre-processing on the classification performance: Can pan-sharpening improve the accuracy of urban structure type pattern recognition?
- Spatial distribution of classification uncertainties: How are the classification uncertainties distributed in space and are they correlated with segment heterogeneity?

The paper is structured as follows. In Section 2 the study area, datasets and pre-processing steps are introduced. Section 3 outlines the method and describes the modules of the processing and analysis chain. Results are presented in Sections 4 and 5 with a particular focus of Section 4 being on performance evaluation of the learning machines under varying conditions. A comprehensive validation with in-situ and OpenStreetMap data is presented, before the study closes with a discussion and conclusion Section.

2 Study area and data

The study area covers 606,185 km² spanning large parts of Central Asia, namely Southern Kazakhstan (S-KZ), Kyrgyzstan (KG), Tajikistan (TJ) and Eastern Uzbekistan (E-UZ) (Fig. 1). The landscape of the study area can be divided into the flat steppes of Kazakhstan in the North, desert areas mainly in Uzbekistan and the mountainous regions of Kyrgyzstan and Tajikistan with the major mountain ranges being Tien Shan in Kyrgyzstan and Pamir in Tajikistan. Along the margins of the Amur Darya and Syr Darya rivers which flow in North-West direction through Kyrgyzstan, Tajikistan, and Eastern Uzbekistan vast agricultural areas can be found. Currently approximately 40% of the population in Central Asia live in urban areas while economies remain mostly agrarian-industrial. According to the Population Division of the UN Department of Economic and Social Affairs, the population living in urban areas will reach 45% by 2030 and exceed 55% by 2050 [20]. Despite a rapidly changing built-environment, especially since the end of the Soviet-rule in 1990, no unified map of up-to-date settlement extents and compositions existed for the region at the time of writing this manuscript.

43 multi-spectral image scenes from Landsat-8 with a ground coverage of 185 x 180 km each were acquired over the study area between April and June 2013. The Landsat Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS) systems on board the satellite have a ground sampling distance of 30 m in the visible, near-infrared, short wave infrared and 100 m respectively in the thermal infrared with a spectral resolution of 10 bands plus an additional panchromatic band with 15 m spatial resolution. The images are delivered in processing Level 1 T-Terrain Corrected, which means they are orthorectified, radiometrically corrected and the TIRS bands are resampled to 30 m by using cubic convolution. A transformation of the images into the WGS84 coordinate reference system was performed. The images were converted to top-of-atmosphere reflectance, mosaicked and split into 494 regular tiles of 1,500 x 1,500 px. A Bayesian data fusion as implemented in OTB [23] and proposed by Fasbender et al. (2008) [24] has been performed to pan-sharpen image tiles that intersect with reference datasets (see below for further explanation) in order to test the influence of the spatial resolution on the performance of classifying urban structures (Section 4.2). Moreover, the Shuttle Radar Topography Mission (SRTM) digital elevation model version 4 with 90m horizontal resolution has been used for the area covered by these test image tiles in order to further evaluate the influence of features derived from elevation data on the classification performance.

3 Method

A processing pipeline has been developed, entirely based on free and open-source tools, in order to extract settlement patterns from multi-spectral satellite images (Fig. 2). The proposed approach combines object-based image analysis with statistical learning. Four main processing modules can be distinguished, namely image preprocessing (Section 2), segmentation (Section 3.1), feature-based description (Section 3.2) and classification (Section 3.3). The classifier performance evaluation is presented as part of the results section (Section 4). In this study, the image processing and analysis chain is trained and tested for the tasks of built-up area extraction and urban structure types detection from Landsat-8 data. However, it aims at being flexible enough to be applicable also to other tasks and image types such as building footprint extraction from very high-resolution satellite images [18]. The system architecture is largely based on Python and custom C++ routines with a PostgreSQL/PostGIS database back-end that stores the image tiles and manages the input and output data. In the following the main processing and analysis modules are explained in more detail.

3.1 Segmentation

An efficient graph-based image segmentation algorithm [25] is used to cluster the original image pixels into segments. The choice of the algorithm is based on experience from previous work of the authors [26]. The algorithm compares the brightness values of the image pixels in all spectral bands of the satellite image between neighbouring pixels within each segment with those across the boundary in order to measure the evidence for a boundary between two segments. As distance measure the L2 Euclidean distance is used. A function of the segment size controlled by a scale parameter k and a merge parameter m allow to effectively set a scale of analysis, where larger values cause a preference for larger segments. Segmentation parameters are tuned in this study by a comparison of intra-segment homogeneity and inter-segment heterogeneity [27] with global Moran's I being used to quantify inter-segment heterogeneity and weighted standard deviation to assess intra-segment homogeneity [28]. In a set of segmentations, the segmentation that minimizes intra-segment variability and maximizes inter-segment separability is defined as being optimal in terms of providing the most appropriate analysis scale for a given image scene.

3.2 Feature-based description

An extended set of 157 image features were calculated at the computational unit of segments and considering all 10 spectral bands of Landsat-8. The features can be grouped into (a) spectral features (51), (b) spectral band indices (18), (c) textural features derived from grey-level co-occurrence matrix (GLCM) [27] (80) and (d) digital elevation model features (8). A summary list of the features is provided in Tab. 1. The feature values were normalized to zero mean and unit variance to account for the requirements of the distance-based SVM classifier (Section 3.3). To reduce the dimensionality of the feature space and preselect the most significant features for a specific classification task, dataset and classifier, a recursive feature selection algorithm as proposed by Guyon *et al.* (2002) [30] is used in this study. During the training stage, SVM themselves already select features and ignore irrelevant or redundant ones. Given a SVM that assigns weights to features, recursive feature selection considers iteratively smaller and smaller sets of features. With each iteration a feature set is used to train the SVM and assign weights according to their discriminating power. The features with the smallest weights are eliminated from the feature set for the next iteration. The feature selection has been performed in a five-fold cross-validation loop using classification accuracy as scoring value. The feature space that maximizes the scoring value is selected.

3.3 Classification

Support Vector Machines (SVM) has been selected as the most promising classifier to be used for the tasks of built-up area and urban structure type pattern recognition. The classifier choice is based on a previous study by the authors that assessed the performance of different learning machines to distinguish built-up from not built-up areas [18]. SVM is a non-parametric binary classifier which exploits non-linear kernel functions to project the feature vectors describing the instances being classified into a multi-dimensional feature space, where the two classes can be separated by a suitable hyper plane. SVM compute the optimal separating hyper plane between two classes by maximizing the margin between the hyper plane itself and the closest feature vectors (hence called support vectors). Extension to multi-class application are usually based on a one-against-one tournament scheme. In this study we use the scikit-learn [31] implementation based on the well-known libsvm software library [32]. SVM parameters (kernel function Φ , kernel coefficient γ and penalty or regularization parameter C) are tuned for each classification task by using a grid-search in a five-fold cross-validation during the training phase of the classifier. An optimal parameter selection is reached when the cross-validation estimate of the test samples error is minimal. Once the SVM classifier has been trained, the actual labelling of previously unseen instances is easily obtained (by a set of scalar products), and thus can be efficiently performed with a batch processing scheme.

In order to get an estimate of the spatial variability of the classification uncertainties, the Shannon entropy [33] has been computed from the soft answers of the SVM. The standard formulation of SVM does not provide class membership probability estimates which are needed to calculate entropy. However, the probabilities can be calibrated through logistic regression on the SVM scores, fit by an additional cross-validation on the training data as is described in Platt (1999) [34]. For each sample x the class probability is computed from its distances to the optimal separating hyper planes for each of the $n(n - 1) / 2$ binary SVMs by fitting a sigmoid function to the decision values of each of the binary classifiers. The probabilistic output of the binary classifiers is then combined as proposed in Wu *et al.* (2004) [35]. The result is a vector containing the estimated class memberships associated with the sample defined as

$$pk(x) = \{pk_1(x), pk_2(x), \dots, pk_i(x), \dots, pk_n(x)\} \quad (1)$$

where $pk_i(x)$ is the estimated membership degree of x to class i , and n is the number of classes. From the probability vector $pk(x)$ the Shannon entropy can be calculated as follows.

$$H(x) = - \sum_{i=1}^n pk_i \log_2 pk_i \quad (2)$$

A flexible classification scheme (Tab. 2) has been designed that consists of three hierarchical levels [36]. The first level is used for the extraction of built-up areas. In order to get a more differentiated picture of the built environment, the class “built-up area” is further refined and split into more detailed classes “residential” and “industrial/commercial”. The classes in the first two levels are fixed and general enough to be applied in other geographical contexts worldwide. The selection of level 3 classes needs to be adjusted for the specific study area based on local expert knowledge and existing urban patterns, as is done in this study for Central Asia based on previous work of the authors [37].

3.4 Generation of training and testing sets

Reference datasets for two different classification tasks, namely the identification of built-up areas and the further stratification of these into urban structure types, have been derived. An overview of the training and testing datasets is given in Fig. 1. Following an object-based approach to image analysis, segments have been used as sample type. Each sample is composed of a georeferenced segment geometry, a manually assigned class label and an image feature vector describing it. At least 75 % of a segment need to be covered by a particular class in order to be assigned this class label.

For the recognition of built-up areas, 3,200 training samples structured in 8 land-use/land-cover classes (400 samples per class) have been compiled through visual image interpretation by a human operator (TRAIN_BU). For a further stratification of the built-up area in terms of urban structure type patterns, an additional training dataset has been compiled (TRAIN_UST). 800 samples divided in 4 urban structure type classes have been manually labeled based on visual interpretation of Landsat and very high-resolution satellite images, ground-truth knowledge and in-situ images acquired by a mobile mapping system [38]. The training samples have been selected in a guided manner and are distributed in patches over the study area.

In order to assess the final map accuracies, stratified random samples on the resulting classifications have been taken. This resulted in a total of 770 test samples for the built-up area classification (TEST_BU) and 100 test samples for the urban structure types classification (TEST_UST). For the assessment of the settlement type classification, 100 settlement locations from OpenStreetMap have been randomly sampled (TEST_ST). The settlements were filtered prior to the sampling according to the tags “city” and “village”, which were used as indication for urban and rural settlements respectively. None of the test samples were included in training the classifier.

Moreover, independent test datasets have been compiled for a comprehensive quality assessment of the urban structure types classification. For this purpose 5,000 per-building samples from in-situ surveys [37] in several sites in Kyrgyzstan and Tajikistan have been used (TEST_UST_INSITU). These samples include the location and footprint of single buildings along with an extended set of structural and non-structural attributes in a standardized format following the building taxonomy of the Global Earthquake Model (GEM) [39]. These include the building height above ground, building footprint

area, construction material and occupancy type. Furthermore, for 370 image segments labeled as 'built-up', a dataset of building footprints (49'000 instances) has been extracted from OpenStreetMap and checked for coverage and consistency (TEST_UST_OSM).

3.5 Performance evaluation

To assess the classification performance standard accuracy measures have been derived from error matrices including precision, recall, accuracy and F1 score [40]. To increase the reliability of the measures and to reduce the bias resulting from specific training-testing scenarios five-fold cross-validation has been used on the reference datasets (TRAIN_BU, TRAIN_UST). Performance measures are therefore reported as average over the results of the cross-validation iterations. To account for the multiple classes, the weighted average of the performance measures of each class are provided. Additionally, the final map accuracies have been evaluated by deriving error matrices and standard accuracy measures in a non-cross-validated manner from independent test datasets (TEST_BU, TEST_UST, TEST_ST). In order to provide a thorough accuracy assessment that goes beyond a simple hard classification comparison, the results have also been compared with ground-truth data from in-situ surveys (TEST_UST_INSITU). Summary statistics have been compiled for main building features in order to get a better insight into inter-class homogeneity and inter-class heterogeneity. Moreover, the urban structure types classification has been evaluated by a comparison with landscape metrics [21] that were computed from per-building geometries (TEST_UST_OSM). In this context, the classified image segments form the landscape units that are being described by metrics that quantify the spatial characteristics of the buildings inside the unit. In this study, we compare four well known landscape metrics (patch density, Euclidean nearest neighbor distance, shape index and landscape index) across the urban structure type classes. The landscape metrics were calculated using the public domain software FRAGSTATS Version 4.2 [41].

4 Results

4.1 Built-up areas

The Landsat image tiles have been segmented by the previously introduced graph-based segmentation algorithm. For the study area 1,261,877 segments have been produced with a parameter combination of $k=25$, $m=75$ which provided the best overall quality score of 384 segmentation runs on a number of test image tiles.

i) Performance evaluation and the influence of digital elevation features

A performance evaluation of SVM has been carried out by using cross-validation on the reference data with iteratively splitting the data into training and testing subsets. In order to test the influence of features derived from the SRTM digital elevation model, two different feature subsets have been considered for evaluation – first an extended feature space including 8 elevation features, and second a reduced feature space excluding the elevation features. The reference samples were kept identical between the two feature sets. Recursive feature selection was performed independently on the two feature spaces and receiver operating characteristic (ROC) curves, learning curves and performance measures derived from aggregated error matrices have been computed. The resulting feature subsets for each feature space are depicted in Fig. 3 a and b. It can be seen that elevation features provide significant discriminating power for the classification task at hand and main elevation features include the standard deviation (height_std), minimum (height_min) and maximum (height_max) elevation value.

However, when looking at the classification performances (Fig. 4a and b) good results can be measured on both feature spaces with largest difficulties arising from the separation of shrubs (220) and grasslands (230), which show smallest area under the curve (AUC) values. Built-up areas (100) are very well separable from the other classes. With mean precision, recall, F1 and accuracy values of 0.84 to 0.85 for both feature spaces the overall classification performance is very good. Standard deviations of the measures are small with maximum 0.03 and are comparable between the two feature spaces. Fig. 4c and d depict the learning curves, which indicate that enough samples are available in order to design a robust classifier with good generalization ability. The gap between training and cross-validation score is very small for large sample sizes, the cross-validation score in both cases is sufficiently high with values above 0.8 and adding more samples does not significantly improve the cross-validation score.

It becomes evident from the comparison of the classification results between the two feature spaces that, despite the class separability power of elevation features, the extended feature space does not provide a significant gain in the overall classification performance with respect to the reduced feature space (Fig. 4). The computation of additional elevation features would, however, trigger higher (pre-) processing and storage costs (pre-processing time and storage space of the SRTM tiles) for the classification framework. Due to the obvious cost-benefit reasons, elevation features were not further considered in the classification.

ii) Classification, refinement and accuracy assessment

The trained and tuned SVM classifier has been applied to all the 1,261,877 segments using the feature space identified by recursive feature selection. The accuracy measures derived from the independent test dataset (TEST_BU) are provided in Tab. 3a. It can be seen that high average precision, recall and F1 score values are reached. However, the built-up class shows a low precision of 0.52 indicating a large number of false positives in the classification result. In order to improve the built-up area mask for the succeeding analysis steps, the classification output has been validated and refined with geo-information from other freely available sources. In a first step, a connected components analysis has been carried out to dissolve the classified segments. The so derived built-up area units have then been validated against point locations of settlements from OpenStreetMap [42] and Geonames [43] by checking whether a settlement location falls into a classified built-up area unit or not. The units that were not intersecting with settlement locations were identified as being potential false positives and were discarded. Besides the comprehensive list of settlement locations, also few manually digitized settlement outlines existed in OpenStreetMap for the study area. These were integrated with the classification in order to decrease the number of false negatives. Tab. 3b shows the updated accuracy measures for the refined built-up area map. It can be seen that the refinement improved clearly the precision, recall and F1 score values respectively.

4.2 Settlement types

Within this study we use a strongly simplified approach to distinguish settlement types. Image segmentation density of the previously identified built-up areas is used as proxy measure to classify urban from rural areas. More specifically, we extract points within the segments that are closest to the segments centroids and compute their density distribution over the study area by using a quartic (bi-weight) kernel with 5,000 m radius. A subset of the resulting built-up segment density map is depicted in Fig. 5 a. For each built-up segment mean density values have been computed from the underlying density distribution. Based on the mean values two clusters with equal interval have been classified indicating low-density (<0.5) and high-density (≥ 0.5) segments. Low-density segments were considered being rural whereas segments with high density values were classified as being urban. Fig.

5 b shows the urban-rural classification for the same subset as depicted in Fig. 5 a. Despite its simplicity, the results for the study area are satisfactory and a comparison with the independent test dataset (TEST_ST) indicates precision, recall and F1 score values of 0.94 (Tab. 4).

4.3 Urban structure types

The delineation of a settlement into areas of relatively homogeneous urban structure types aims at providing a generalized but spatially defined description of a building stock in terms of its predominant building types and their spatial alignment. In this context, a higher spatial resolution of the satellite image could potentially be beneficial for class separability. Therefore, before running the classification on the whole study area, the influence of the pan-chromatic band of Landsat-8 on the classification performance was evaluated and pan-sharpened test image tiles (15 m GSD) were compared with not pan-sharpened ones (30 m GSD).

i) Performance evaluation and the influence of pan-sharpening

In order to compare the pan-sharpened with the not pan-sharpened image tile, the segmentation of the pan-sharpened image has been recomputed under consideration of scene specific segmentation parameters. The reference samples were transferred to the new segmentation and the feature vectors were recalculated. After performing recursive feature selection, ROC curves, learning curves and performance measures derived from aggregated error matrices have been computed with cross-validation for the multi-spectral and for the pan-sharpened image tiles separately. The respective feature subsets for each feature space are depicted in Fig. 3 c and d. It can be seen from Fig. 6a and b that good performance can be measured on both image sets. With mean precision, recall, F1 and accuracy values above 0.75 for both image types the overall classification performance is good. Standard deviations of the measures are slightly higher on the multi-spectral image indicating a possibly higher sensitivity to changes in the training samples with respect to the pan-sharpened image. Fig. 6c and d depict the learning curves, which show a large gap between validation and training score at small training sample sizes indicating a tendency of the classifier to over-fit the data. Increasing the training sample size can improve the validation score, lower the variance and therefore increase the generalization ability of the classifier. However, on the pan-sharpened data more training samples are needed to constrain a classification model that reduces the gap between training and cross-validation scores at high score values. More than 350 samples are needed on the pan-sharpened dataset to achieve validation score values of 0.7 and above. On the multi-spectral dataset the solution converges faster and only 200 samples are needed to achieve a similar classification performance.

Due to the negligible performance improvement through pan-sharpening, only the multi-spectral images are used in the following for urban structure type classification. This is further justified by the results from the learning curves and the additional computational efforts and storage capacities that would be involved when using pan-sharpened images.

ii) Classification and accuracy assessment

The previously trained and tuned SVM classifier has been applied to 37,733 segments that were classified as being built-up. Fig. 7 shows the computed classification output and the spatial distribution of entropy values for a subset of the study area. The accuracy measures derived by comparison with the independent test data (TEST_UST) give an average precision of 0.81, recall of 0.77 and an F1 score of 0.78 (Tab. 5). Major difficulties arise from detecting residential high rise (113) and commercial / industrial (120) classes. From the close-up in Fig. 7 higher entropy values can be observed for segments that appear to be more heterogeneous in terms of their building composition. This, however,

is just a first observation and will be evaluated further below.

iii) Comparison with in-situ reference data

A comparison of the classification results with in-situ reference data (TEST_UST_INSITU) is depicted in Fig. 8. It can be observed that the different urban structure type classes can be clearly distinguished from each other by the distribution of the observed building height measured in number of storeys above ground. The structure types show relatively homogeneous intra-class distributions and represent well the according storey ranges from the classification scheme. Similarly, the footprint area distributions distinguish well the low-rise (111, 112) from the high-rise residential (113) and the commercial / industrial classes (120). Also the vulnerability class distributions show significant distinctions between the urban structure types, where higher values indicate higher seismic resistance and therefore lower vulnerability. Accordingly, low-rise residential types (111, 112) host the most vulnerable buildings, followed by commercial / industrial areas (120) and high-rise residential types (113) which accommodate the largest spread of vulnerability classes but show a generally lower vulnerability. Moreover, the compositions of construction material types between the classes differ and for most of the classes a predominant material type can be identified, which further underlines the intra-class homogeneity and inter-class heterogeneity. Also, the occupancy distribution indicates a clear separation between different urban structure type classes, where predominant occupancy types can be identified for each class.

iv) Comparison with landscape metrics

From Fig. 9 it can be observed that the different classes derived from Landsat-8 can be distinguished well by the landscape metrics for the segments for which building footprints from OpenStreetMap were available (TEST_UST_OSM). To this regard, *patch density* equals the number of patches of a specific land cover class (buildings) divided by the total segment area. The density distribution is well in line with the classification scheme (Tab. 2) in that class 112 (residential, low-rise, high density) shows highest density of buildings, followed by class 111 (residential, low-rise, low density), whereas classes 113 (residential, high-rise) and 120 (commercial / industrial) appear to be areas of low building density. This is confirmed by the *Euclidean nearest neighbor distance* which provides the shortest edge-to-edge distance from cell center to cell center between each patch of a class and its nearest neighbor patch. The area weighted mean of the measure reported per segment is used which can be interpreted as a measure of patch isolation per segment. Classes 111 and 112 (residential, low-rise) show small distance values indicating a spatial clustering of buildings whereas classes 113 (residential, high-rise) and 120 (commercial / industrial) show larger distances indicating stronger patch isolation.

The complexity of a patch shape compared to a square shape of the same size is measured by the *shape index*. The index is computed at the patch (per-building) level and the area weighted mean is reported per segment. Across the identified urban structure type classes the buildings in class 112 (residential, low-rise, high-density) seem to be the most complex in shape. Buildings in the other classes show more regular shapes with the most regular constructions belonging to class 120 (commercial / industrial). A similar distribution can be observed when looking at the *landscape shape index*. It measures the perimeter-to-area ratio for the landscape (segment) as a whole and can be interpreted as a measure of the overall geometric complexity of a segment. More complex segment shapes can be found in classes in which also more irregular buildings are present. This highlights also the sensitivity of the segmentation algorithm to observable structural differences within built-up areas even when applied to medium resolution Landsat-8 images.

v) Spatial distribution of classification uncertainties

Fig. 10 a shows the distribution of classification uncertainties for all 37,733 built-up segments grouped by settlement type and by urban structure type class. The uncertainty is measured by the Shannon entropy and computed from the soft answers of the SVM as described above. It can be seen that in rural areas the classifier shows slightly higher entropy values with respect to urban areas. The entropy distribution for the urban structure type classes largely reflects the results of the accuracy assessment in Tab. 5. Classes 111 and 112 (residential, low-rise) show lowest entropy and highest accuracy values, whereas class 113 (residential, high rise) shows highest entropy and lowest accuracy values. Moreover, Fig. 11 shows that within the 100 test samples (TEST_UST) the entropy values are clearly higher for falsely classified segments than for the correctly classified ones.

In order to test the previously mentioned hypothesis that building heterogeneity within segments can negatively influence the classification result, an ordinary least squares regression has been performed with entropy as dependent and landscape metrics as independent variables. The coefficient of variation was computed per metric for each test segment (TEST_UST_OSM) to provide measures of segment's building heterogeneity. A linear regression model combining the coefficient of variation of the Euclidean nearest neighbor distance and the shape index was found to provide statistically significant explanations of the entropy values (Tab. 6). The model explains 83 % of the variance of the entropy values (adj. $R^2 = 0.831$). The Durbin-Watson test indicates that the model is not affected by autocorrelation and also multicollinearity is not present according to the condition index. The signs of the partial regression coefficients and the according plots (Fig. 10 b) indicate a positive relationship between building heterogeneity within segments and entropy. This means that larger heterogeneity in the composition of buildings within segments – here defined in terms of variations of the building shape and variations of the relative distance between buildings – increases the classification uncertainty.

5 Summary of results

The built-up area that has been identified within this study from Landsat-8 data of 2013 covers 2 % (11,182 km²) of the total study area. 78 % of this were identified as being rural, whereas only 22 % can be considered urban areas with respect to the simplified definitions of urban and rural given within this work (Fig. 12a). Besides the overall rural characteristics of the study area, differences of settlement type distribution can be observed between the neighboring countries (Fig. 12b). Eastern Uzbekistan (E-UZ) shows the largest degree of urbanization whereas on the contrary in Kyrgyzstan (KG) a negligible percentage of the built environment can be regarded as being urban. With more than 80 % total coverage (Fig. 12c), the clearly dominant urban structure type classes are residential low-rise (classes 111 and 112). Residential high-rise (class 113) and commercial / industrial (class 120) classes cover both less than 10 % of the built-up area. Whereas the dominance of residential low-rise classes can be observed throughout all the neighboring countries, differences arise with respect to commercial / industrial and residential high-rise classes (Fig. 12d). Commercial / industrial classes make for a higher percentage of the built-up area in Eastern Uzbekistan (E-UZ) and Tajikistan (TJ), whereas residential high-rise areas can be mainly found in South Kazakhstan (S-KZ), Kyrgyzstan (KG) and Uzbekistan (E-UZ). When looking at the urban structure types distribution further grouped by settlement type and depicted in absolute terms (Fig. 12e), it can be observed that E-UZ covers the largest built-up area followed by KG, S-KZ and TJ in decreasing order. Residential low-density classes can be mainly found in rural areas throughout all the countries, whereas high-density classes dominate in urban areas. Also commercial / industrial classes are stronger represented in rural than in urban settlement types.

6 Discussion and conclusion

This study described a regional classification of settlement patterns from Landsat-8 data by using a SVM in the context of an object-based image analysis. Classification tasks that were tackled go beyond a simple binary built-up area extraction and include settlement type classification and urban structure types recognition. Moving towards a pattern recognition approach, which effectively makes use of image domains beyond the spectral content on the pixel level, proved to be a valuable solution when extracting information on the built environment.

The information content of Landsat-8 images was sufficient for the classification tasks and even detailed urban structures could be extracted with a satisfying accuracy. Recursive feature selection identified additional features from a digital elevation model as being important for class separability. However, adding them to the feature space did not significantly improve the overall classification performance for built-up area extraction with respect to an image feature space only. The conclusion of excluding digital elevation features from the final feature space in this study holds for this particular case where the aim is also to reduce computational classification costs. However, the decision on whether to keep or discard this feature type may change for other applications and system requirements, given the observed importance of digital elevation features for class separability in the general case. The most important spectral bands of Landsat-8 for class separability over all classification tasks were band 1 (coastal / aerosol, 0.43-0.45 μm) and band 3 (green, 0.53-0.59 μm). Texture features, especially textures derived from Landsat-8 bands 1 (coastal / aerosol, 0.43-0.45 μm) and 6 (short wave infrared, 1.57-1.65 μm), showed a significant importance for class separability throughout all the classification tasks. The influence of the coastal / aerosol band indicates the significance of the improved spectral resolution of Landsat-8 with respect to its predecessors. NDBI, MNDWI and SAVI appeared to be the most important band indices. Landscape metrics proved to be valuable features to characterize the urban structure. In this study the metrics were used for validation of the classification results. Where very high-resolution per-building data is available, such metrics could be used as additional features to assist the classification of segments into urban structure types, as is for example proposed in Herold *et al.* (2003) [21].

It was also found that pan-sharpening could improve classifier performance for urban structure types recognition. However, the difference with respect to not pan-sharpened images was found to be nearly negligible. A more comprehensive study on the effects of pan-sharpening on the classifier performance would be needed to better understand the observations. In this context, especially different pan-sharpening algorithms should be compared with each other.

Shannon entropy computed from the soft answers of the SVM proved to be a valid method in order to measure the spatial distribution of classification uncertainties. A positive correlation could be observed between entropy and the heterogeneity of the urban structure inside segments (here measured by the coefficient of variations of the Euclidean nearest neighbor distance and the shape index). Therefore, the entropy distribution could potentially be used to prioritize further training data selection as part of an active learning approach, or it could guide manual classification refinements, or the sampling of in-situ surveys to acquire more in-depth information about the segment composition.

SVM performed well for the identification of urban structure type patterns. Performance evaluation of the classifier highlighted the importance of SVM parameter tuning, feature selection and size of the training data. Low standard deviations of accuracy values from cross-validation during the training stage indicated a good transferability of the learning machine. This was further confirmed by the final map accuracy which showed acceptable to good precision and recall values for all classes over the

whole study area. With respect to the extraction of built-up areas, the trained learning machine showed low precision values and thus a large number of false positives when applied to the full study area and compared to an independent testing dataset. However, cross-validation as part of the training stage showed good performance of the learning machine, and low standard deviations of the cross-validation scores indicated a potentially good transferability. This discrepancy between training and testing performance indicates a possible mismatch between source and target domain and / or a problem with the training sample approach. The training data were sampled from three different image scenes in a guided manner, where most representative samples were being selected based on expert judgment and comparison with very high resolution satellite images from Google Earth. Even though, the selection of the training samples and image scenes tried to account for the heterogeneity of land-cover types that are present in the study area, it is possible that the not built-up land-cover classes are not properly described by the training samples. Despite the fact that efforts were being made to account for domain-shifts (atmospheric correction, across-scene training) and class heterogeneity, still it seems that either more robust features and / or different training approaches are needed to improve the classification performance. In order to proof this and to check which classes cause confusion, further tests are needed and a more detailed test dataset that covers all the land-cover classes needs to be derived.

Alternatively, in this study a significant improvement could be achieved by validation of the built-up area mask against available global settlement location datasets. However, such a refined mask does not present anymore a purely statistical classification but is the result of a more complex operation that incorporates ancillary information. The success of such an operation is therefore strongly dependent on the availability of the settlement datasets and a successful transfer to other regions is not guaranteed.

Given the above observations, future work will focus on testing domain adaptation [44], the implementation of other robust image features like texture-based built-up presence indices [45] and different training sample approaches, to improve the transferability of trained learning machines between image scenes and especially to increase the precision of the built-up area extraction without an integration of ancillary data. Also a comparison of the derived settlement products with existing global land use / land cover products is envisaged. The settlement products are applied as part of a larger strategy on seismic risk assessment in Central Asia. They are used as base-layers to prioritize in-situ surveys and to derive up-to-date information about exposed assets and their seismic vulnerability over large areas [37].

Acknowledgments

This work was supported by the Earthquake Model Central Asia (EMCA) and the SENSUM project (Grant Agreement Number 312972). The authors would like to thank the editors and the anonymous reviewers for suggestions that helped to improve this paper and K. Fleming for English language revision.

References

- [1] M. Salvatore, S. Pozzi, E. Ataman, B. Huddleston, and M. Bloise, *Mapping global urban and rural population distributions*. Rome: FAO, 2005.
- [2] *Principles and Recommendations for Population and Housing Censuses*. New York: United Nations, 1998.
- [3] C. Geiß and H. Taubenböck, “Remote sensing contributing to assess earthquake risk: from a literature review towards a roadmap,” *Natural Hazards*, pp. 1–42, 2013.
- [4] M. Pittore, M. Wieland, and K. Fleming, “Perspectives on Global Dynamic Exposure Modelling for Geo-risk Assessment: from Remote Sensing to Crowd-Sourcing,” *Natural Hazards*, under review.
- [5] M. Pittore, D. Bindi, J. Stankiewicz, A. Oth, M. Wieland, T. Boxberger, C. Milkereit, and S. Parolai, “Towards a new concept of optimized, loss-driven earthquake early warning and rapid response systems,” *Seismological Research Letters*, vol. 85, no. 6, pp. 1328–1340, 2014.
- [6] P. Gamba and M. Herold, Eds., *Global Mapping of Human Settlement: Experiences, Datasets, and Prospects*. Boca Raton: CRC Press, 2009.
- [7] D. Potere, A. Schneider, S. Angel, and D. L. Civco, “Mapping urban areas on a global scale: which of the eight maps now available is more accurate?,” *International Journal of Remote Sensing*, vol. 30, no. 24, pp. 6531–6558, 2009.
- [8] M. Pesaresi, D. Ehrlich, S. Ferri, A. Florczyk, S. Freire, F. Haag, M. Halkia, A.M. Julea, T. Kemper, and P. Soille, “Global human settlement analysis for disaster risk reduction,” *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-7/W3, pp. 837-843, 2015.
- [9] T. Esch, M. Marconcini, A. Felbier, A. Roth, W. Heldens, M. Huber, M. Schwinger, H. Taubenböck, A. Muller, and S. Dech, “Urban Footprint Processor: Fully Automated Processing Chain Generating Settlement Masks From Global Data of the TanDEM-X Mission,” *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 6, pp. 1617–1621, 2013.
- [10] S. S. Bhatti and N. K. Tripathia, “Built-up area extraction using Landsat 8 OLI imagery,” *GIScience & Remote Sensing*, vol. 51, no. 4, pp. 445–467, 2014.
- [11] D. Poursanidisa, N. Chrysoulakisa, and Z. Mitrakaa, “Landsat 8 vs. Landsat 5: A comparison based on urban and peri-urban land cover mapping,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 35, no. 2, pp. 259–269, 2015.
- [12] D. P. Roy, M. A. Wulder, T. R. Loveland, W. C.E., R. G. Allen, M. C. Anderson, D. Helder, J. R. Irons, D. M. Johnson, R. Kennedy, T. A. Scambos, C. B. Schaaf, J. R. Schott, Y. Sheng, E. F. Vermote, A. S. Belward, R. Bindaschadler, W. B. Cohen, F. Gao, J. D. Hipple, P. Hostert, J. Huntington, C. O. Justice, A. Kilic, V. Kovalsky, Z. P. Lee, L. Lymburner, J. G. Masek, J. McCorkel, Y. Shuai, R. Trezza, J. Vogelmann, R. H. Wynne, and Z. Zhu, “Landsat-8: Science and product vision for terrestrial global change research,” *Remote Sensing of Environment*, vol. 145, pp. 154–172, 2014.
- [13] M. C. Hansen and T. R. Loveland, “A review of large area monitoring of land cover change using Landsat data,” *Remote Sensing of Environment*, vol. 122, pp. 66–74, 2012.
- [14] J. Chena, J. Chenb, A. Liaoa, X. Caob, L. Chena, X. Chenb, C. Hea, G. Hana, S. Penga, M. Lua, W. Zhanga, X. Tongc, J. Mills, “Global land cover mapping at 30 m resolution: A POK-based operational approach,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 103, pp. 7–27, 2015.
- [15] Y. Ban, P. Gong, C. Giri, “Global land cover mapping using Earth observation satellite data:

Recent progresses and challenges,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 103, pp. 1–6, 2015.

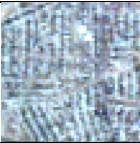





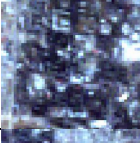





- [16] Q. Weng, “Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends,” *Remote Sensing of Environment*, vol. 117, no. 0, pp. 34–49, 2012.
- [17] G. Yan, J. -F. Mas, B. H. P. Maathuis, Z. Xiangmin, and P. M. Van Dijk, “Comparison of pixel-based and object-oriented image classification approaches—a case study in a coal fire area, Wuda, Inner Mongolia, China,” *International Journal of Remote Sensing*, vol. 27, no. 18, pp. 4039–4055, 2006.
- [18] M. Wieland and M. Pittore, “Performance evaluation of machine learning algorithms for urban pattern recognition,” *Remote Sensing*, vol. 6, no. 4, pp. 2912–2939, 2014.
- [19] E. Banzhaf and R. Hofer, “Monitoring Urban Structure Types as Spatial Indicators With CIR Aerial Photographs for a More Effective Urban Environmental Management,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 1, no. 2, pp. 129–138, 2008.
- [20] Taubenbock H, Wurm M, Roth A, Schmidt M, Mehl H, Dech S, Esch T, Thiel M, Ullmann T, and G. A. Remote Sensing for Environmental Monitoring, “Urban structure analysis of mega city Mexico City using multi-sensoral remote sensing data,” *Proc SPIE Int Soc Opt Eng Proceedings of SPIE - The International Society for Optical Engineering*, vol. 7110, 2008.
- [21] M. Herold, X. Liu, and K. C. Clarke, “Spatial metrics and image texture for mapping urban land use,” *Photogrammetric Engineering & Remote Sensing*, vol. 69, no. 9, pp. 991–1001, 2003.
- [22] A. Hashimov, I. Pathiddinov, M. Makhmutova, A. Akhmetova, B. Raki-Sheva, K. Berentaev, L. Torgasheva, M. Suyunbaev, A. Mamytova, T. Burzhubaev, M. Karimova, B. Khabibov, R. Zoyirov, and M. Sobirov, “Urbanization in Central Asia: challenges, issues and prospects,” UN ESCAP, Tashkent, Analytical Report 2013/03, 2013.
- [23] “Orfeo Toolbox Cookbook,” 2015. [Online]. Available: <https://www.orfeo-toolbox.org/SoftwareGuide/SoftwareGuidech13.html#x41-21500013.2>. [Accessed: 13-May-2015].
- [24] D. Fasbender, J. Radoux, and P. Bogaert, “Bayesian Data Fusion for Adaptable Image Pansharpening,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 6, pp. 1847–1857, 2008.
- [25] P. Felzenszwalb and D. Huttenlocher, “Efficient Graph-Based Image Segmentation,” *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.
- [26] M. Wieland, M. Pittore, S. Parolai, and J. Zschau, “Exposure Estimation from Multi-Resolution Optical Satellite Imagery for Seismic Risk Assessment,” *ISPRS International Journal of Geo-Information*, vol. 1, no. 3, pp. 69–88, 2012.
- [27] V. Dey, Y. Zhang, and M. Zhong, “A review on image segmentation techniques with remote sensing perspective,” In: Wagner W., Székely, B. (eds.): *ISPRS TC VII Symposium – 100 Years ISPRS*, Vienna, Austria, July 5–7, 2010.
- [28] B. Johnson and Z. Xie, “Unsupervised image segmentation evaluation and refinement using a multi-scale approach,” *ISPRS Journal of Photogrammetry and Remote Sensing*, 2011.
- [29] R. Haralick, K. Shanmugam, and I. Dinstein, “Textural Features for Image Classification,” *IEEE Transactions on system, man and cybernetics*, vol. 3, no. 6, 1973.
- [30] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene selection for cancer classification using support vector machines,” *Machine learning*, vol. 46, no. 1–3, pp. 389–422, 2002.
- [31] “Scikit-learn,” 08-Jan-2015. [Online]. Available: <http://scikit-learn.org>. [Accessed: 08-Jan-2015].
- [32] C.-C. Chang and C.-J. Lin, “LIBSVM: a library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

- [33] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the IEEE*, vol. 86, no. 2, pp. 447–457, 1998.
- [34] J. Platt, "Probabilistic outputs for support vector machines and comparison to regularized likelihood method," *Advances in large margin classifiers*, vol. 3, no. 10, pp. 61–74, 1999.
- [35] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *The Journal of Machine Learning Research*, vol. 5, pp. 975–1005, 2004.
- [36] J. R. Anderson, *A land use and land cover classification system for use with remote sensor data*, vol. 964. US Government Printing Office, 1976.
- [37] M. Wieland, M. Pittore, S. Parolai, U. Begaliev, P. Yasunov, S. Tyagunov, B. Moldobekov, S. Sai, I. Ilyasov, and T. Abakanov, "A multi-scale exposure model for seismic risk assessment in Central Asia," *Seismological Research Letters*, vol. 86, no. 1, pp. 1–13, 2015.
- [38] M. Wieland, M. Pittore, S. Parolai, J. Zschau, B. Moldobekov, and U. Begaliev, "Estimating building inventory for rapid seismic vulnerability assessment: Towards an integrated approach based on multi-source imaging," *SDEE Soil Dynamics and Earthquake Engineering*, vol. 36, pp. 70–83, 2012.
- [39] S. Brzev, C. Scawthorn, A. W. Charleson, L. Allen, M. Greene, K. Jaiswal, and V. Silva, "GEM Building Taxonomy v2.0," GEM Building Taxonomy Global Component, Global Earthquake Model, Pavia, 2013.
- [40] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, Jun. 2006.
- [41] K. McGarigal, S. Cushman, and E. Banzhaf, *FRAGSTATS: Spatial Pattern Analysis Program for Categorical and Continuous Maps*. 2012.
- [42] "OpenStreetMap," 10-Jan-2011. [Online]. Available: <http://www.openstreetmap.org/>. [Accessed: 10-Jan-2011].
- [43] "Geonames," 2015. [Online]. Available: <http://www.geonames.org/>. [Accessed: 13-May-2015].
- [44] L. Bruzzone and M. Marconcini, "Toward the Automatic Updating of Land-Cover Maps by a Domain-Adaptation SVM Classifier and a Circular Validation Strategy," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 4, pp. 1108–1122, 2009.
- [45] M. Pesaresi, A. Gerhardinger, and F. Kayitakire, "A Robust Built-Up Area Presence Index by Anisotropic Rotation-Invariant Textural Measure," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 1, no. 3, pp. 180–192, 2008.

Tab. 1: List of image features derived from Landsat-8 and SRTM. All image features are computed per segment and all 10 spectral bands of Landsat-8 are used.

Feature ID	Image feature description	Feature class
mean_bx	Mean brightness value in image band x	spectral
std_bx	Standard deviation of brightness values in image band x	spectral
mode_bx	Mode of brightness values in image band x	spectral
max_pixel_bx	Maximum brightness value in image band x	spectral
min_pixel_bx	Minimum brightness value in image band x	spectral
wB	Weighted brightness	spectral
mean_std_mode_ndvi	Mean, standard deviation and mode values of normalized difference vegetation index (NDVI)	band index
smean_std_mode_savi	Mean, standard deviation and mode values of soil adjusted vegetation index (SAVI)	band index
mean_std_mode_mndwi	Mean, standard deviation and mode values of modified normalized difference water index (MNDWI)	band index
mean_std_mode_ndbi	Mean, standard deviation and mode values of normalized difference built-up index (NDBI)	band index
mean_std_mode_nbi	Mean, standard deviation and mode values of normalized built-up index (NBI)	band index
mean_std_mode_ndisi	Mean, standard deviation and mode values of normalized difference impervious surface index (NDISI)	band index
asm_bx	Angular Second Moment derived from the grey-level co-occurrence matrix (GLCM) in band x	textural
contrast_bx	Contrast derived from the GLCM in band x	textural
dissimilarity_bx	Dissimilarity derived from the GLCM in band x	textural
entropy_bx	Entropy derived from the GLCM in band x	textural
glcm_correlation_bx	Correlation derived from the GLCM in band x	textural
glcm_meani_bx	Mean derived from the GLCM in band x	textural
glcm_varianci_bx	Variance derived from the GLCM in band x	textural
homogeneity_bx	Homogeneity derived from the GLCM in band x	textural
cont_area	Area of the segment in pixel counts	geometrical
cont_perimeter	Perimeter of the segment	geometrical
convhull_perimeter	Perimeter of the convex hull of the segment	geometrical
cont_convexity	Convexity of the contour outlining the segment;	geometrical
cont_paxe1	Principal axes 1 of the contour outlining the segment	geometrical
cont_paxe2	Principal axes 2 of the contour outlining the segment	geometrical
cont_paxes_ratio	Ratio of principle axes	geometrical
cont_compactness	Compactness of the contour outlining the segment;	geometrical
rect_width	Rectangular width	geometrical
slope_mean_std	Mean and standard deviation of slope values	elevation
slope_max_min	Minimum and maximum slope values	elevation
height_mean_std	Mean and standard deviation of height values	elevation
height_max_min	Minimum and maximum height values	elevation

Tab. 2: Hierarchical classification scheme adapted to the study area and its building stock.

Class ID	Level 1	Level 2	Level 3	Appearance in Landsat	Appearance in WorldView-2	Ground-reference picture
111	Built-up	Residential	Low rise, low density			
112	Built-up	Residential	Low rise, high density			
113	Built-up	Residential	High rise			
120	Built-up	Commercial / Industrial	-			
210	Vegetation	Forrest	-			
220	Vegetation	Shrubs	-			
230	Vegetation	Grassland	-			
240	Vegetation	Agriculture	-			
300	Bare soil, rock, sand	-	-			
400	Water	-	-			
500	Snow, Ice	-	-			

Tab. 3: Accuracy measures for the classification of built-up areas derived from comparison with independent test samples (TEST_BU). Measures are provided for a) the raw classification output and b) the refined classification result.

(a) Raw classification					(b) Refined classification			
Class	Precision	Recall	F1 score	support	Precision	Recall	F1 score	support
Built-up	0.52	0.68	0.59	25	0.88	0.78	0.83	37
Not built-up	0.99	0.98	0.98	745	0.99	0.99	0.99	733
Avg / Total	0.97	0.97	0.97	770	0.98	0.98	0.98	770

Tab. 4: Accuracy measures for the classification of settlement types derived from comparison with independent test samples (TEST_ST).

Class	Precision	Recall	F1 score	support
rural	0.96	0.96	0.96	84
urban	0.81	0.81	0.81	16
Avg / Total	0.94	0.94	0.94	100

Tab. 5: Accuracy measures for the classification of urban structure types derived from comparison with independent test samples (TEST_UST).

<i>Class</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 score</i>	<i>support</i>
111	0.92	0.73	0.82	49
112	0.70	0.91	0.79	33
113	0.40	0.57	0.47	7
120	0.88	0.64	0.74	11
Avg / Total	0.81	0.77	0.78	100

Tab. 6: Ordinary least squares regression model (TEST_UST_OSM).

<i>Dep. Variable</i>	y (Entropy)	<i>Adj. R-squared</i>	0.831	
<i>Model</i>	OLS	<i>F-statistic</i>	916	
<i>Method</i>	Least Squares	<i>Durbin-Watson</i>	1.83	
<i>No. Observations</i>	370	<i>Cond. no.</i>	3.60	
	<i>coef</i>	<i>std_err</i>	<i>t</i>	<i>sig</i>
x1 (CV Eucl. Nearest N.)	0.004	0.001	5.275	0.005
x2 (CV Shape)	0.012	0.001	17.477	0.014

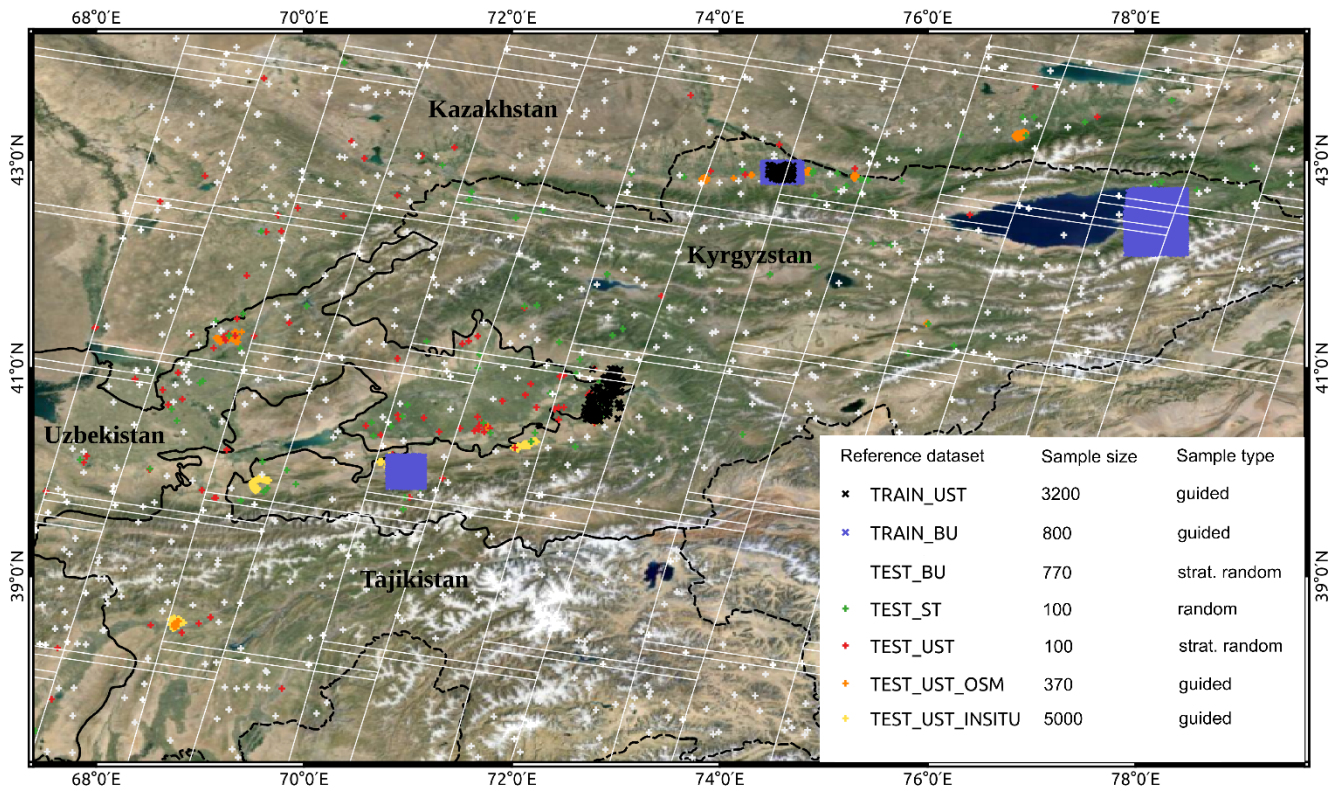


Fig. 1: Study area covering large parts of Central Asia. The Landsat-8 image mosaic and outlines of individual scenes are displayed in the background, superimposed by the training (TRAIN) and testing (TEST) datasets for built-up areas (BU), settlement types (ST) and urban structure types (UST).

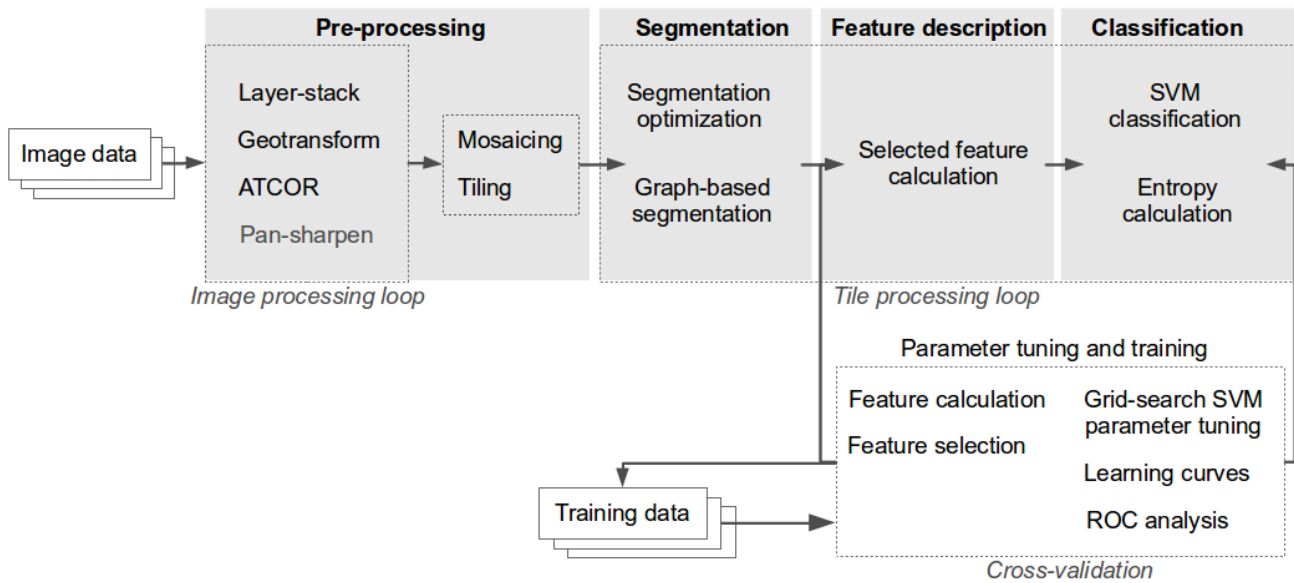


Fig. 2: Processing and analysis pipeline for settlement pattern recognition from multi-spectral satellite images.

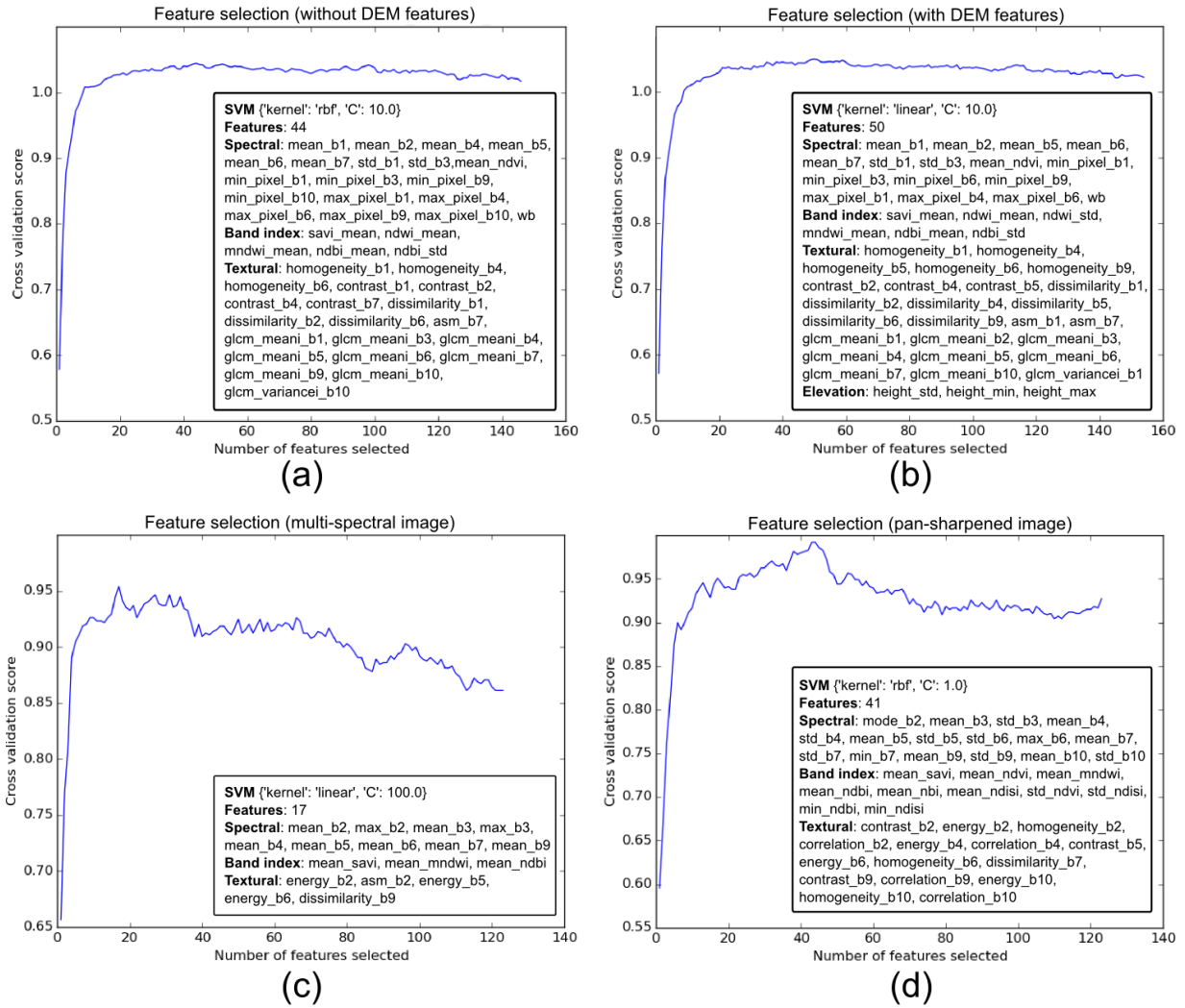


Fig. 3: Selected feature subsets for the classification of built-up areas on a) the Landsat image feature space and b) the extended feature space including SRTM elevation features. Feature subsets that were identified for the classification of urban structure types on the c) the images with 30 m and d) the pan-sharpened images with 15 m spatial resolution.

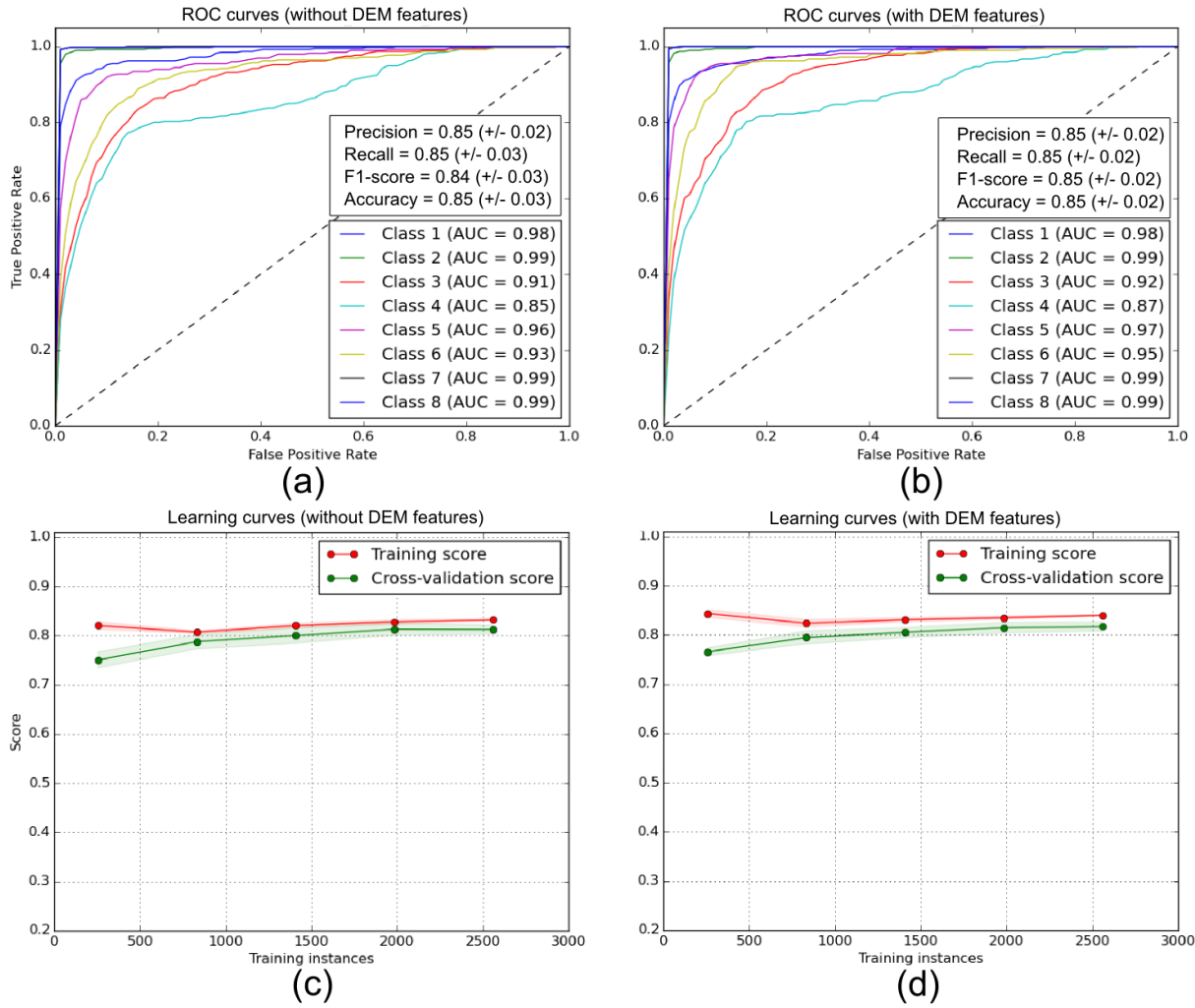


Fig. 4: ROC curves with performance measures and learning curves derived for the task of built-up pattern recognition for a) and c) the multi-spectral, b) and d) the extended feature space with elevation features.

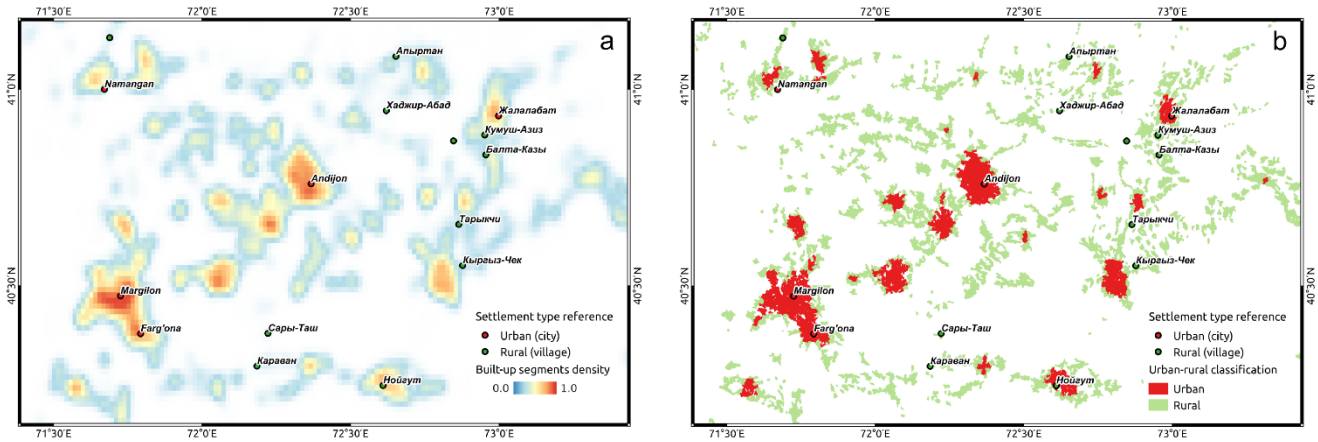


Fig. 5: Subset of the a) built-up segment density map and b) urban-rural classification with superimposed reference settlement types from OpenStreetMap for the Fergana valley in Uzbekistan and

Kyrgyzstan.

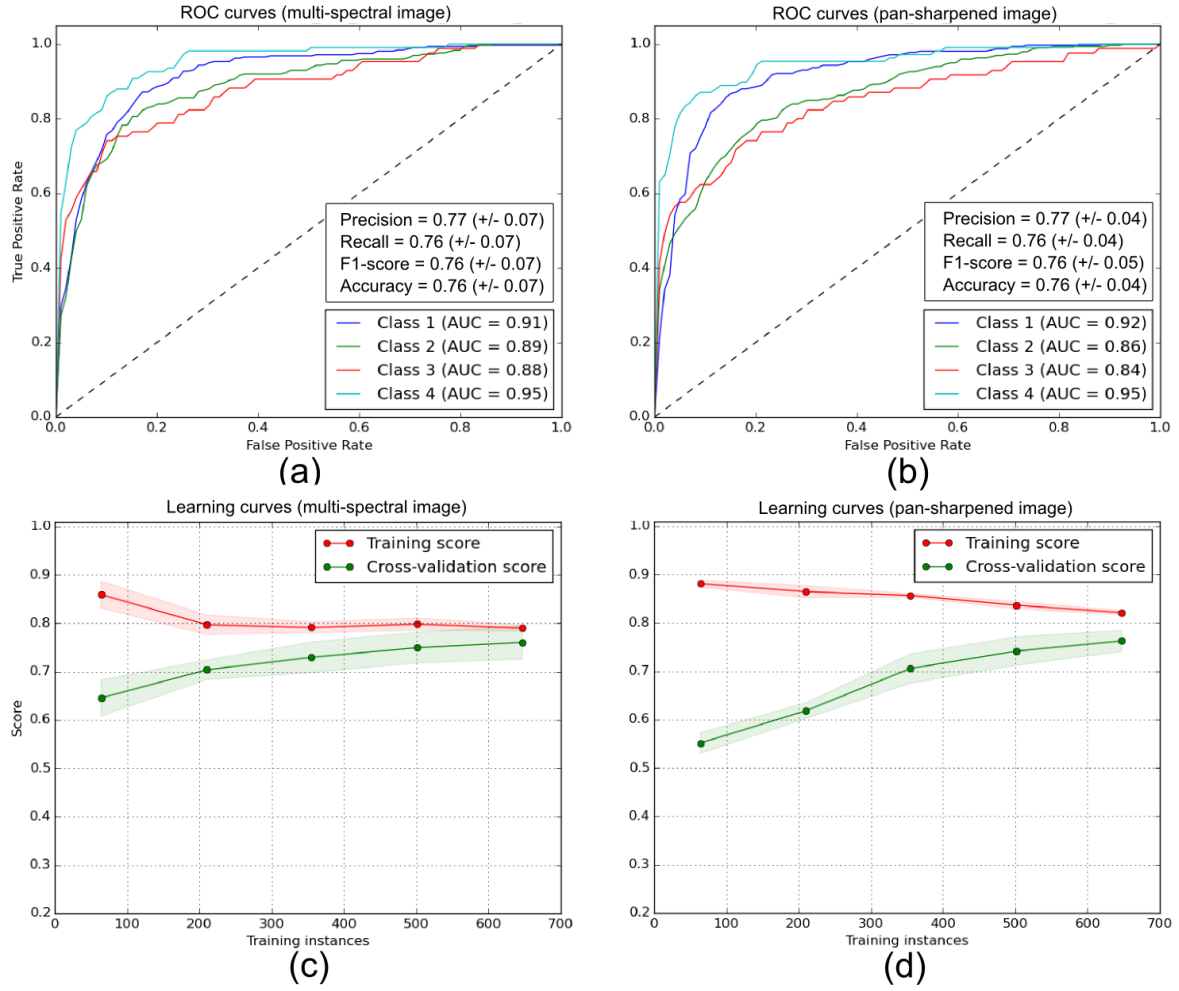


Fig. 6: ROC curves with performance measures and learning curves derived for the task of urban structure types classification for a) and c) the multi-spectral, b) and d) the pan-sharpened image.

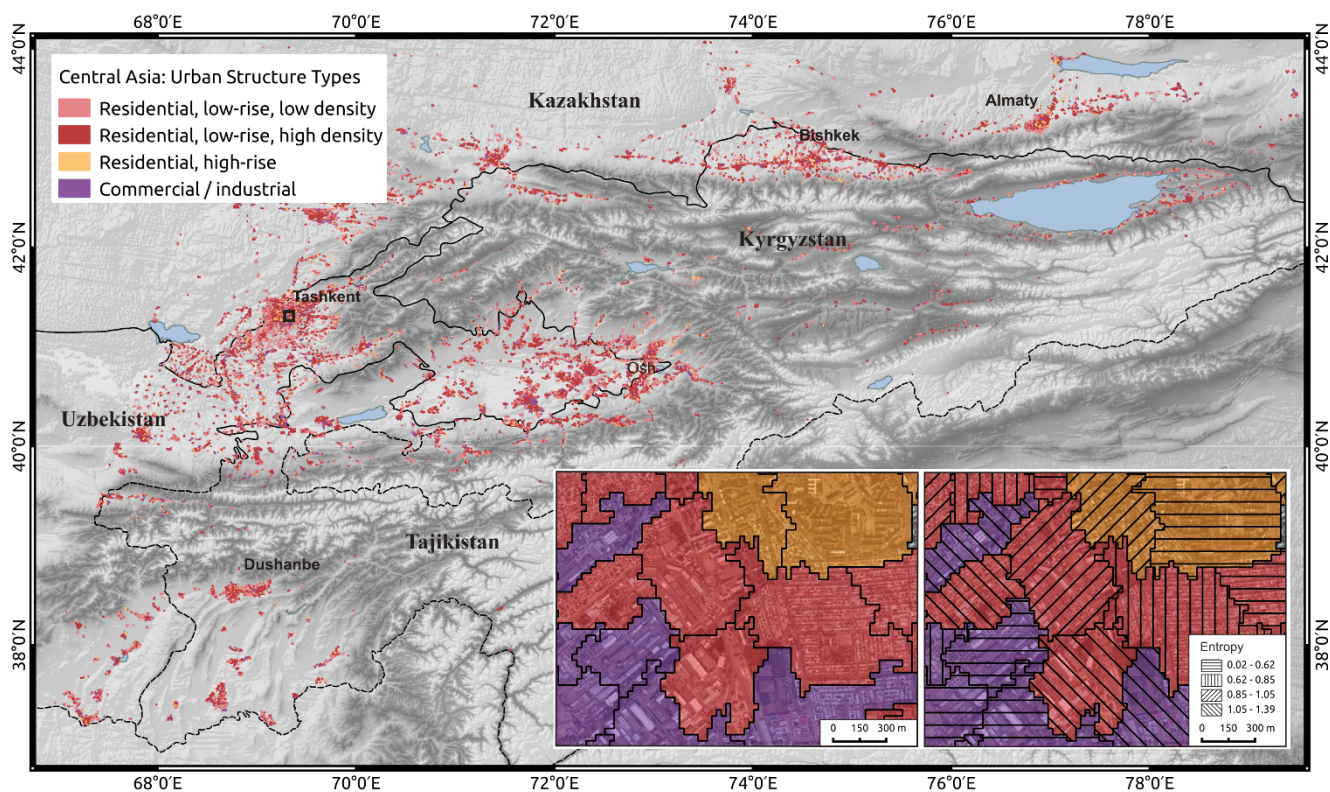


Fig. 7: Urban structure types classification from Landsat-8. The magnified view shows a close-up of the classification and the calculated entropy values for a neighborhood in Tashkent, Uzbekistan superimposed on a very high-resolution WorldView-2 image for visualization.

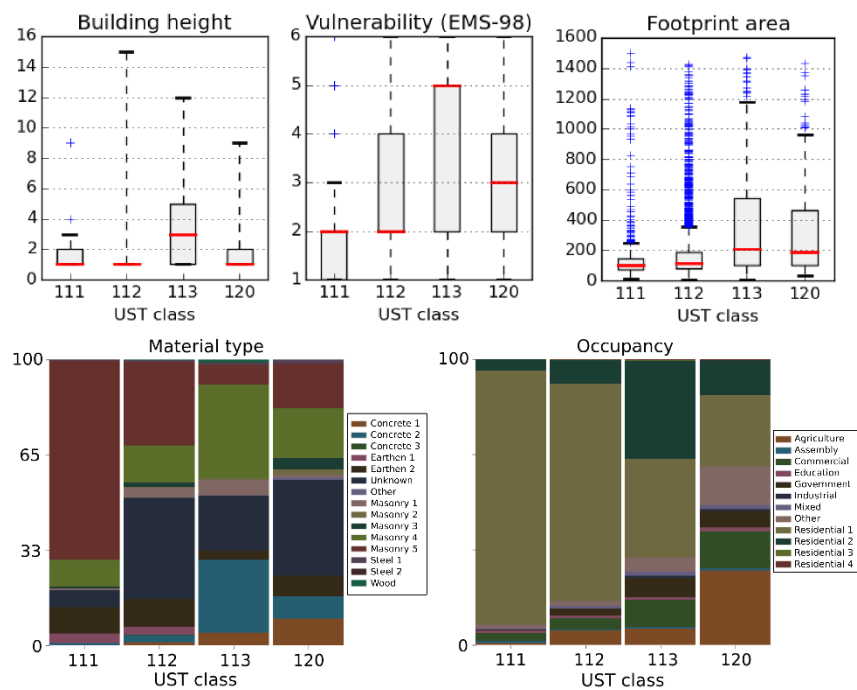


Fig. 8: Comparison of the urban structure types classification with ground-truth data from per-building surveys.

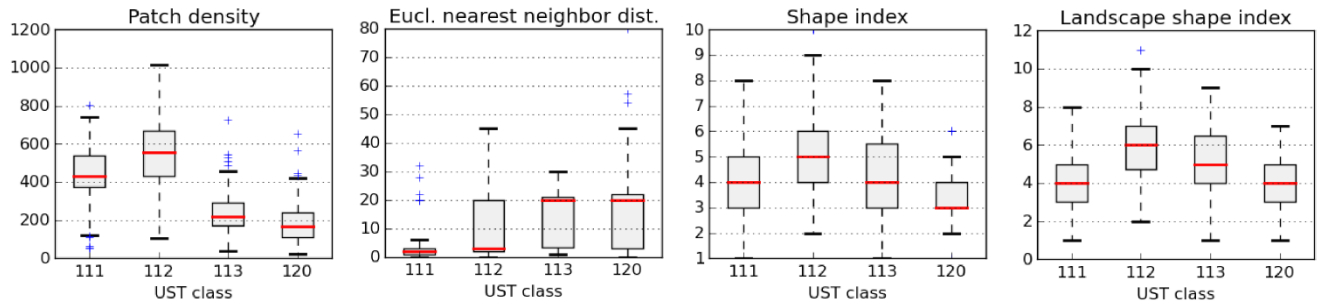


Fig. 9: Comparison of the urban structure types classification with landscape metrics computed from OpenStreetMap building data for sample segments.

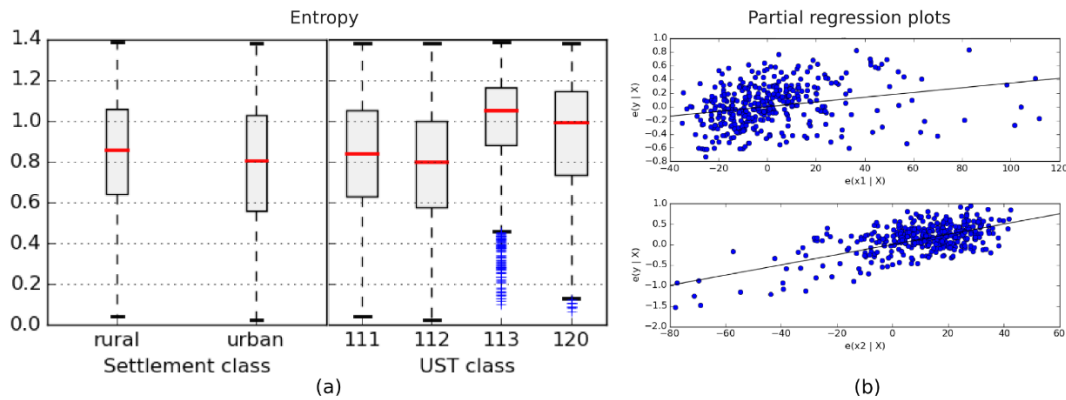


Fig. 10: a) Comparison of the entropy values with settlement and urban structure types classes for sample segments. b) Partial regression plots of an ordinal least square model with the coefficient of variation of the euclidean nearest neighbor distance (x1) and the shape index (x2) being the independent and entropy (y) being the dependent variables.

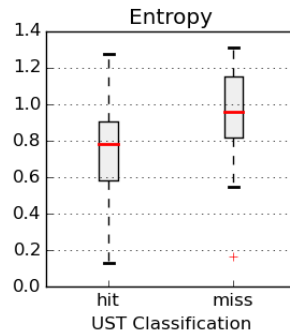


Fig. 11: Entropy value distribution between correctly classified (hit) and misclassified (miss) UST classes for 100 test samples (TEST_UST).

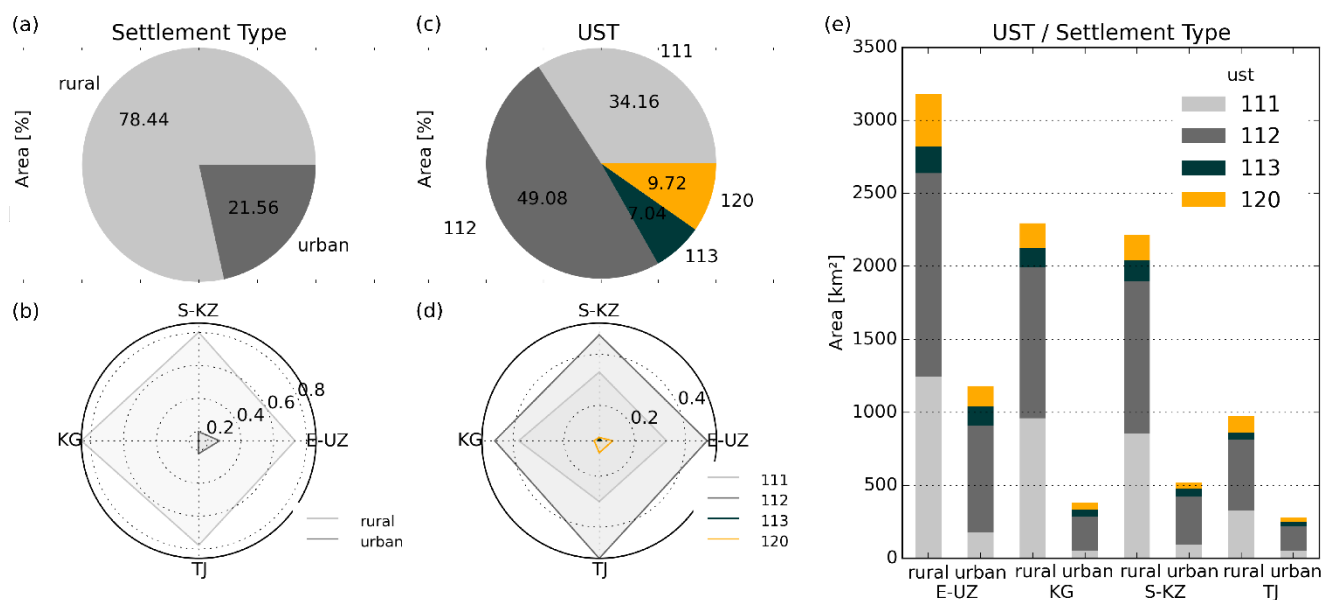


Fig. 12: a) Settlement type distribution for the study area. b) Settlement type distribution between the neighboring countries. c) Urban structure type distribution for the study area. d) Urban structure type distribution between the neighboring countries. e) Urban structure type distribution grouped by settlement type and country.