Article

Exposure Estimation from Multi-Resolution Optical Satellite Imagery for Seismic Risk Assessment

Marc Wieland *, Massimiliano Pittore, Stefano Parolai and Jochen Zschau

Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences, Telegrafenberg,
D-14473 Potsdam, Germany; E-Mails: pittore@gfz-potsdam.de (M.P.);
parolai@gfz-potsdam.de (S.P.); zschau@gfz-potsdam.de (J.Z.)

* Author to whom correspondence should be addressed; E-Mail: mwieland@gfz-potsdam.de;
Tel.: +49-331-288-1283; Fax: +49-331-288-1204.

Abstract: Given high urbanization rates and increasing spatio-temporal variability in many present-day cities, exposure information is often out-of-date, highly aggregated or spatially fragmented, increasing the uncertainties associated with seismic risk assessments. This work therefore aims at using space-based technologies to estimate, complement and extend exposure data at multiple scales, over large areas and at a comparatively low cost for the case of the city of Bishkek, Kyrgyzstan. At a neighborhood scale, an analysis of urban structures using medium-resolution optical satellite images is performed. Applying image classification and change-detection analysis to a time-series of Landsat images, the urban environment can be delineated into areas of relatively homogeneous urban structure types, which can provide a first estimate of an exposed building stock (e.g., approximate age of structures, composition and distribution of predominant building types). At a building-by-building scale, a more detailed analysis of the exposed building stock is carried out using a high-resolution Quickbird image. Furthermore, the multi-resolution datasets are combined with census data to disaggregate population statistics. The tools used within this study are being developed on a free- and open-source basis and aim at being transparent, usable and transferable.

Keywords: remote sensing; disaster/risk management; exposure estimation; earthquakes

1. Introduction

A standard seismic risk model is composed of hazard, exposure and vulnerability modules [1]. Exposure within the context of seismic risk assessment refers to the assets or population that is at risk when an earthquake occurs. Depending on the scale of analysis, exposure data may vary from detailed descriptions of characteristics and locations of structures that may be damaged, such as buildings or transportation lines, to aggregated composite models for larger geographical entities, such as administrative units, cities or countries. When significant structural characteristics are available (e.g., construction type, building age or building height), the vulnerability of structures can be assessed using one of the different methods that have been proposed in literature [1]. The location and characteristics of structures therefore provide a crucial basis for damage and economic loss calculations in the event of an earthquake. This furthermore gives information about human exposure since inhabited structures in hazardous areas imply that people are likely to be exposed to injury or death, for example, when a building collapses [2]. Due to the high spatio-temporal variability in many present-day cities, local governments are often unable to keep track of the exposed building stock and its population in order to adjust disaster risk reduction efforts accordingly. This is especially the case in developing countries where rapid urban growth is often accompanied by unplanned settlements, which rapidly change over short periods and are moreover often highly vulnerable to natural hazards in terms of both structural and social aspects. Within this context, satellite remote sensing is increasingly being recognized as a valuable addition to established, but time- and cost-intensive, ground-based screening procedures [3] to provide exposure data for disaster risk assessment. The main benefits of a satellite-based approach are its capabilities of analyzing the elements of interest at various spatial and temporal scales, while covering large geographical areas at comparatively low costs [2]. The possibility of process automation furthermore allows for rapid analyses.

Two principle strategies for estimating exposure information can be distinguished in the literature: sampling and full enumeration. Sampling has the advantage that only small subset areas of an image need to be analyzed in detail to allow for estimating summary statistics for the entire city or for well defined strata, if a stratified sampling is used [4]. Ehrlich *et al*. [5] use a combination of texture-based image processing, statistical sampling and photo-interpretation to quantify a building stock in terms of the number of buildings, the distribution of built-up areas and building size. Full enumeration from satellite imagery refers to the detection and definition of every building within a study area. It can therefore potentially achieve high accuracies and levels of detail, but usually requires more work. However, this can be reduced by the automation of image processing chains. Approaches for localizing buildings and extracting their footprints from high-resolution satellite data range from manual digitization [6] to automated object-based classification methods [7]. The extraction and derivation of additional characteristics of the exposed building stock from remotely sensed data, such as the age of buildings, construction type or building density, enable a more refined definition of the exposed structures. In this context multi-sensor/multi-resolution approaches [8], which combine satellite images of different types and spatial resolutions, seem a promising addition to the purely high-resolution sensor-oriented approaches. For an estimation of human exposure in terms of population distribution many methods have been reported in the literature [9]. Besides bottom-up approaches, which make use of *in situ* survey data, studies on population estimation generally use ancillary data such as census

records as the primary data source. Remote sensing products (e.g., landuse/landcover classifications) can be used to disaggregate and extrapolate the census data, which are generally aggregated to artificial administrative units, to more meaningful geographical entities. A widely used approach in combination with remote sensing is the dasymetric method, which is commonly respected as a stable and accurate model that depends less on the imagery classification error than on the quality of the ancillary data used as input [9].

The aim of this study is to analyze the urban environment with respect to its composition and temporal evolution patterns, and to extract structures together with their main characteristics from multi-sensor/multi-resolution satellite imagery for the case of Bishkek, Kyrgyzstan. At a neighborhood scale, medium-resolution optical satellite images are used to delineate the urban environment into areas of relatively homogeneous urban structure types, which can provide a first estimate of the exposed building stock. The approximate age of structures is derived from a multi-temporal change-detection analysis of a time-series of Landsat images. Composition and distribution of predominant building types is extracted from a recent Landsat image using supervised image classification. At a building-by-building scale, a full enumeration of the exposed building stock using high-resolution optical satellite images is carried out. This includes a detailed delineation of structures using a semi-automatic object-based image analysis procedure supported by statistical learning. Furthermore, the remote sensing products are integrated and combined with ancillary data to derive additional exposure information and to disaggregate population statistics coming from a census report. The tools used within this study are being developed on an open-source basis to allow for a high degree of transparency, usability and transferability.

2. Study Area and Data Sets

Bishkek, the capital of Kyrgyzstan, is the largest city in the country [10] and is located at about 750 m above sea level at the northern fringe of the Kyrgyz Ala-Too mountain-range in the centre of the Chu basin. The city is placed in one of the most seismically hazardous zones in Central Asia (Figure 1(a)). The GSHAP Global Seismic Hazard Map [11] shows a peak ground acceleration (PGA) of 4 m/s ²with a probability of 10% to be exceeded in 50 years for the area of Bishkek. A seismic risk scenario carried out at a city scale for Bishkek, which considers a magnitude 7.5 earthquake occurring along the Issyk-Ata fault along the southern outskirts of the city (Figure 1(b)), estimates that 30% of the buildings will collapse and 63% of the buildings will be severely damaged [12]. Over the latest decades, Bishkek has rapidly expanded and neither up-to-date exposure nor vulnerability data are available for large parts of the city. Also, the data used in Bindi *et al*. [12] are strongly aggregated for the city as a whole and do not allow for a spatially more detailed risk assessment that would be necessary to sufficiently plan and adjust risk reduction efforts. Therefore, remote sensing methods are proposed as a way to update and extend existing exposure data and to provide a basis for improved seismic risk assessment in Bishkek.

Figure 1. (**a**) Location and seismic hazard of the study area (GSHAP [11]); (**b**) Bishkek urban area overview, showing the extent of the Landsat and Quickbird images used (background image: Landsat TM, 2009).
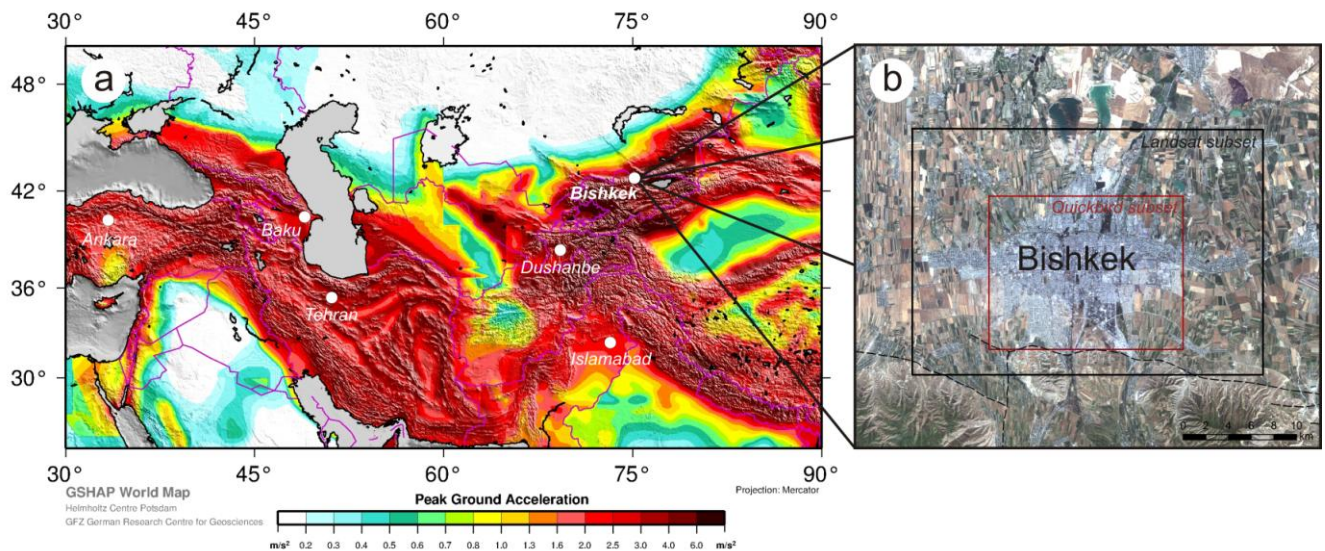

The analysis of remotely sensed data in this study is a multi-staged process using satellite sensors with different geometric, spectral and radiometric characteristics. For an analysis of urban structures at the neighborhood scale, we use images from the Landsat Thematic Mapper (TM) (0.45–12.5 μm in 7 spectral bands; 8 bit radiometric resolution; 30 m (120 m in thermal band) geometric resolution) and Multi-spectral Scanner (MSS) (0.5–1.1 μm in 4 spectral bands; 6–7 bit radiometric resolution; 60 m geometric resolution). Both satellite sensors have a large swath-width that allows a city like Bishkek to be captured within just one image. A freely accessible data archive that dates back until 1972 in combination with a revisit period of the satellites of 16 days makes the Landsat series a valuable dataset for time-series analysis. For a more detailed building-by-building analysis of the exposed building stock a Quickbird image (0.45–0.90 μm in 4 spectral bands, plus 1 panchromatic band; 11 bit radiometric resolution; 0.61 m geometric resolution for panchromatic band and 2.4 m for multi-spectral bands) covering a spatial extent of 230 km ²over the main urban area of Bishkek is used.

3. Analysis of Medium-Resolution Optical Satellite Imagery

Exposure information related to the building stock can be extracted either directly at a building-by-building scale from high-resolution satellite images, or indirectly based on the identification of the urban structure types in which the buildings are located [13]. Urban structure types are in this study defined as areas that are relatively homogeneous in medium-resolution satellite images in terms of their physical appearance (landcover) and usage (landuse) as well as their approximate age. Exposure estimates from medium-resolution satellite imagery can also be seen as a problem of estimating an unknown population [14]. In the case of exposure information, the population of interest is linked to the composition of the urban structure and will therefore most likely show great variability related to differences in landuse/landcover (LULC), age of structures and other socio-economic parameters. A stratification of the population into sub-populations as homogeneous as

possible could therefore provide at a neighborhood scale a first reasonable estimate of an exposed building stock over large areas. The derivation of urban structures from remote sensing is carried out by combining a LULC classification with the age of structures into a new thematic layer by spatial intersection. The final output of this analysis step therefore represents urban structures that are at an aggregated neighborhood scale relatively homogeneous in terms of pre-dominant LULC and approximate age. When superimposed on a high-resolution Quickbird image, it can be seen that the approach is capable of delineating areas of different urban structure in the city based on an analysis of medium-resolution satellite images (Figure 2). Descriptions of the processing steps, which have been carried out to derive urban structures, are given in the following. A more detailed description of the workflow and the different analysis steps can be found in Wieland *et al*. 2012 [4].

Figure 2. Subset of urban structure types stratification from an analysis of medium-resolution satellite images (superimposed on the Quickbird image).
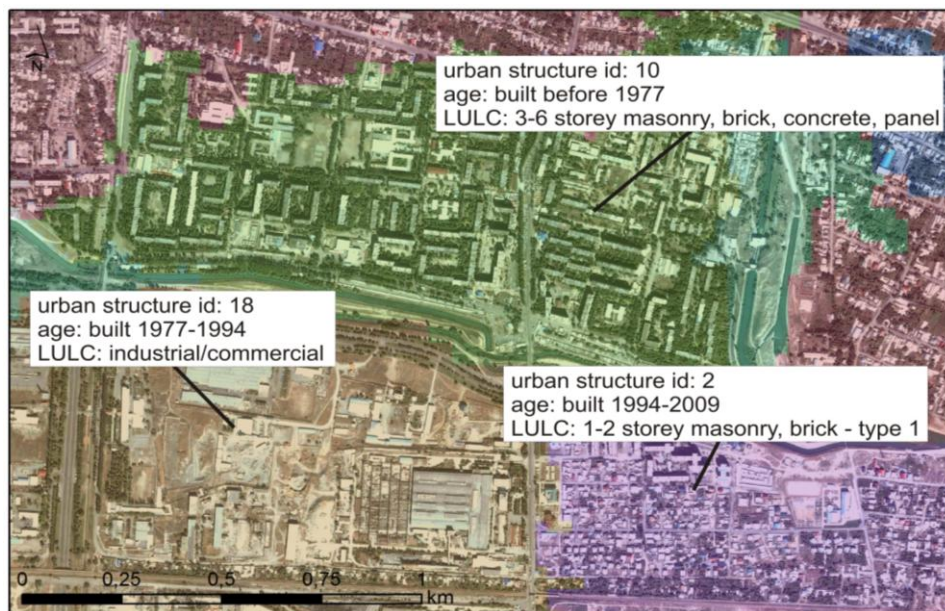

3.1. Age of Structures

To quantify changes in the urban environment and to extract information about the approximate age of structures, we apply change-detection to a time-series of Landsat TM and MSS images that cover the same spatial extent at different times (1977, 1994 and 2009). All satellite images have been geometrically corrected and co-registered to each other. Post-classification comparison was selected as the preferred change-detection method, because of the different input image types and the fact that it directly provides information about the change rate, direction and distribution [15]. For the post-classification comparison, binary LULC classifications that distinguish built-up from non-built-up areas for the three input images are generated and compared to each other on a pixel-by-pixel basis to derive a change matrix. The change matrix provides the "from-to" information of every LULC class, which allows one to identify and analyze change. The individual binary LULC classifications have been carried out using an object-based approach. The images were segmented using a graph-based segmentation algorithm [16] and automatically labeled by a Support Vector Machines (SVM)

statistical learning model. SVM is a classifier derived from statistical learning theory [17]. SVMs partition a feature space in a computationally efficient way by identifying an optimal separating hyperplane according to the properties of selected training-instances [18]. The classifier has been specifically trained to distinguish between built-up and non-built-up areas in Landsat images (Figure 3).

Figure 3. Urban-extent of Bishkek in 1977, 1994 and 2009. The binary maps have been extracted from Landsat TM and MSS images using supervised object-based image classifications.
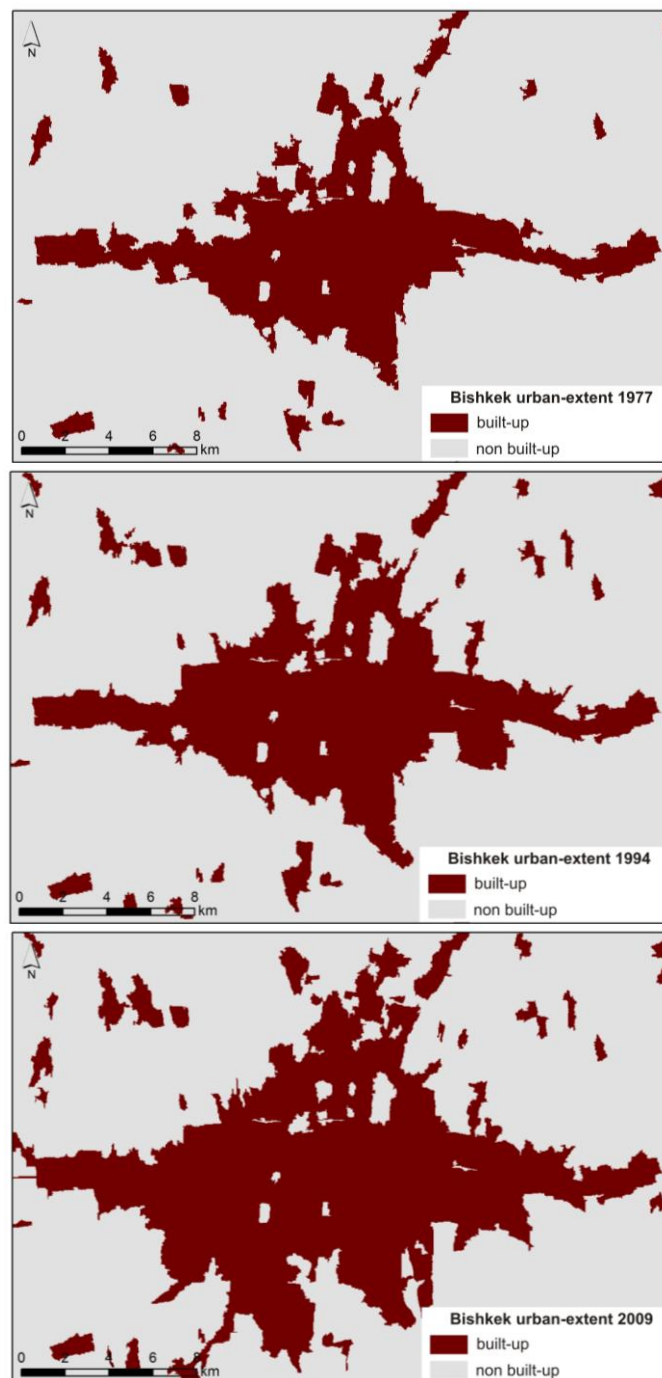
Since the accuracy of a post-classification comparison depends mainly on the accuracy of the individual LULC classifications of the input images, receiver operating characteristics curves (ROC) [19]

have been derived from error matrices for the different urban-extent classifications (Figure 4). Test-instances have been randomly selected for each classification output following a stratified sampling with proportional allocation depending on the area of the classes. The identified test-instances have been manually labeled based on visual interpretation of the input images. Differences in performance of the classifier depend both on differences in the spectral and spatial resolution between Landsat TM (2009, 1994) and MSS (1977), and on the training-instances used to train the learning machine (availability of reliable ground-truth data for 1977 is limited, for instance, with respect to more recent images).

Figure 4. ROC curves for urban-extent classifications from Landsat TM and MSS.
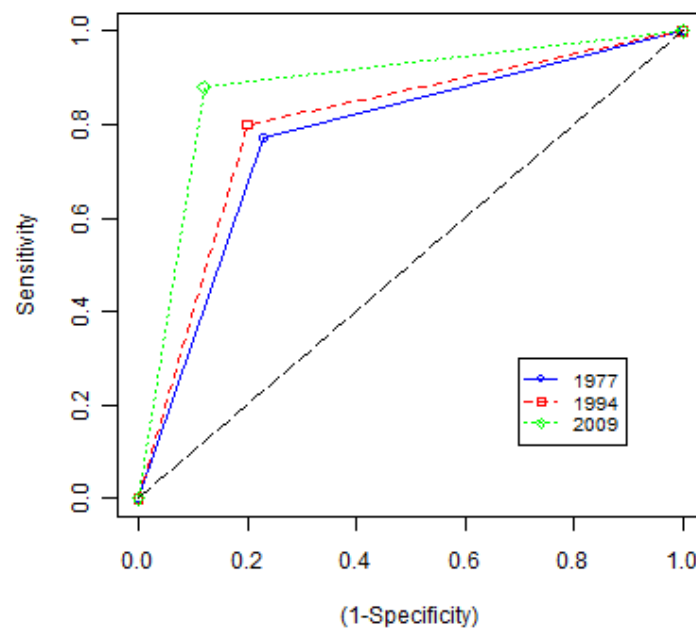
3.2. Building Types

LULC provides information about the predominant building type. An analysis of existing building-inventory data and classification schemes for the building-stock of Kyrgyzstan [20] allows for the identification of 7 built-up LULC classes of interest for Bishkek that can be identified and distinguished from medium-resolution satellite imagery (Table 1). We use a SVM statistical learning algorithm to recognize the identified LULC classes and determine decision boundaries to partition the feature space according to the properties of selected training-samples. Since we label segments instead of individual pixels, the input feature vector is composed of both spectral and textural features, which also takes into account neighborhood relationships between pixels. This is particularly important when built-up areas need to be further distinguished by their pre-dominant building types. At the medium spatial resolution of the Landsat TM, individual buildings cannot be detected in the image, but the classes of interest can be defined by the spatial alignment and composition of buildings, streets and open-spaces. Therefore, different urban LULC classes can be adequately described by a combination of spectral responses and textures at the segment level. The input feature vector, which describes any segment to be classified, consists of 26 features, including mean and standard deviation of the values of 6 spectral bands of the input image, mean and standard deviation of the Normalized Difference Vegetation Index (NDVI) and two band-specific texture descriptors derived from the Gray-Level

Co-occurrence Matrix (GLCM). Training-instances for the learning algorithm have been manually selected based on local expert knowledge on construction practices and the distribution of building types as well as using visual image interpretation supported by GPS-photos and a high-resolution satellite image of the same acquisition year. The resulting LULC classification with the distribution of pre-dominant building types for Bishkek is presented in chapter 5.

Table 1. List of identified predominant building types for Bishkek.

Class	Number of Storeys	Main Construction Materials	Description
1	1–2	Masonry, brick	Individual apartment house—detached alignment
2	1–2	Masonry, brick	Individual apartment house—attached alignment (2 sides)
3	1–2	Masonry, brick	Individual apartment house—attached alignment (3 sides)
4	3–6	Brick, concrete, panel	Multi-family apartment blocks
5	7–9	Concrete, panel, frame, monolithic	Multi-family apartment blocks
6	1–2	Brick, concrete	Industrial, commercial
7	1–9	Mixed	Mixed built-up areas
8	-	-	Non built-up area

The results of an accuracy assessment for the LULC classification of Bishkek show an overall accuracy of 81% (overall accuracy is defined as the total number of correctly classified segments divided by the total number of test-instances). A user accuracy of 50% for 7–9 storey, concrete, panel, frame, monolithic, multi-family apartment blocks (class 5) indicates a deficiency of the classifier to correctly detect this particular class. The other classes have been accurately classified with user accuracies of 70% and higher. The full error matrix is presented in Table 2. The test-instances (reference) used to create the error matrix have been randomly selected for each class following a stratified sampling with proportional allocation. The identified test-instances have been manually labeled based on the visual interpretation of the input image and a high-resolution image, as well as on the basis of GPS-georeferenced ground-based photos.

Table 2. Error matrix for LULC classification of Landsat TM (2009). A description of the classes is given in Table 1.

	Class	Reference								User's Accuracy (%)
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	
Classification	**1**	9	0	0	0	0	0	1	0	90
	2	0	8	0	1	0	0	1	0	80
	3	0	0	8	0	0	0	2	0	80
	4	0	0	0	9	1	0	0	0	90
	5	0	0	0	0	5	1	2	2	50
	6	0	0	0	0	0	9	0	1	90
	7	0	0	1	2	0	0	7	0	70
	8	0	0	0	0	0	0	0	10	100
Producer's Accuracy (%)		100	100	89	75	83	90	54	77	**81**

4. Analysis of High-Resolution Optical Satellite Imagery

Image processing applied to high-resolution satellite images is a potential source of exposure information on a building-by-building basis. The development of automated image processing chains allows for a full enumeration of the objects of interest over large areas while overcoming the time and cost constraints usually associated with manual processing. A crucial step in estimating exposure information from high-resolution satellite imagery is the extraction of a detailed built-up mask. The built-up mask provides information about the location of buildings and the area occupied by buildings. Furthermore, it can function as basis for an estimation of the number of buildings or, if it includes information about the height of structures, can be used to calculate building volume and floor space and therefore allow for a detailed disaggregation of population statistics.

4.1. Detailed Built-Up Mask

Following an object-based approach to image analysis, we deployed an automated image processing chain to extract a detailed built-up mask from the image data. In the initial stage of image analysis, the processing chain uses an efficient graph-based image segmentation [16] and combines it with a multi-scale optimization procedure to improve the delineation of building footprints of varying scales within the same image. The optimization creates a hierarchical set of segmentations and merges them into a single multi-scale segmentation based on the mean percentage difference of the weighted brightness values between sub-segments and super-segments. Optimal segmentation parameters are selected in an iterative process based on a supervised evaluation of segmentation quality using manually digitized building footprints as reference-objects. The optimization procedure is able to reduce over- and under-segmentation as well as the number of segments needed to describe buildings of different sizes. The outlined segments are labeled using a SVM statistical learning algorithm, which has been trained for 5 classes (built-up area, other sealed surfaces, bare soil, vegetation, shadows) on an extensive set of training-instances. The input feature vector used to describe the feature space is composed of spectral, textural and geometrical features. Feature selection has been carried out systematically and quantitatively using a Relief algorithm [21]. The trained SVM model is applied to each segment of the full Quickbird image of the Bishkek scene to automatically derive a LULC classification. In a post-classification stage, segments labeled as built-up are extracted from the 5 class output to derive a binary classification with built-up segments and non built-up segments. The final built-up unit mask is then derived from a connected components analysis of the built-up segments. This means that built-up units are composed of connected built-up segments. Figure 5 shows the resulting detailed built-up unit mask for Bishkek as well as the underlying built-up segments and the original input image for the classification.

An accuracy assessment has been carried out on the resulting built-up mask using 978 manually digitized and evenly distributed building footprints to select the positive test-instances (built-up). Proportional to the area of the built-up and non built-up classes, 4986 negative test-instances (non built-up) have been randomly selected. The resulting built-up unit mask proves to be highly accurate with an overall accuracy of 90.24%, which equals 5283 correctly classified test-instances out of 5964 (Figure 6 and Table 3).

Figure 5. Results of the analysis of high-resolution satellite imagery. (**a**) Built-up units for 2009 in Bishkek and magnification of the rectangular area of (**b**) the built-up units, (**c**) the built-up segments composing the built-up units, (**d**) the input Quickbird image from 2009.
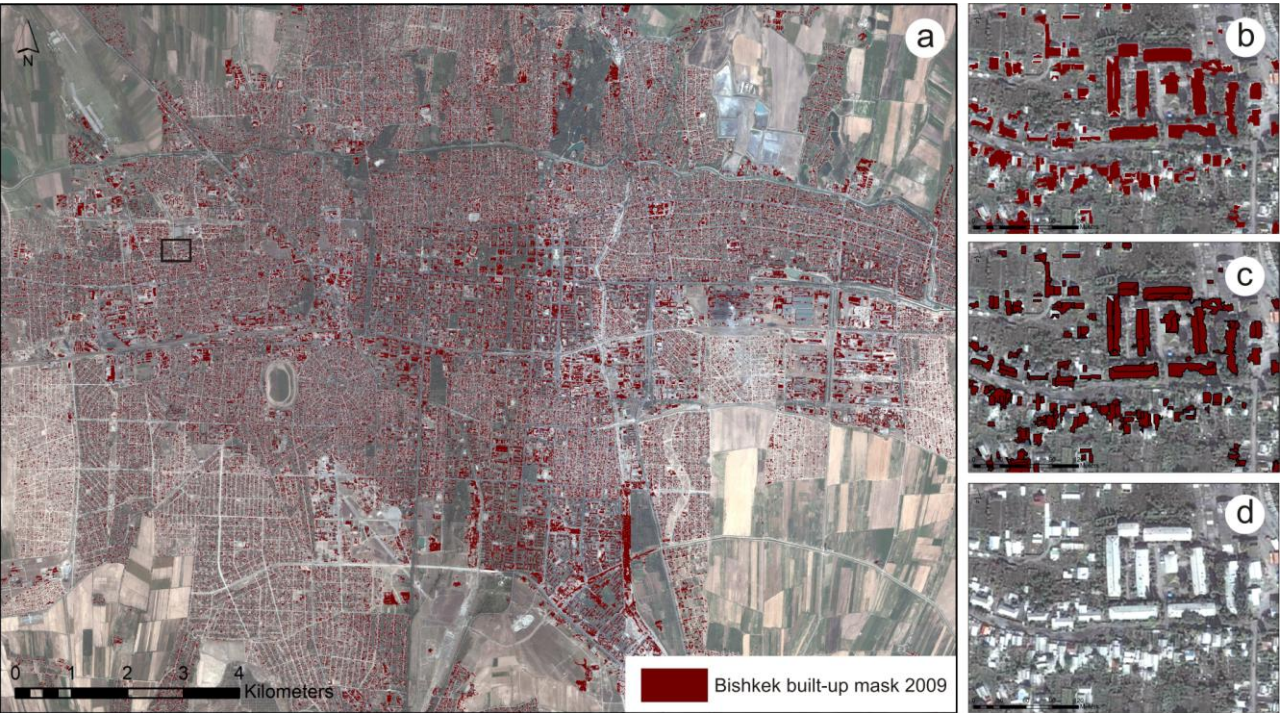

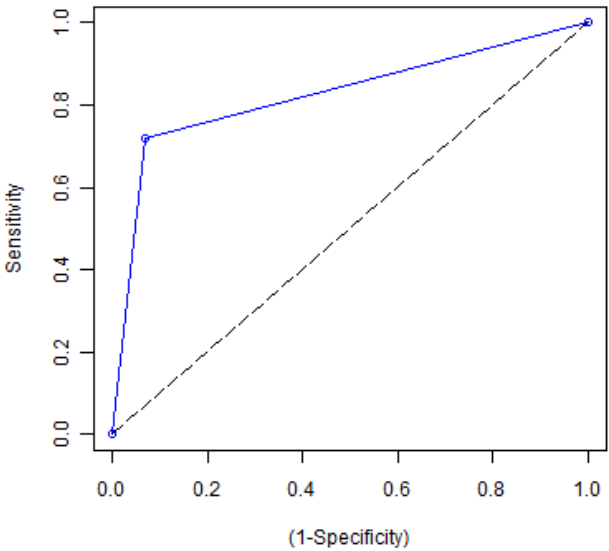**Figure 6.** ROC curve for high-resolution built-up mask, which has been extracted from Quickbird.

Table 3. Error matrix for the built-up mask, which has been extracted from Quickbird.

| Class		Reference		
		Built-up	**Non built-up**	**User's Accuracy (%)**
Classification	**Built-up**	706	272	72.19
	Non built-up	310	4,676	93.78
	Producer's Accuracy (%)	69.49	94.50	**90.24**

4.2. Number of Buildings

Even though the detailed built-up mask that has been generated is sufficiently accurate for delineating built-up segments, it does not directly provide the exact footprints for all buildings and can therefore not be directly used to count the number of buildings. As can be seen in Figure 7(a), the algorithm is able to accurately outline and detect detached single buildings, even in complex urban settings with high building densities, single buildings can be detected and outlined by the algorithm (Figure 7(b)). However, we observe that the number of built-up segments is correlated with the size of buildings and larger buildings are composed of a larger number of built-up segments (Figure 7(c)). In these cases, a simple count of the built-up segments would lead to an overestimation of the number of actual buildings. Counting the built-up units, on the other hand, would lead to an underestimation of the number of buildings, due to the clustering of attached buildings to larger units in more complex urban settings. Given these observations, we empirically derived for each building type a conversion factor to estimate the actual number of buildings from the number of built-up segments. Therefore, we used a set of 1950 manually digitized reference-buildings with an equal proportion for each building type and compared for each building type separately the number of built-up segments per reference-building. To calculate the actual number of buildings the conversion factors are applied to all built-up units that consist of more than one built-up segment. The built-up units, which are not further subdivided by built-up segments, can be assumed to be single detached buildings and can therefore be counted separately as individual buildings without applying a conversion factor.

Figure 7. (**a**) Detached single buildings detected and outlined by built-up mask; (**b**) Appearance of complex urban settings with attached buildings in built-up mask; (**c**) Large buildings in built-up mask.

Figure 8 shows the results of a comparison between the estimated number of buildings from the built-up mask and 15,098 manually digitized buildings for 49 test-areas. The test-areas were selected following a stratified sampling approach with proportional allocation using the urban structure types as strata [4]. They are therefore evenly distributed over the whole study area and are representative for different urban structure types. A comparison between the estimated number of buildings and the real number of buildings in the test-areas shows a good fit (Figure 8). In total, 15,031 buildings are estimated for the test-areas by applying the conversion factors to the detailed built-up mask. In comparison to the total number of buildings from manual digitization, the automatic procedure slightly under-estimates the total by 67 buildings. Linear regression indicates that 94% of the variance of the real number of buildings in the 49 test-areas was correctly estimated by the proposed procedure.

Figure 8. Comparison of the estimated number of buildings from the built-up mask with the real number of buildings in 49 test-areas from manual digitization. Test-areas have been selected following a stratified sampling approach using the urban structure types as strata.
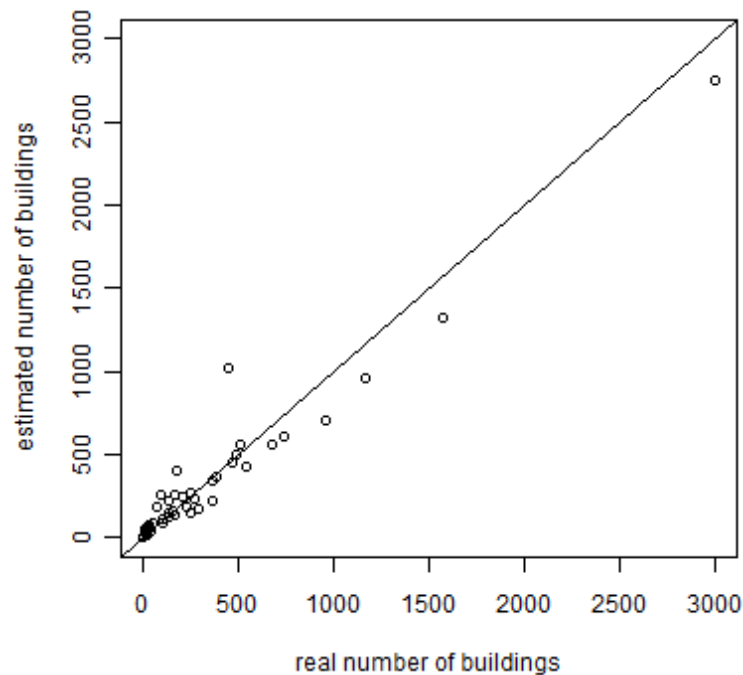
4.3. Spatial Disaggregation and Estimation of Population Statistics

Earthquakes may damage the building stock in a selective way, depending on site-specific local ground-motion amplification [22] or the structural vulnerability of buildings and their distribution. Therefore, only fine scale population datasets can provide an accurate estimate of the population actually exposed in the event of an earthquake. Census information is usually aggregated in administrative units such as city districts, which are often too large to sufficiently take into account local ground-motion amplifications. Administrative units furthermore represent artificial boundaries that are not necessarily related to the actual urban structure and therefore do not allow for a more detailed link between the structural vulnerability of the building stock and its population. However, it is possible to spatially disaggregate the census information from a set of larger spatial units (source zones) to smaller spatial units (target zones) using remote sensing products such as LULC classifications or detailed built-up masks as a proxy for population presence [23]. Using a detailed built-up mask as proxy for population presence has the benefit that the built-up area and its distribution is clearly outlined and quantifiable (and not aggregated as, for example, in medium-resolution LULC classifications) and therefore allows for the direct calculation of population density.

In this study we use population data from the Population and Housing Census of Bishkek carried out in 1999 by the National Statistical Committee of the Kyrgyz Republic [10]. The data refer to the resident population and are aggregated to administrative units (source zones), but are available only for parts of the urban area of interest. A spatial disaggregation is carried out for the areas where census data are available using the built-up units of the detailed built-up mask as target zones. For each built-up unit the most probable building type and number of storeys can be inferred from the urban structures layer based on the spatial location. Taking the average number of storeys in combination with

the built-up area derived from the high-resolution image analysis, it is possible to compute for each built-up unit i the average floor area FA_i. The population density D_j can be calculated for each source zone j, which is defined as the administrative unit that fully overlaps with the detailed built-up mask and for which there is the total population P_j from the census report available, using

$$D_j = \frac{P_j}{\sum_{i \in J} FA_i} \tag{1}$$

Applying Equation (2) to each built-up unit in a source zone provides a proportional spatial disaggregation of the total population of the source zone to the built-up units. It provides the average population per built-up unit P_i as a product of floor area FA_i and population density D_j of the specific source zone where $i \in J$ and,

$$P_i = D_j * FA_i \tag{2}$$

Population density and average population per built-up unit can be disaggregated from administrative units to built-up units for areas where census data are available. To estimate the population density of built-up units for parts of the city where no or incomplete population statistics are available, a regression model is deployed, using FA_i and P_i as independent and dependent variables, respectively. A total of 70,768 built-up units, which overlap with the source zones and for which population numbers could be directly disaggregated from the census data, were used as input for the regression. This means that the population over all the source zones is used in the regression, therefore providing an estimation of the average population per floor area, which is independent of a specific source zone. With the regression model, 79% of the variance of the dependent variable can be explained for the 70,768 built-up units of the overlapping areas. Using the regression function, the population per built-up unit can also be estimated for the 42,088 built-up units not covered by census data.

In this study only residential areas have been taken into account for population estimation, because the census data refers to the resident population. This provides information about the total estimated number of inhabitants per built-up unit of the residential areas and can therefore be seen as an estimate of the night-time population distribution in the study area. The day-time population distribution, at least during working hours, may differ from the night-time estimate due to the commuting of the working population between residential areas and industrial/commercial areas. Since information about the percentage of working people per source zone is currently not available, these population dynamics could not be taken into account at this stage.

5. Results

The analysis of multi-resolution satellite images allowed for an automated estimation of the physical and human exposure in the study area. Exposure characteristics extracted from remote sensing in this study include predominant building types, building footprints, number and distribution of buildings, number and distribution of population and approximate age of structures. For the study area, covered by the Quickbird image (Figure 1(b)), a total of 112,293 buildings with 847,639 inhabitants have been estimated (Table 4). With 77% of the building stock (86,842 buildings), 1–2 storey masonry, brick individual houses (including three sub-types depending on the spatial alignment of the buildings) are clearly the dominating building type in the study area. 3–6 storey brick, concrete, panel

multi-family building-blocks account for 8% of the building stock (8,469 buildings) and form the second most widespread residential building type followed by 7–9 storey concrete panel, frame and monolithic building-blocks with 2% of the total building stock (2,271 buildings). Industrial and commercial buildings account for 5% (5,583 buildings) of the buildings in Bishkek. The remaining 8% of the building stock (9,128 buildings) could not be attributed to a specific class based on remote sensing, and were classified as mixed building types (Figure 9). Despite the clear dominance of 1–2 storey masonry, brick individual houses in the study area, the population is equally well distributed between this building type (292,207 people) and 3–6 storey brick, concrete, panel building-blocks (288,030 people). Focusing only on residential buildings for population estimates, 13% of the population (107,936 people) are furthermore estimated to live in 7–9 storey concrete, panel, frame and monolithic buildings and 19% (159,466 people) live in mixed building types (Figure 9).

Table 4. Estimated number of buildings and population per building type for the Quickbird subset in 2009 (see Figure 1(b)).

Building Type	Estimated Number of Buildings	Estimated Population
1–2 storey masonry, brick individual house, type 1,2,3	86,842	292,207
3–6 storey brick, concrete, panel multi-family block	8,469	288,030
7–9 storey concrete, panel, frame, monolithic block	2,271	107,936
1–2 storey brick, concrete industrial, commercial	5,583	-
1–9 storey mixed built-up	9,128	159,466
TOTAL	**112,293**	**847,639**

Figure 9. Composition of the number of buildings and population per building type for the Quickbird subset in 2009 (see Figure 1(b)).
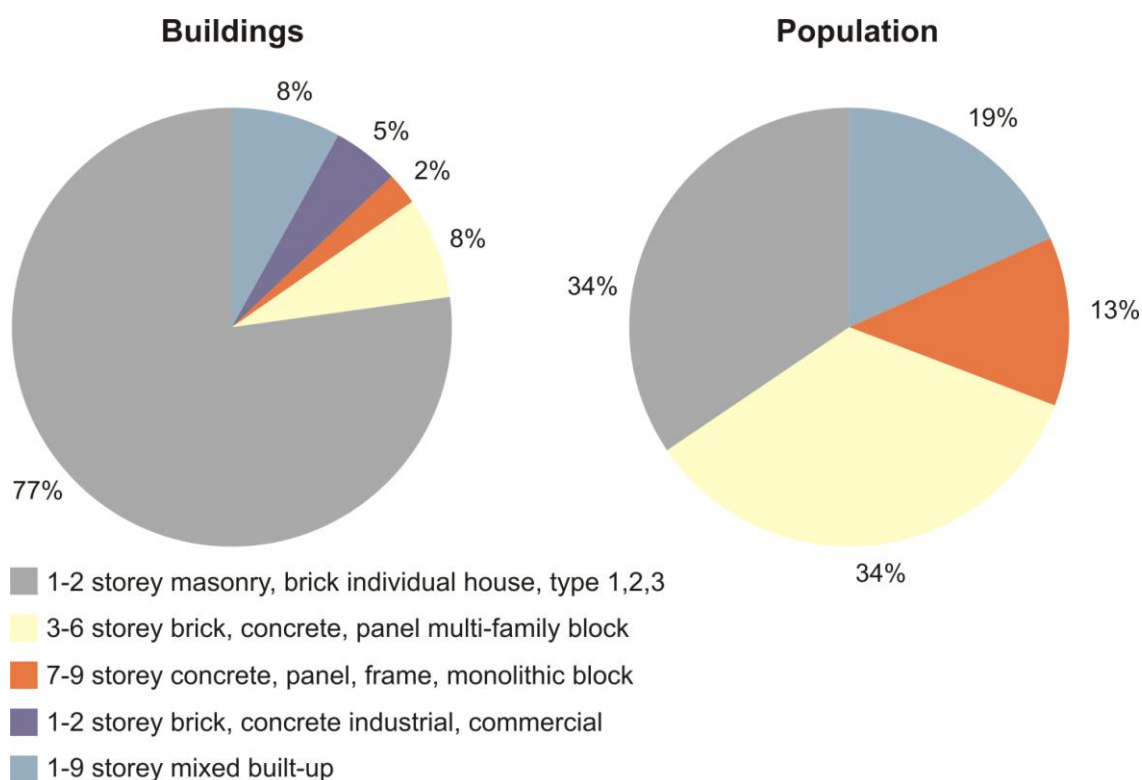
Figure 10. LULC classification of Bishkek showing the distribution of pre-dominant building types in 2009 (see Tables 1 and 4). The numbers are the estimated numbers of buildings per stratum.
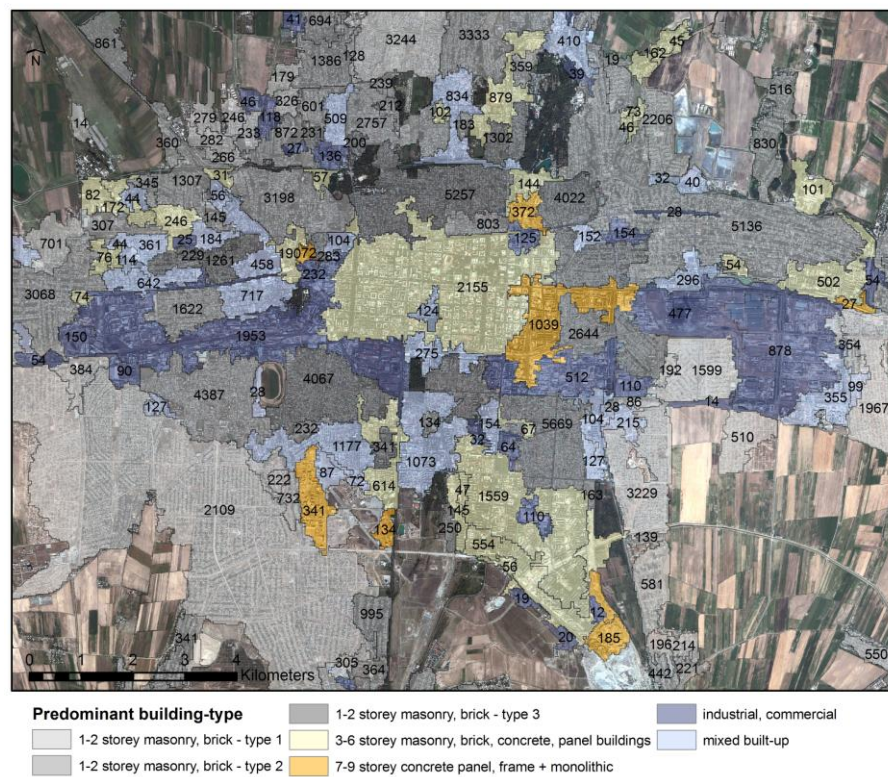

Figure 11. Population density per stratum in Bishkek. The numbers are the estimated population. Industrial areas are excluded.
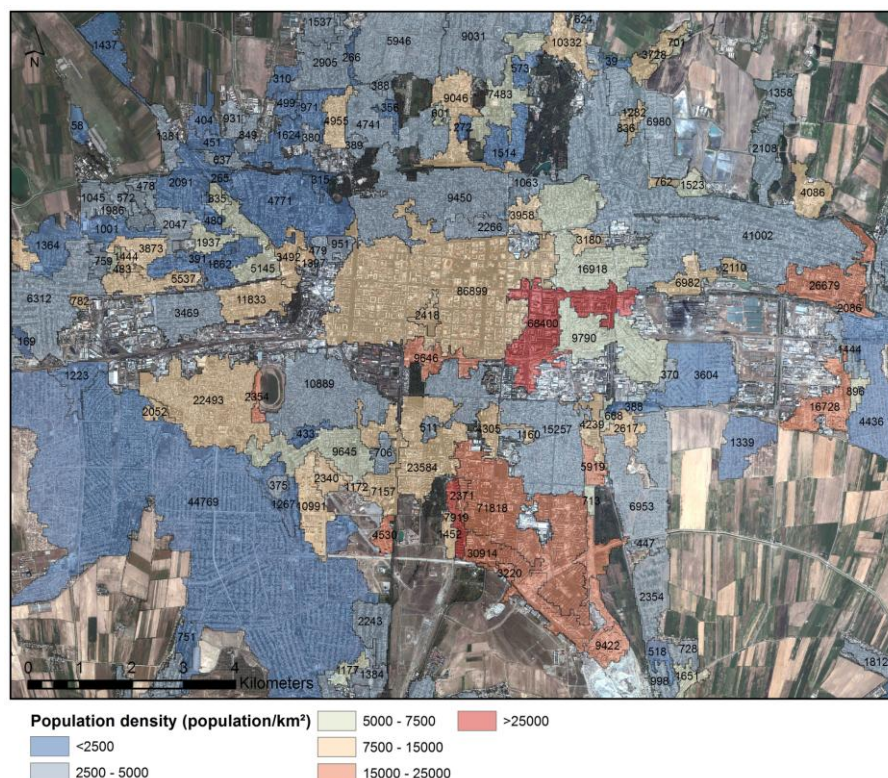
Figure 10 shows the distribution of the different building types in the study area in 2009. The stratification has been derived from medium-resolution satellite images. For each stratum the number of buildings has been calculated from the detailed built-up mask extracted from the high-resolution satellite image. Furthermore, population data have been aggregated from the detailed built-up mask to derive population density and the estimated number of residents per stratum (Figure 11).

An analysis of a time-series of Landsat images provides a temporal dimension of the exposure analysis and gives information about the approximate age of structures, change rate and change directions. Combined with the outputs of the high-resolution analysis, more detailed conclusions can be drawn about the spatio-temporal variability in the city, including information about changes in the number of buildings and the population. Of the building stock under observation, 77,292 buildings were classified as being built before 1977, with an estimated 579,594 inhabitants living in these buildings in 2009. A total of 93,497 buildings were classified as being built between 1977 and 1994, with 695,572 inhabitants. 112,293 buildings could be identified as being built between 1994 and 2009, housing 847,639 people. Looking at the spatio-temporal patterns of the urban sprawl (Figure 12), the urban expansion between 1977 and 1994 concentrated mainly on the suburban eastern and northern parts of the city. From 1994, three years after the Kyrgyz Republic became independent of the former Soviet Union, to 2009, the city of Bishkek expanded rapidly in the southern parts towards the Issyk-Ata fault. These areas close to the fault system, where many new buildings are still being constructed, show the highest seismic hazard in the study area [24]. Large quarters have also been built recently in the far northern parts of the city.

Figure 12. Spatio-temporal development of the building stock in Bishkek from 1977 to 2009. The approximate age of structures has been derived from a change-detection analysis of Landsat images and joined to the detailed built-up mask extracted from a Quickbird image.
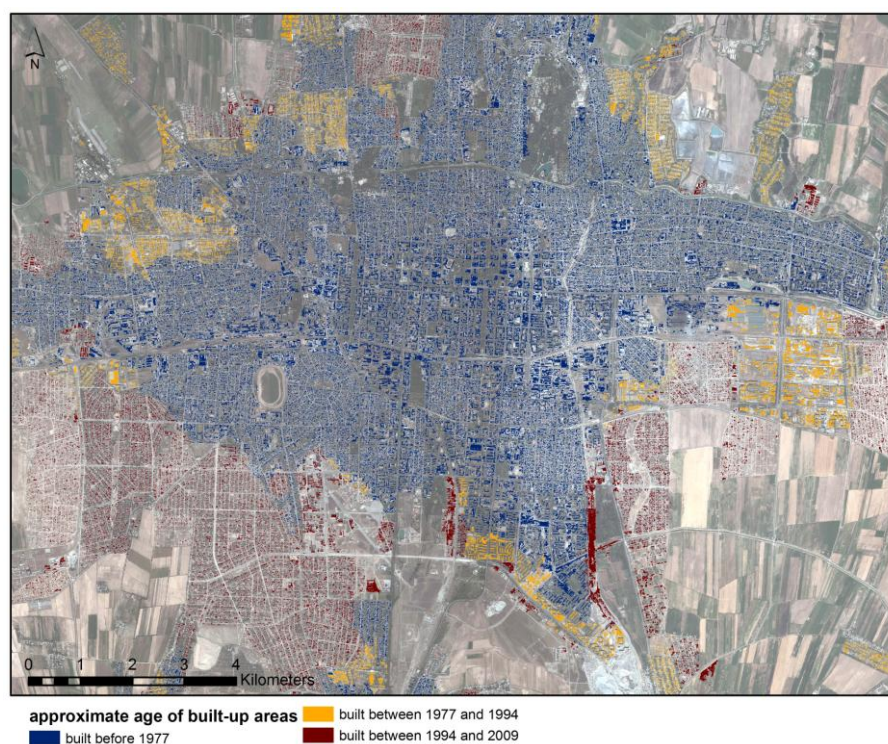
6. Discussion

Using satellite images as the basis for information extraction allows for a rapid exposure estimation over large areas at comparatively low costs. Apart from its benefits, a purely satellite based approach shows limitations, in that it can only provide information about exposure characteristics that can be assessed from the top view. For example, the building height, which is one of the most significant features for the assessment of the vulnerability of a particular structure, can be estimated from a single high-resolution optical satellite image only under specific conditions using shadow information [25]. Other vulnerability relevant features that cannot directly be analyzed for a single building from satellite imagery include detection of soft-storeys or vertical irregularities. Therefore, to create a basis for a more detailed assessment of the structural vulnerability and occupancy of an exposed building stock, an analysis of the facades of buildings from a street view should be taken into account. Using rapid visual ground-based surveys, such as omnidirectional imaging, within the framework of an integrated sampling scheme together with the proposed satellite-based approach, can further enhance the description of the exposed building stock, both in terms of the level of detail and accuracy, while still keeping a reasonable cost-benefit ratio [4]. Especially when integrating data coming from different sources with varying accuracies, uncertainties should be adequately taken into account. A transformation from a deterministic to a probabilistic description of exposure elements can be achieved by the use of Bayesian networks [26]. A probabilistic description of the exposure model and its vulnerability would therefore allow, in combination with probabilistic seismic hazard analysis (PSHA), for a fully probabilistic seismic risk assessment. The proposed methods and tools in this study have been developed on a free and open-source basis. Using free and open-source software (FOSS) can further improve the capabilities of less equipped institutions in developing countries, which are often especially vulnerable to natural hazards, to carry out risk-preventive analysis by themselves and therefore allow for a sustainable development in the sense of the United Nations International Strategy for Disaster Reduction (UN-ISDR). An important global initiative in this context is the Global Earthquake Model (GEM), which aims at providing open tools and standards to calculate and communicate earthquake risk worldwide [27].

7. Conclusions

This paper proposes an approach to analyze an urban environment with respect to its composition and temporal evolution patterns and to extract structures, together with their main characteristics, based on multi-sensor/multi-resolution satellite imagery. The use of image segmentation as an initial step in an object-based approach to image processing provided the basis for a rich description of the objects of interest in terms of the dimensionality of feature vectors. In combination with quantitative feature selection and SVM statistical learning a detailed differentiation of the urban environment, even from medium-resolution satellite images, proved feasible. The use of statistical learning models allowed on the one hand for a high degree of process automation while on the other, it showed great flexibility and potential for the transferability of the method to other areas and/or image types. Local expert-knowledge can easily be integrated into the training-phase of the SVM learning models and the

selection of training-instances is a straight forward process that can be adjusted depending on the input image type and study area.

For the city of Bishkek, the approach could provide a detailed spatio-temporal description of the exposed building stock and its population that can be used as input for seismic vulnerability and risk assessments. The results clearly show a dominance of 1–2 storey masonry, brick individual houses, which are mainly classified by local engineers as highly vulnerable (class A to B following the EMS-98 vulnerability classification [28]). Moreover, one third of the population live in this specific building type. Change-detection analysis furthermore shows that, especially in recent years, a majority of the newly constructed buildings of this highly vulnerable type expand towards the most hazardous areas in the region. This indicates a trend towards increasing seismic risk for Bishkek and emphasizes the need for a continuous update of the exposure and vulnerability datasets for the city.

Current research work focuses on a refinement and extension of the exposure data using ground-based omnidirectional imaging and its integration into a probabilistic risk assessment framework by the use of Bayesian networks [26]. In this context it is planned to derive a probabilistic vulnerability map for Bishkek and other Central Asian towns and to combine it with the results of seismic hazard analysis [22] to update and improve risk scenarios [12] in a probabilistic framework.

Acknowledgements

References

1. Calvi, G.M.; Pinho, R.; Magenes, G.; Bommer, J.J.; Restrepo-Velez, L.F.; Crowley, H. Development of seismic vulnerability assessment methodologies over the past 30 years. *ISET J. Earthq. Technol.* **2006**, *43*, 75–104.
2. Deichmann, U.; Ehrlich, D.; Schmal, C.; Zeug, G. *Using High Resolution Satellite Data for the Identification of Urban Natural Disaster Risk*; Global Facility for Disaster Reduction and Recovery: Washington, DC, USA, 2011.
3. Federal Emergency Management Agency. *Rapid Visual Screening of Buildings for Potential Seismic Hazards: A Handbook*, 2nd ed.; FEMA 154; ATC: Washington, DC, USA, 2002.
4. Wieland, M.; Pittore, M.; Parolai, S.; Zschau, J.; Moldobekov, B.; Begaliev, U. Estimating building inventory for rapid seismic vulnerability assessment: Towards an integrated approach based on multi-source imaging. *Soil Dyn. Earthq. Eng.* **2012**, *36*, 70–83.
5. Ehrlich, D.; Zeug, G.; Gallego, J.; Gerhardinger, A.; Caravaggi, I.; Pesaresi, M. Quantifying the building stock from optical high-resolution satellite imagery for assessing disaster risk. *Geocarto Int.* **2010**, *25*, 281–293.

6.   Voogt, J.A. Image representations of complete urban surface temperatures. *Geocarto Int.* **2000**, *15*, 21–31.

7.   Taubenböck, H.; Esch, T.; Wurm, M.; Roth, A.; Dech, S. Object-based feature extraction using high spatial resolution satellite data of urban areas. *J. Spat. Sci.* **2010**, *55*, 117–132.

8.   Wurm, M.; Taubenböck, H.; Dech, S. Quantification of urban structure on building block level utilizing multisensoral remote sensing data. *Proc. SPIE* **2010**, *7831*, 78310H.

9.   Wu, S.; Qiu, X.; Wang, L. Population estimation methods in GIS and remote sensing: A review. *GISci. Remote Sens.* **2005**, *42*, 58–74.

10. National Statistical Committee of the Kyrgyz Republic. *Population and Housing Census of the City of Bishkek*; Bishkek City: Bishkek, Kyrgyzstan, 2001.

11. Shedlock, K.; Giardini, D.; Grünthal, G.; Zhang, P. The GSHAP global seismic hazard map. *Seismol. Res. Lett.* **2000**, *71*, 679–686.

12. Bindi, D.; Mayfield, M.; Parolai, S.; Tyagunov, S.; Begaliev, U.; Abdrakhmatov, K.; Moldobekov, B.; Zschau, J. Towards an improved seismic risk scenario for Bishkek, Kyrgyz Republic. *Soil Dyn. Earthq. Eng.* **2011**, *31*, 521–525.

13. Müller, M.; Segl, K.; Heiden, U.; Kaufmann, H. Potential of high-resolution satellite data in the context of vulnerability of buildings. *Nat. Hazards* **2006**, *38*, 247–258.

14. Cochran, W. *Sampling Techniques*, 3rd ed.; Wiley: New York, NY, USA, 1977.

15. Lu, D.; Mausel, P.; Brondizio, E.; Moran, E. Change detection techniques. *Int. J. Remote Sens.* **2004**, *25*, 365–401.

16. Felzenzwalb, P.; Huttenlocher, D. Efficient graph-based image segmentation. *Int. J. Comput. Vis.* **2004**, *59*, 67–81.

17. Vapnik, V. *The Nature of Statistical Learning Theory*, 2nd ed.; Springer: New York, NY, USA, 2000.

18. Burges, C.J.C. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* **1998**, *2*, 121–167.

19. Bradley, A. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recogn.* **1997**, *30*, 1145–1159.

20. *World Housing Encyclopedia*. Available online: http://www.world-housing.net (accessed on 10 March 2012).

21. Koprinska, I. Feature Selection for Brain-Computer Interfaces. In *New Frontiers in Applied Data Mining*; Theeramunkong, T., Nattee, C., Adeodato, P.J.L., Chawla, N., Christen, P., Lenca, P., Poon, J., Williams, G., Eds.; Springer: Heidelberg, Germany, 2010; pp. 100–111.

22. Parolai, S.; Orunbaev, S.; Bindi, D.; Strollo, A.; Usupaev, S.; Picozzi, M.; Di Giacomo, D.; Augliera, P.; D'Alema, E.; Milkereit, C.; Moldobekov, B.; Zschau, J. Site effects assessment in Bishkek (Kyrgyzstan) using earthquake and noise recording data. *Bull. Seismol. Soc. Am.* **2010**, *100*, 3068–3082.

23. Flowerdew, R.; Green, M. Data Integration: Statistical Methods for Transferring Data between Zonal Systems. In *Handling Geographical Information: Methodology and Potential Applications*; Masser, L., Blakemore, M., Eds.; Longman: Harlow, UK, 1991, pp. 38–54.

24. Erdik, M.; Rashidov, T.; Safak, E.; Turdukulov, A. Assessment of seismic risk in Tashkent, Uzbekistan and Bishkek, Kyrgyz Republic. *Soil Dyn. Earthquake Eng.* **2005**, *25*, 473–486.

25. Shao, Y.; Taff, G.N.; Walsh, S.J. Shadow detection and building-height estimation using IKONOS data. *Int. J. Remote Sens.* **2011**, *32*, 6929–6944.

26. Pittore, M.; Wieland, M. Towards a rapid probabilistic seismic vulnerability assessment using satellite and ground-based remote sensing. *Nat. Hazards* **2012**, submitted.

27. *Global Earthquake Model*. Available online: http://www.globalquakemodel.org (accessed on 15 March2012).

28. Grünthal, G.; Musson, R.; Schwarz, J.; Stucchi, M. *European Macroseismic Scale 1998 (EMS-98)*; Cahiers du Centre Européen de Géodynamique et de Séismologie 15, Centre Européen de Géodynamique et de Séismologie: Luxembourg, 1998; p. 99.