Investigating the Applicability of Cartosat-1 DEMs and Topographic Maps to Localize Large-Area Urban Mass Concentrations

Michael Wurm, Pablo d'Angelo, Peter Reinartz, *Member, IEEE*, and Hannes Taubenböck

***Abstract*—Building models are a valuable information source for urban studies and in particular for analyses of urban mass concentrations (UMCs). Most commonly, light detection and ranging (LiDAR) is used for their generation. The trade-off for the high geometric detail of these data is the low spatial coverage, comparably high costs and low actualization rates. Spaceborne stereo data from Cartosat-1 are able to cover large areas on the one hand, but hold a lower geometric resolution on the other hand. In this paper, we investigate to which extent the geometric shortcomings of Cartosat-1 can be overcome integrating building footprints from topographic maps for the derivation of large-area building models. Therefore, we describe the methodology to derive digital surface models (DSMs) from Cartosat-1 data and the derivation of building footprints from topographic maps at 1:25 000 (DTK-25). Both data are fused to generate building block models for four metropolitan regions in Germany with an area of $\sim 16\,000\ \mathrm{km}^2$. Building block models are further aggregated to 1×1 km grid cells and volume densities are computed. Volume densities are classified to various levels of UMCs. Performance evaluation of the building block models reveals that the building footprints are larger in the DTK-25, and building heights are lower with a mean absolute error of 3.21 m. Both factors influence the building volume, which is linearly lower than the reference. However, this error does not affect the classification of UMC, which can be classified with accuracies between 77% and 97%.**

***Index Terms*—Building model, cartosat, digital elevation models, topographic maps.**

I. Introduction

In SPATIAL urban research, knowledge on the spatial distribution and the quantity of the urban area and urban objects is of fundamental importance. The urban mass concentration (UMC) or volume density (VD) is a key variable in the physical characterization of a city's shape and structure [1]–[3]. Identification and discrimination of these urban masses can be performed using earth observation data. Remote sensing offers the possibility for a large-area and cost-effective derivation of an inventory of urban features. Recent developments in sensor and image analysis technology exhibit a wide spectrum of data acquisition options at various spatial resolutions available. Naturally, for the quantification of urban masses, three-dimensional (3-D) data are of particular relevance. For the creation of 3-D data, height information is inevitable.

In remote sensing, height information in urban areas is derived from stereoscopic aerial images [4]–[6], satellite-based stereo images [6]–[9], or spaceborne SAR data [10], [11]. Furthermore, light detection and ranging (LiDAR) represents a traditional data set, used to derive building models [12]–[16]. LiDAR data can be used to automatically derive physical parameters of urban areas [17], but it underlies the spatial limitations of airborne platforms and is therefore subject to comparably high costs in comparison with spaceborne data acquisition. The spatial limitation, however, also applies to satellite-based stereo images from very high resolution (VHR) sensors like, e.g., IKONOS, Quickbird, WorldView-2, or Pléiades. The shortcomings for digital surface model (DSM) generation from these satellites in urban studies are the high efforts needed for the acquisition of a stereo pair or triplet with appropriate viewing geometry. Furthermore, the spatial coverage of the resulting DSM is limited and dependent on the swath width of the satellite images, which varies between 11 and 20 km. Thus, these data are not applicable for large-area stereo mapping.

In the current analysis, we aim at obtaining large-area information of urban masses for the detection of physical polycentricity of urban regions. Physical polycentricity relates to an increased concentration of urban masses and can be derived from building volume densities or the floor space index [18], [19]. A thorough investigation of the degree of polycentricity implies the analysis of very large regions such as entire cities or metropolitan areas including their sub- and peri-urban and also rural surroundings. This results in very large investigation areas, exceeding the spatial coverage of traditional data sets by a multiple. Besides the large area, a further requirement for the data for such analysis is the high level of detail in terms of individual buildings including their volume. In our case, the aim is to investigate and compare the degree of physical polycentricity of four metropolitan areas in Germany. These areas include: Munich, Stuttgart, Frankfurt, and Cologne (Fig. 1).

The spaceborne sensor Cartosat-1 (IRS-P5) was designed particularly for large-scale stereo mapping. It provides stereo images with 2.5 m pixel spacing at a swath width of about 27 km [20] and data can be exploited to generate DSMs with

Manuscript received November 12, 2013; revised July 22, 2014; accepted July 28, 2014. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) under Grant TA 800/1-1 and under Grant SI 932/4-1.

M. Wurm and H. Taubenböck are with the Deutsches Zentrum für Luft- und Raumfahrt (DLR), Deutsches Fernerkundungsdatenzentrum (DFD), Wessling 82234, Germany (e-mail: michael.wurm@dlr.de; hannes.taubenboeck@dlr.de).

P. d'Angelo and P. Reinartz are with the Deutsches Zentrum für Luft- und Raumfahrt (DLR), Remote Sensing Technology Institute (IMF), Wessling 82234, Germany (e-mail: pablo.dangelo@dlr.de; peter.reinartz@dlr.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2014.2346655

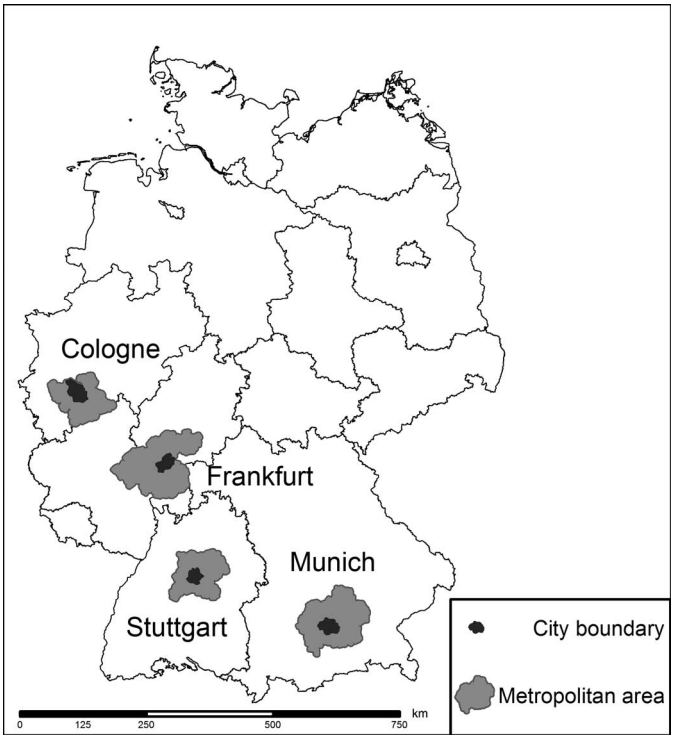

Fig. 1. Spatial coverages of the investigated metropolitan areas: Munich, Stuttgart, Frankfurt, and Cologne/Bonn. The area covered by the Cartosat-1 data sums up to a total of almost 16 000 km^2 (Stuttgart: 3010 km^2; Munich: 4698 km^2; Frankfurt: 5398 km^2; Cologne: 2836 km^2).

5 m spacing [21]. Since the launch of Cartosat-1 in May 2005, large parts of the Earth have been mapped with stereo images, including an almost entire coverage of Europe [22]. Therefore, the availability, spatial, and temporal coverage (5 days revisiting rate) meet the theoretic requirements for an area-wide and cost-effective derivation of height information of very large urban areas. A limited number of studies have investigated the accuracy of Cartosat-1 DSMs for building height detection. These studies retrieve errors between 4.14 m root mean square error (RMSE) and standard deviation (SD) of 3.59 m [23], a mean error of 1.88 m with an SD between 2.1 and 2.6 m [6] and a mean absolute error (MAE) of 2.76 m and an SD of 3.04 m [24]. The latter, however, investigates not individual buildings but deviations of the DSM over urban areas compared to a LiDAR reference. These papers have compared height accuracies from Cartosat-1 DSMs based on a small number (less than 200) of reference samples. However, no large-area assessment of the suitability of Cartosat-1 DSMs for the quantification of urban masses has been conducted. Summing up, point spacing of 5 m of Cartosat-1 DSMs leads to a lower building height and not satisfactory preconditions for precise delineation and reconstruction of buildings [6], [21].

In this context, image information and data fusion allow for data enhancement through combination of different data sources for the characterization of complex urban environment [25]. Against this background, we fuse height information from Cartosat-1 DSMs with building footprints extracted from topographic maps for the generation of a large-area 3-D building model. Extraction of building footprints from topographic maps is a valuable technique for the generation of an area-wide data source of building footprints. Image segmentation methods [26] can be applied to extract objects representing the building footprints from digital raster maps at a scale of 1:25 000 [27], [28]. 3-D building reconstruction is performed with standard information fusion techniques. For the combination of building footprints and height data, mostly from LiDAR, a vast number of techniques exists, e.g., [16], [29]–[33]. The described techniques focus on complex building and roof reconstruction. Our approach, however, has the focus on solving a gap for large-area applications of Cartosat-1 DSMs and thus presenting and evaluating a workflow to derive an estimate of building volume densities and subsequently the classification of UMCs for entire metropolitan areas.

Therefore, the objective of this study is to evaluate the performance and overall suitability of fully automatic derivation of UMCs. In this perspective, the output of the procedure is tested within a thorough validation exercise firstly on the derived building block models and secondly on the derived UMCs. The major contribution of this article is to present a simple, fast and cost-effective method which allows the spatial structuring of entire metropolitan areas into areas of very high, high, medium, low and very low UMCs.

The remainder of this article is organized as follows. Section II introduces the study area and data collection and in Section III, methods for the derivation of UMCs are described. In Section IV, the validation strategy and experiments for the performance evaluation of the derived building block models as well as evaluation of the UMCs are presented. Finally, Section V contains a summary of the main research contribution and concluding remarks.

II. Study Area and Data Collection

The sites analyzed in this study are the four metropolitan areas of the cities Munich, Stuttgart, Frankfurt, and Cologne in Germany. The metropolitan areas are defined as the administrative areas of these four cities including the surrounding municipalities which are located within the employment market region of the cities [34]. The total population in the investigated area is about 12 million and the total investigation area cumulates to almost 16 000 km^2, which is about the area of Connecticut. We use Cartosat-1 stereo pairs for the generation of large-area DSMs of the metropolitan regions. DSMs are used to extract building heights for the total area and building footprints are extracted from topographic maps at a scale of 1:25 000.

A. Cartosat-1

The Cartosat-1 (P5) sensor is mounted on-board the Indian Remote Sensing Satellite (IRS) which is a dual-optics two-line along-track stereoscopic pushbroom scanner. It collects panchromatic stereo data with two cameras (Fore camera and Aft camera) at a resolution of 2.4 (Fore) and 2.2 m (Aft), respectively. The stereo angle between the images is 31° for an area of 26.8 × 30 km (Fore) and 26.8 × 26.8 km (Aft) in a single scene [20]. The sensor was designed for large-scale topographic mapping. The geometric resolution does not allow for a detailed

extraction of individual buildings, but advanced stereo matching algorithms allow to extracting 3-D information over urban areas. In combination with data from topographic maps, a 3-D building block model can be generated.

The Cartosat-1 data for the test sites were acquired between January 24, 2008 and July 25, 2012. All in all we use 129 Cartosat-1 stereo pairs (Munich: 28; Stuttgart: 28; Frankfurt: 34; Cologne: 39). Each stereo pair is used for the generation of a DSM applying an adapted semiglobal matching procedure (SGM) [35]. To establish a consistent geometry for all stereo pairs, a block adjustment is performed using ground control points (GCP), resulting in bias-corrected rational polynomial coefficient (RPC). The matching results and RPC parameters are used to generate the final DSM with 5-m spacing. The method for DSM generation is described in detail in Section III.

B. Digital Topographic Map 1:25 000 (DTK-25) and Ancillary Vector Data

DTK-25 is used in our study to extract building block footprints. Maps are scanned and georeferenced raster data of the topographic map 1:25 000 (TK25). For the entire area of Germany, a total number of 4070 seamless tiles with each covering 10 km $\times$ 10 km are available. For our study areas, we processed a total of 239 tiles (Munich: 64; Stuttgart: 48; Frankfurt: 82; Cologne: 45). Data are georeferenced to Universal Transverse Mercator (UTM) zone 32 with ellipsoid World Geodetic System 84 (WGS84). Positional accuracies of the maps vary between 10 and 20 m [36].

Maps have been scanned at a resolution of 254 dpi and stored as TIFF with a color depth of 1 bit (black-and-white). Data includes the raster data of DTK-25 without terrain representation by shadow relief. The German Federal Agency for Cartography and Geodesy (BKG) provides DTK-25 as single layers which correspond to the colors of the printed maps. Thus, the layer containing building footprints contains also lettering and boundaries [37] (cp. Fig. 3). Actualization cycles of topographic maps are up to 3 years making this data more or less up-to-date for less dynamic urban regions.

For the extraction of building footprints from the DTK-25, additional vector data are used. In this context, urban block, street and railway geometries from the digital basic landscape model (Basic-DLM) are integrated in the building extraction procedure [38]. The Basic-DLM describes the topographic features of the landscape in vector format and has a basic topicality of 5 years and a top up-to-dateness of 3–12 months for certain feature types.

C. Digital Terrain Model

A digital terrain model (DTM) with a spacing of 25 m provided by the Federal Agency for Cartography and Geodesy (BKG) is used for the block adjustment (see Section III) and for the generation of a normalized DSM (nDSM). Location correctness of the BKG-DTM is given with an RMSE of 1–3 m in position and elevation, depending on terrain type [39].

D. Building Cadaster

The accuracy of the derived building block models is assessed using a cadastral building model for the city of Cologne. The cadaster holds precise geometric information on the location, the size, median roof height, and usage of each individual building. These cadastral building data are used for performance evaluation in Section IV.

E. Multisource Considerations of the Data

For the building block model generation, two different data sets, acquired at different points in time, are used: the topographic maps were updated on January 1, 2008 and P5-DSM was acquired in the years between 2008 and 2012. As the observed metropolitan areas show very low urban growth dynamics, considerations regarding the time lags between the data can be neglected.

All data have been transformed into Transverse Mercator projection with WGS84 ellipsoid and UTM Zone 32 as projected coordinate system.

III. Methods for the Derivation of UMCs

In this section, we describe the generation of the building block models from Cartosat-1 DSMs and topographic maps for the identification of UMCs. In the first part, the process for the derivation of the large-area DSMs for the four metropolitan areas is described in detail. In the second part, we describe the fusion of the derived DSM with the topographic maps for the generation of the building block models (Fig. 2).

A. DSM Generation From Cartosat-1 Stereo Images

1) Stereo Matching: Within the stereo matching procedure, at the beginning, quasi-epipolar images are generated to restrict the search range for the dense stereo matching into one dimension. The SGM algorithm [35] is used to perform dense and reliable stereo matching. SGM is able to reconstruct sharp object boundaries whereby it avoids using matching windows which would lead to fuzzy edges. Instead of strong local constraints in a window, a (semi-) global energy function E is minimized for all disparities D (local shifts within disparity matrixes). SGM performs a semiglobal optimization by aggregating costs from 16 directions, which finds an approximate solution to the global energy function E:

$$E(D) = \sum_{p} \left(C(p, D_p) + \sum_{q \in N_p} P_1 T\left[|D_p - D_q| = 1\right] + \sum_{q \in N_p} P_2 T\left[|D_p - D_q| > 1\right] \right). \quad (1)$$

The function C defines the matching cost between the image pixels for each pixel location p in the first image. The simplest cost function is the absolute gray value difference of pixel p and the corresponding pixel in the second image, as defined by

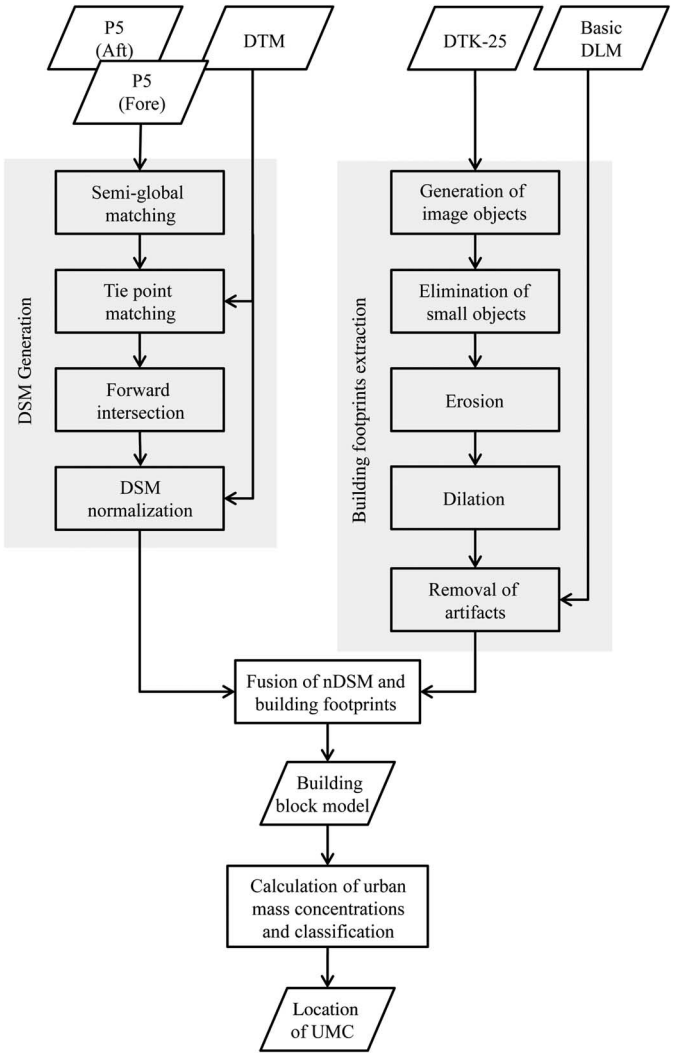

Fig. 2. Flowchart of the approach to derive building volume densities on 1×1 km grid cells based on Cartosat-1 stereo images and DTK-25 topographic maps. The processes on the left represent the generation of the nDSM from Cartosat-1 data and the processes on the right represent the generation of the building footprints by extraction from topographic maps at a scale 1:25 000. In the lower part of the flowchart, both resulting data sets from the two processes are merged into a building block model. The building block model is further used for the calculation and localization of UMCs.

the disparity map D_p. The cost function used in this paper is a sum of mutual information [35] and census [40]. These cost functions adapt to brightness changes in the stereo images and allow matching of images with large viewing angle differences. The second and third terms of E penalize disparity changes in the neighborhood N_p at each position p. T is 1 if the argument is true, and 0 otherwise. The penalty P_1 is added for all disparity changes equal to one pixel. At larger discontinuities (disparity change > 1 pixel), a fixed cost P_2 is added. This cost function favors similar or slightly changing disparities between neighboring pixels, and thus stabilizes not only the matching in image areas with weak contrast, but also allows large disparity jumps at e.g., building edges.

Minimizing (1) for two-dimensional neighborhoods N_p is an NP-complete problem, for which no efficient algorithms exist. In SGM, the minimization is performed by aggregating the cost along 8 or 16 paths, and tabulating the aggregated costs for all pixels and possible disparities in a 3-D array. The disparity map D is computed by choosing the disparities with the smallest aggregated cost. Matching is performed from first to the second and second to the first image, and only consistent disparities are kept. Finally, small, independent disparity segments are identified and removed as outliers. A detailed explanation of the used algorithms is given in [35] and [41].

2) Block Adjustment: Cartosat-1 stereo scenes are provided with the RPCs sensor model [42] derived from orbit and attitude information. Using the RPCs directly results in ortho-images and DSMs with shifts of 15 m CE90 according to [43], however, we have identified single absolute shifts of 100 m or more. Thus, good relative and absolute orientation of the DSMs are achieved by a block adjustment of all urban regions consisting of 129 Cartosat-1 stereo pairs.

The first step of automatic block adjustment is tie point matching between all overlapping images in the block. For performance enhancement, a coarse matching of all overlapping images is applied using scale-invariant feature transform (SIFT) [44]. The resulting tie points can connect multiple overlapping images. The applied SIFT method is an enhancement of the standard procedure as it was modified to allow feature extraction in areas of low contrast with an increased matching ratio threshold. This makes it possible to match images which do not match with the standard parameters; however, it also leads to an increased number of outliers. Those blunders are removed using random sample consensus (RANSAC). The remaining tie point image coordinates are refined to an accuracy of approximately 0.1 pixels using local least squares matching. In total, 366 747 high-quality tie points with 998 032 image projections were found.

Usually, highly accurate GCPs identified in each image are used during the block adjustment. However, GCPs of suitable accuracy are often not available. In this work, the BKG-DTM is used as horizontal and vertical reference in the block adjustment. Integration of this DTM allows fully automatic adjustment without GCPs. We apply standard affine RPC block adjustment [42], which has been extended with the constraint that the tie points have to lie on the surface of the reference (BKG-DTM) [45]. The use of the BKG-DTM as reference surface during the block adjustment effectively performs a DSM alignment. Thus, both BKG-DTM and Cartosat-1 DSM are in the same horizontal and vertical reference frame, allowing valid comparison between the two data sets. This approach requires a dense tie point collection and works well in regions with some terrain variation, which is present in the four urban areas. For each image, the block adjustment estimates a six-parameter affine correction in image space [42]. The block adjustment yielded a tie point image reprojection RMSE of 0.16 pixels indicating a very good relative orientation of the entire block. The difference between tie point height and reference DTM height shows an RMSE of 4.24 m. The BKG-DTM does only contain bare ground heights, but some of the automatically matched tie points are located on the forest canopy or building roofs. Thus, an overall discrepancy of more than 4 m can be expected, and is mostly caused by comparison of a surface model with a terrain model. We have visually checked the P5-DSM and P5 ortho

Fig. 3. Extraction of building footprints from DTK-25. The left part of the figure displays the black layer of the digital topographic map 1:25 000 (DTK-25) with the map content including buildings, streets, railways, and lettering. After application of the building extraction procedure, only building footprints remain (right part of the figure). These extracted building footprints are further used for the generation of the building block models in combination with the Cartosat-1 DSM. Topographic map: GeoBasis-DE/BKG (2010).

images against high resolution aerial ortho imagery from BKG and found no shift between these data sets, indicating a good horizontal geolocation of the DSM.

After block adjustment, DSMs for each stereo pair are generated by forward intersection of the SGM matching results. The pairwise DSMs are merged using the median elevation value for each pixel. Discrepancies between DSMs from overlapping stereo pairs are usually below 2 m, due to the subpixel accurate matching. Higher deviations exist in areas that have changed, e.g., mining areas, cut down forests, or between stereo pairs that were acquired with leaf ON and leaf OFF conditions. For occluded areas, regions where the matching failed or outliers are removed, a filling procedure is applied. Small holes are filled using inverse distance weighted (IDW) interpolation [46], while large holes due to clouds or water bodies are filled with the BKG-DTM using the delta surface fill method [47]. The resulting DSMs have been checked against the BKG-DTM with a large number of reference points to calculate the mean error: for the Cologne test site, for 42 008 reference points, a mean error of 0.55 m is computed for the entire DSM, in Stuttgart 0.02 m (78 417 points), in Frankfurt −0.27 m (341 624 points), and in Munich 0.10 m (266 015 points).

For the generation of the building block models, terrain heights in the DSM are removed to retrieve relative heights of the elevated objects above ground such as buildings. Therefore, the P5-DSM is normalized using the BKG-DTM applying a subtraction of DTM elevation values from the DSM resulting in the normalized P5-BKG.

B. Building Footprint Extraction From Topographic Maps

The layer containing the black elements from the DTK-25 (cp. Section II-B) is used for the extraction of building footprints. The extraction procedure of all 239 map tiles has been achieved fully automatic and consists of several steps which are described in the following in detail. The object-based workflow is implemented in the software package eCognition 8.0 [48]. In a first step, all connected black pixels are converted to cohere image objects. This step is followed by the elimination of very small point objects based on an area threshold. Narrow elongated objects like streets, railway tracks, and administrative boundaries are eliminated in a subsequent step using the morphological operator *erosion*. Implementation of this operator in eCognition decreases the size of the objects from the outer object boundary by a certain distance. Thus, in our case, the majority of line objects can be removed by applying erosion. The remaining objects of an area smaller than 100 pixels are further investigated based on their shape (width and compactness) to discriminate between small buildings and fragments of the erosion process. The latter are removed iteratively from the data set. For the image objects which remain in the data set and which have been affected by the erosion operator, a *dilation* operator is applied to restore the initial objects' sizes. Dilation is the opposite process of erosion, meaning that the sizes of the image objects are enlarged by the same distance as the erosion. In the subsequent step, ancillary data from the Basic-DLM [38] are used for the removal of artifacts from the preceding procedure and to reduce the data set to the areas dominated by buildings (Fig. 3).

C. Generation of Building Block Models

Height information from the nDSM P5-BKG and extracted building footprints are combined to generate the targeted building block models. The building block models are reconstructed at the level-of-detail 1 (LOD-1) [49]. Thus, the height of each building is represented by one single height value. This specific height value is calculated as the area-weighted median value of P5-BKG and is derived as follows. To each building footprint, the median value of all P5-BKG height values which are covered by the building footprint is assigned. Because the point spacing of 5 m is very large for small or narrow buildings, the point spacing was altered to 1 m using nearest neighbors resampling. In this way, the original height values are not modified

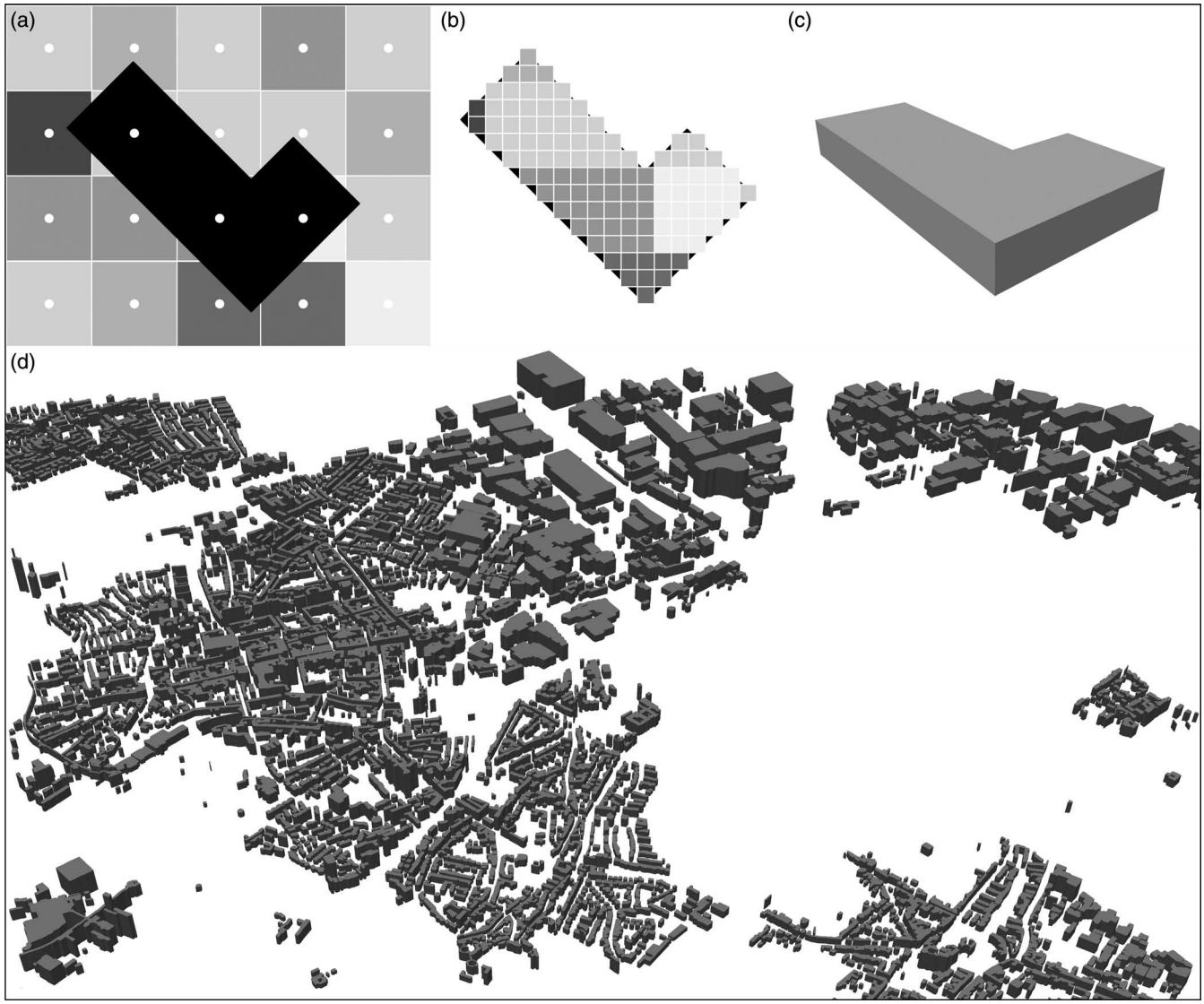

Fig. 4. Generation of building block models by fusion of height values from P5-BKG and extracted building footprints from DTK-25. (a) A building footprint overlaid on the height values of P5-BKG. Center points of P5-BKG are represented as white points and height values are represented as gray values. The higher the value, the brighter the color and vice versa. In this example, only three height values at a spacing of 5 m would be fused with the represented building footprint and thus the building height modeled by only three observations (three points intersect with the building footprint). To overcome this problem, point spacing is altered to 1 m and thus the number of observations (height values per building footprint) is increased in (b). The median of all height values covered by the building footprint is calculated and assigned as the building block height in (c) and a depiction of a perspective view of the building block model for the same area as in Fig. 3 above is presented in (d) Topographic map: GeoBasis-DE/BKG (2010).

but the number of height values per building footprint can be increased. This leads to a more stable calculation of the resulting median height value per building. A minimum threshold of 2 m was applied on the height values to reduce errors from the stereo image matching. A graphical depiction of the generation of the building block model is presented in Fig. 4.

D. Derivation of UMC

The building block models form the basis for the derivation of UMCs. Concentration is a measure of density and thus relates to a proportion between objects and a reference plane [1]. In general, these objects can be composed by manifold variables ranging from physical objects like buildings, various land-cover types [50], [51] to activities such as job density, etc. [34]. In our case, we want to locate concentrations of urban masses, thus we focus on the physical quantifiable elements of urban space—buildings. Density is described as the fraction of the number of observed elements in the numerator and the denominator, in our case the reference plane. The correct size and location of the reference plane is a crucial variable in the context of density measuring [52] and it is subject to the *modifiable areal unit problem* (MAUP) [53]. MAUP describes the variability of density dependent on the size and location of the reference plane. It is a well-known issue in quantitative geography and has to be considered when interpreting results. For the derivation of UMC, we relate to the INSPIRE grid [54] which is a standardized data set containing 1×1 km grid cells covering the Europe. Thus, for the derivation of the UMC, we relate building volume from the derived building block models with

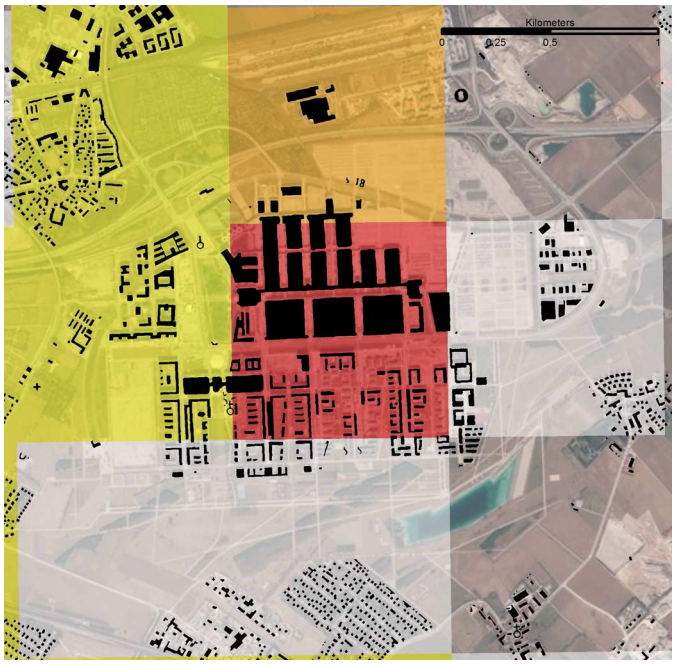

Fig. 5. Derivation of UMCs. For each of the 1 × 1 km grid cells the volume of the corresponding building blocks is cumulated. In this manner, only the share of the respective part of the building volume is used for the calculation of UMC for each grid cell. The figure shows also the classified grid cells which were classified with an SD-based classification procedure (red = very high UMC; orange = high UMC; yellow = medium UMC; white = low UMC; no color = very low UMC). Source background image: Esri, DigitalGlobe, GeoEye, i-cubed, USDA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community. Topographic map: GeoBasis-DE/BKG (2010).

the 1 × 1 km grid cells as reference plane and perform a classification procedure. Thus, in the remainder of this article, UMC is also referred to as classified VD. Building volume is calculated separately for each individual building block as the product of the building footprint area and its height. VD is further calculated by the total sum of the volume of all buildings which are contained by a grid cell and then divided by the area of the grid cell (1 km^2). For buildings which are crossed by the border of one or more grid cells, only the volume for the respective part of the building is used for the calculation of VD (cp. Fig. 5).

For the identification and location of areas comprising different UMCs with a special focus on very high UMCs, a classification procedure is applied on the volume densities for each of the grid cells. In this manner, five classes of UMCs are distinguished: *very low (vl), low (l), medium (m), high (h),* and *very high (vh)* UMCs. Definition of the class borders is designed to be flexible between various test sites and to be able to cope with various levels of scale. Thus, a transferable classification procedure based on the log-normal distribution (for all grid cells with a VD > 0) is applied because the values of VD show approximately a logarithmic distribution. For the classification procedure, the log-normally distributed VD values are standardized (VD_{st}) to a mean (μ) of 0 and an SD (σ) of 1

$$VD_{st} = \frac{VD - \mu}{\sigma}. \tag{2}$$

In the following, UMCs are classified based on the following equation:

$$UMC = \begin{cases} vl, & VD_{st} \leq -1.5 \\ l, & -1.5 < VD_{st} \leq -0.5 \\ m, & -0.5 < VD_{st} \leq 0.5 \\ h, & 0.5 < VD_{st} \leq 1.5 \\ vh, & 1.5 < VD_{st}. \end{cases} \tag{3}$$

A graphical depiction of the classification of UMCs based on VD is presented in Fig. 5.

IV. Experiments

The above described framework for the derivation of UMCs aims at the application on extremely large areas at the regional scale in the urban context. For performance evaluation of the derived UMCs, validation of the generated data has been performed. In this manner, we describe a thorough evaluation of the quality of the UMCs. For a detailed assessment, however, it is also important to inspect not only the result of the calculations but also the individual elements of the calculation which correspond to building area and building height in the context of UMCs.

The building height is derived from Cartosat-1, thus special emphasis is given assessing the quality of the building height. The second variable, the building area is directly related to the quality of the underlying topographic map (cf. Section II-B), which itself represents only a generalized image of reality and is subject to map production processes such as generalization, simplification, or selection. But against the background of the applicability and transferability of the described workflow on other regions or even the entire area of Germany, we also assess the quality of the building area. The underlying motivation in this context is to identify and quantify reasons for differences of the derived UMCs compared to reality, since differences in volume cannot be directly connected to either uncertainties in building height or building area.

A. Methodological and Conceptual Considerations for Performance Evaluation

Regarding performance evaluation, in the field of thematic maps, evaluation also referred to as accuracy assessment [55], [56], two maps are compared quantitatively on the basis of an error matrix, distinguishing between *classification* and *reference*, whereas in our case classification corresponds to DTK-25 and reference to the building cadaster. Comparing both data, an entity in the *classification* is called a *True Positive (TP)* when it corresponds to an entity in the *reference*. A *False Negative (FN)* is an entity in the *reference* which does not correspond to an entity in the *classification,* while a *False Positive (FP)* is a *classified* entity which does not correspond to an entity in the *reference*. A *True Negative (TN)* occurs when there is neither an entity in the *reference* nor in the *classification* [57]. Based on these measures, the quality measures *completeness* and *correctness*, also referred to as *Producer's Accuracy* and *User's Accuracy* [58], respectively, are calculated as well as *quality* [59].

TABLE I
PSEUDO-PIXEL-BASED EVALUATION OF BUILDING AREA

	Completeness	Correctness	Quality	Kappa
DTK-25	87.00%	65.30%	59.50%	0.71

Quality includes completeness and correctness and is a measure of "goodness" of the result

$$completeness = \frac{TP}{TP + FN} \tag{4}$$

$$correctness = \frac{TP}{TP + FP} \tag{5}$$

$$quality = \frac{TP}{TP + FP + FN}. \tag{6}$$

In our case, completeness is the percentage of building area in the reference which was detected, while correctness describes the percentage of detected building area matching the *reference*. Completeness and correctness are established measures for the evaluation of thematic accuracy [57], [58]. Furthermore, the kappa value [59] is another established measure of accuracy on the basis of the error matrix. The kappa value is used to define to what extent the outcome differs from a random result. The RMSE is a statistical measure of difference used to characterize the mean value of a series of data. It is very sensitive to statistical outliers. Besides the RMSE, we also use the MAE and SD to measure errors of derived heights and reference heights. These measures allow the comparison of our results with other studies.

As the relevant entity for performance evaluation, we have selected, on one hand, those objects which correspond to building blocks and, on the other hand, the 1×1 km grid cells. Pixels as the spatial entity are discarded, since [55] in general and [57] and [60] in the context of building extraction pointed out some problems in the context of pixel-based evaluation. In our case, the relevant issues are differences between the building outlines caused by vector–raster conversions, sampling effects due to scale/generalization of map elements in DTK-25, systematic errors due to misregistration of the data and the accuracy of the reference data. However, a pseudo-pixel-based performance evaluation of the detected building block footprints from DTK-25 has been carried out, revealing a quality of 59.50% (Table I). The comparably high completeness and low correctness can be explained by the smaller scale and the enlarged spatial representation of buildings in the DTK-25 compared to the building cadaster. Thus, building footprints in the DTK-25 are larger than in the reference data (medium gray color in Fig. 6). Pseudo-pixel-based means that original vector representation of the building cadaster has been kept but the method of performance evaluation is identical to pixel-based.

Pixel or pseudo-pixel-based evaluation, however, is not able to give a structured insight into the performance in dependency on building area or building height. Thus, to overcome conceptual errors in performance evaluation due to problems with pixel-based evaluation, the entity *object* or *building block* is used for further evaluation. This spatial entity also allows

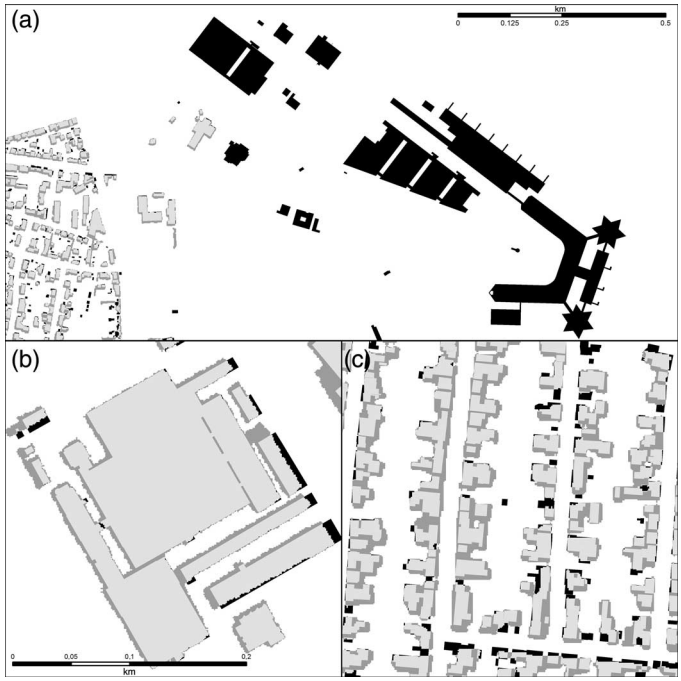

Fig. 6. Pseudo-pixel-based evaluation of the building block representation of DTK-25 compared with the building cadaster for various urban environments: no corresponding object airport (a); industrial (b); suburban neighborhood (c). Light gray: *TPs*; medium gray: *FPs*, black: *FNs*. Topographic map: GeoBasis-DE/BKG (2010).

to evaluate the derived building block heights and to inspect accuracies as a function of object/building block size. The reference data include very small buildings with minimum sizes of $1\,\mathrm{m}^2$. Additionally, 25% of all reference buildings are smaller than $54\,\mathrm{m}^2$. In the DTK-25, the smallest building object is $45\,\mathrm{m}^2$.

These differences in the minimum object size can distort the results of performance evaluation significantly, for which reason the following performance evaluation was conducted only on objects which have a mutual *complete* or *partial overlap* in both data sets and a minimum area of $50\,\mathrm{m}^2$. The mutual overlap is calculated according to [60] and is defined as the ratio of the overlap area of a building object in DTK-25 (B_m) and the reference (B_r) in the way $o_{mr} = a_{m\cap r}/a_m$ and $o_{rm} = a_{m\cap r}/a_m$, where the overlapping regions are elements of the initial building objects $b_m \in B_m$ and $b_r \in B_r$, where $a_{m\cap r}$ is the common area in the overlapping region b_m in the DTK-25 and the overlapping region b_r in the reference data. Because building object representation in the two data sets is not the same, we calculate the overlap o percentages o_{mr} and o_{rm} and classify the overlap o according to [60] into *none* (n), *weak* (w), *partial* (p), and *strong* (s):

$$o = \begin{cases} n, & o_{mr} \vee o_{rm} \leq 10\%\, o_{mr} \\ w, & 10\%\, o_{mr} < o_{mr} \vee o_{rm} \leq 50\%\, o_{mr} \\ p, & 50\%\, o_{mr} < o_{mr} \vee o_{rm} \leq 80\%\, o_{mr} \\ s, & 80\%\, o_{mr} < o_{mr} \vee o_{rm}. \end{cases} \tag{7}$$

The classes correspond to *completely detected* for $o_{rm} = strong$, *partly detected* for $o_{rm} = partial$, *hardly detected*

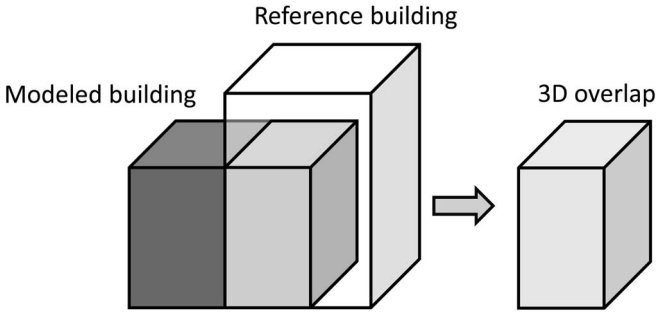

Fig. 7. Schematic illustration for the calculation of 3-D overlap (light gray) between reference building (white) and modeled building (dark gray). Each corresponding building object in both data sets is spatially intersected resulting in an overlapping corpus. Its volume is compared with the volume of the modeled building and to the volume of the reference building.

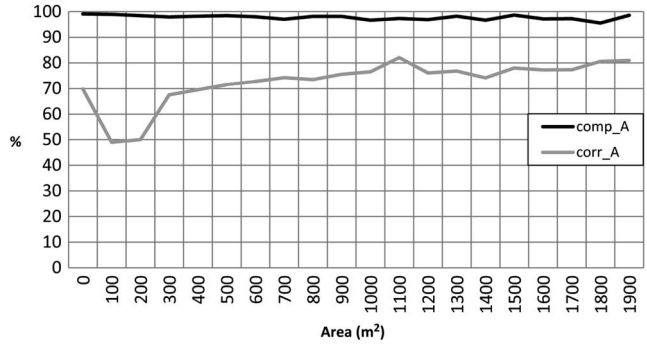

Fig. 8. Completeness (comp_A) and correctness (corr_A) as a function of the building area for 30 741 individual building objects. The accuracy measures are displayed for all buildings with a larger area than the value in the abscissa.

for $o_{rm} = weak$, *not detected* for $o_{rm} = none$ and *completely correct* for $o_{mr} = strong$, *partly correct* for $o_{mr} = partial$, *hardly correct* for $o_{mr} = weak$ and *not correct* for $o_{mr} = none$. With these measures, the numbers for *TP*, *FN* and *FP* can be computed where:

1) *completeness* TP is the number of building regions in the reference that are either *completely* or *partly detected*.
2) *correctness* TP is the number of building regions in the DTK-25 that are either *completely* or *partly correct.*
3) FN is the number of building regions in the reference data that are *hardly detected* or *not detected*.
4) FP is the number of building regions in the DTK-25 that are *hardly* or *not correct*.

For performance evaluation of the variable *height,* we investigate *RMSE*, *MAE,* and *SD* in dependency on several height or area classes. For evaluation of the variable *volume*, a modified strategy of the above described performance evaluation for the area has been adopted. In this manner, the 3-D overlap of the building objects from the reference and the respective modeled building object is calculated (Fig. 7) and further compared to the volumes of the reference and the modeled building. The 3-D overlap is further classified into *strong*, *partial*, *weak,* and *none* using the same thresholds as for the areal overlaps (80%, 50%, and 10%).

At last, the topologies of the reference data were adjusted to match the geometries of the DTK-25 in the sense that adjacent buildings were merged [57]. In this manner, we automatically conduct performance evaluation for the building block model (area, height, and volume) which is followed by performance evaluation of the derived UMCs with the reference data for the entire area of the administrative area of the city of Cologne.

B. Object-Based Evaluation of the Building Block Model by Evaluating the Mutual Overlap

As described in Section IV-A, an object-based performance evaluation for the building footprints has been carried out. Therefore, the mutual overlap between the derived building footprint and the reference data has been calculated and classified according to the thresholds. Based on these thresholds, the objects are classified into *TP*, *FP,* and *FN* and from it the accuracy measures completeness and correctness were computed. Fig. 8 displays both measures as a function of area. Completeness is very high for all groups ranging between 95% and 99%. The reason for these high completeness values is the consistent generalization of building objects in the DTK-25 (cp. Fig. 6) representing the buildings larger than in reality. This effect results in comparably low correctness values between 49% and 81%. Another influencing factor reducing the correctness is the different topological representation of building objects in the reference and in the DTK-25. Buildings at close distance to each other are topologically merged in the DTK-25, while they are still represented as single buildings in the reference. Thus, very large differences in area occur on object level. These topological effects are not corrected here because this would alter the result, and thus the conclusion of performance evaluation would be drawn upon a topologically different result. We aim at providing performance measures for the entire derived building block model, including topological errors due to the differing object representations in the DTK-25 and the reference data. Thus, our investigation is evaluating the applicability of the data "as is." In contrast to this analysis, Fig. 9 displays the performance evaluation for a subset of entire building objects with corrected topology. There, the effects of generalized building footprints in the DTK-25 are visible especially for smaller buildings since only buildings larger than 300 m^2 are more than 90% correct.

Performance of building height is evaluated directly as the *MAE*, the *RMSE*, and the *SD* for all individual building objects. To inspect whether systematics can be found in the derived building heights from Cartosat-1, we perform two different experiments. In the first one, we compute these accuracy measures as a function of building area (Fig. 10). All accuracy measures follow a similar curve and no clear trend or pattern can be observed. The group with the smallest building areas shows very low errors: *MAE* of 2.37 m, *RMSE* of 3.49 m, and *SD* of 3.38 m; for the group with the largest building areas, the errors are: *MAE* of 4.09 m, *RMSE* of 5.91 m, and *SD* of 4.76 m. The overall accuracy measures for all building objects are: *MAE* of 3.22 m, *RMSE* of 4.55 m, and *SD* of 3.71 m (Table II).

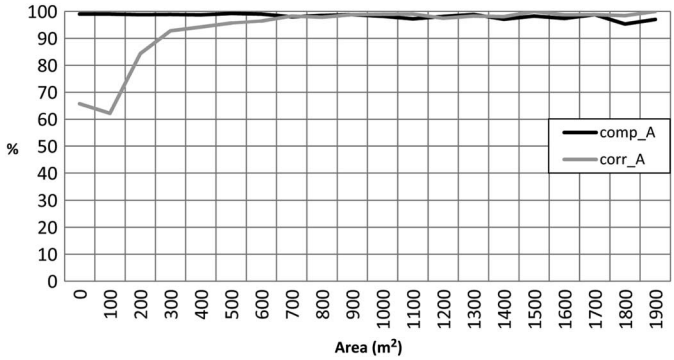

Fig. 9. Completeness (comp_A) and correctness (corr_A) as a function of the building area 21 366 individual building objects with the same topology between reference building footprints and DTK-25 building footprints. The accuracy measures are displayed for all buildings with a larger area than the value in the abscissa.

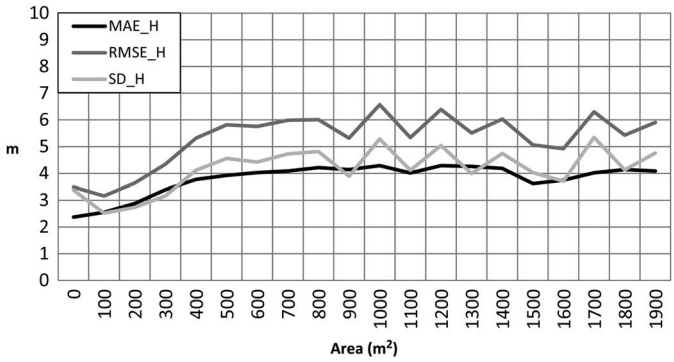

Fig. 10. Mean absolute error (MAE_H), root-mean-squared error (RMSE_H), and standard deviation (SD_H) of the building height as a function of the building area for 30 741 individual building objects. The accuracy measures are displayed for all buildings with a larger area than the value in the abscissa.

TABLE II
PERFORMANCE EVALUATION FOR BUILDING HEIGHT AND VOLUME

	MAE	RMSE	SD
Building height as a function of h	3.22 m	4.55 m	3.71 m
Building height as a function of a	3.21 m	4.48 m	3.63 m
Volume (grid)	0.14 km³	0.23 km³	0.22 km³

In a second performance evaluation for the building height, the structuring variable is the building height. This experiment aims at finding any significances of the derived building height in dependency of the building height itself. Thus, a similar experiment was performed as the previous one. In Fig. 11, a clear trend can be observed in the accuracy measures: with rising building height (abscissa) all difference-based error measures rise as well. Thus, we observe larger errors for higher buildings and smaller errors for lower buildings. For buildings with a height lower than 7 m, we observe an *MAE* of 2.94 m, an *RMSE* of 3.24 m, and an *SD* of 1.68 m. For buildings with a height larger than 21 m, we observe an *MAE* of 11.48 m, an *RMSE* of 12.39 m, and an *SD* of 4.74 m. The fourth curve in the plot displays the SDs of the relative heights. Most interestingly,

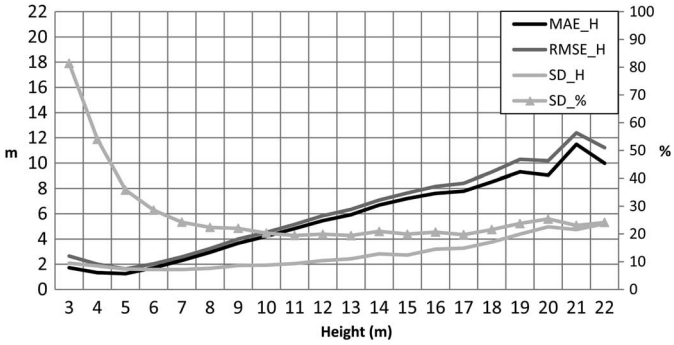

Fig. 11. Mean absolute error (MAE_H), root-mean-squared error (RMSE_H), standard deviation (SD_H), and SD of the relative height (SD_%) of the building height as a function of the building height for 30 741 individual building objects. The accuracy measures are displayed for all buildings with a greater height than the value in the abscissa.

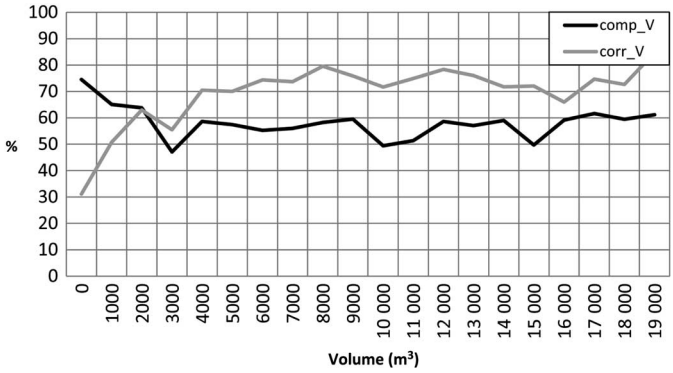

Fig. 12. Completeness (comp_V) and correctness (corr_V) of the building volume as a function of the building volume for 30 741 individual building objects. Accuracy measures are displayed for all buildings with a larger volume than the value in the abscissa.

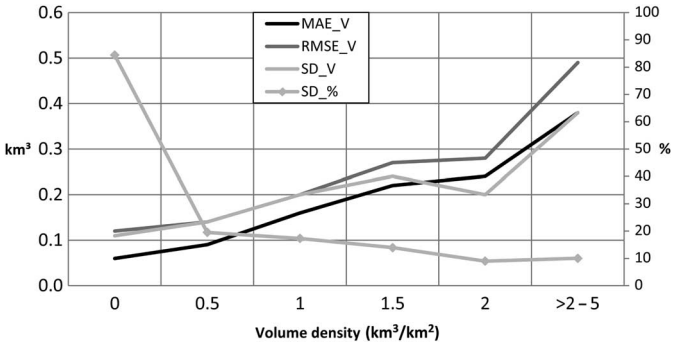

Fig. 13. Mean absolute error (MAE_V), root-mean-squared error (RMSE_V), standard deviation (SD_V), and standard of the relative volume (SD_%) of the volume densities for the grid cells as a function of the volume densities for a total of 324 grid cells. The accuracy measures are displayed for all grid cells with a greater VD than the value in the abscissa.

it shows that for buildings higher than 7 m, the *SD* is more or less constant between 19.5% and 25.4%.

The previous performance analysis reveals that the generated building models are larger in size but lower in height than the reference data. In the following experiment, we aim

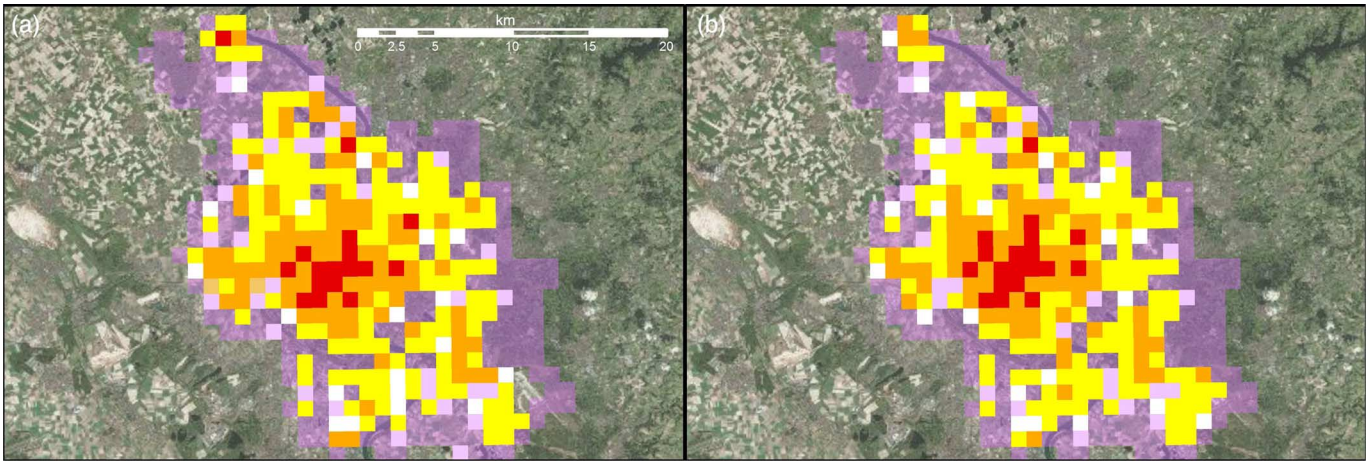

Fig. 14. Classification of UMCs based on SDs of the volume densities for the reference data (a) and for the derived volume densities based on DTK-25 and Cartosat (b) (red = very high UMC; orange = high UMC; yellow = medium UMC; white = low UMC; purple = very low UMC; no color/aerial image = outside the test area of Cologne). Source background image: Esri, DigitalGlobe, GeoEye, i-cubed, USDA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community.

at finding how these errors affect the volumes of the building objects (Fig. 12). The curves reveal that buildings with a volume less than 3000 m^3 show completeness values between 64% and 75%. The completeness decreases for buildings with a larger volume, as well as the correctness. The curve for the correctness is very similar to the curve of the correctness for the building area (cp. Fig. 8). Thus, in an object-based comparison of the volume of individual building objects, the effects of lower building heights and larger areas than the reference are clearly visible.

C. Performance Evaluation of the Derived Volume Densities and Classified UMCs

Errors due to the scale/generalization or systematic errors such as misregistration of the DTK-25 do affect strongly the performance of the derived building models. On the other hand, however, the effects are not so strong for the aggregated information of the VD on the grid cells/UMCs (cp. Fig. 13). The evaluation reveals that the absolute errors increase for increased volume densities with an *RMSE* of 0.5 km^3 for the group with the highest volume densities. But when we inspect the relative errors in terms of SD, it can be observed that the relative error for most of the groups is around 10%. The relative error is very high for very low UMCs with volume densities lower than 0.5 km^3/km^2.

In a final performance evaluation, the procedure for UMC classification is evaluated to assess the impact of the errors in volume densities on the derived UMCs. In this manner, classified UMCs from the reference data were compared with classified UMCs from the derived volume densities. Based on an error matrix, the resulting class of each grid cell is compared for the reference data and the modeled volume densities (Fig. 14, Table III).

Results show good thematic classification accuracies in terms of *completeness* and *correctness* for all classes with kappa values between 0.75 and 0.94 and a *quality* between 63.08% and 93.53%, respectively.

TABLE III
EVALUATION OF SD-BASED CLASSIFICATION OF UMCS

	Completeness (%)	Correctness (%)	Quality (%)	Kappa (%)
Very low	96.91	96.41	93.53	0.94
Low	77.36	77.36	63.08	0.75
Medium	86.36	83.21	73.55	0.77
High	78.57	86.84	70.21	0.84
Very high	88.24	78.95	71.43	0.78

D. Application

For all metropolitan areas of Munich, Stuttgart, Frankfurt, and Cologne (cp. Section II), the described procedure has been applied to derive volume densities and furthermore to localize UMCs. The derived thresholds for the classification are the same for all four metropolitan regions, thus the identified UMCs can be compared between the regions. It is visible that from the four regions, Munich has the highest UMCs grouped around the city center [Fig. 15(d)]. In comparison to the other metropolitan areas, Munich has also the largest connected areas of UMCs.

The characteristic urban structure of the metropolitan area of Stuttgart is the result of urbanization within an area with characteristic orographic features such as hills and valleys [Fig. 15(c)]. The urban center of Stuttgart is classified as class with the highest UMC. Besides that, a second physical center can be found in the southwest of the center. These UMCs belong to automobile industries in Sindelfingen. Characteristic for the metropolitan area of Cologne [Fig. 15(a)] is that it is a widely connected urban area with a clear identified urban center for the city of Cologne and a second urban center for the city of Bonn in the south. The Frankfurt metropolitan area [Fig. 15(b)]

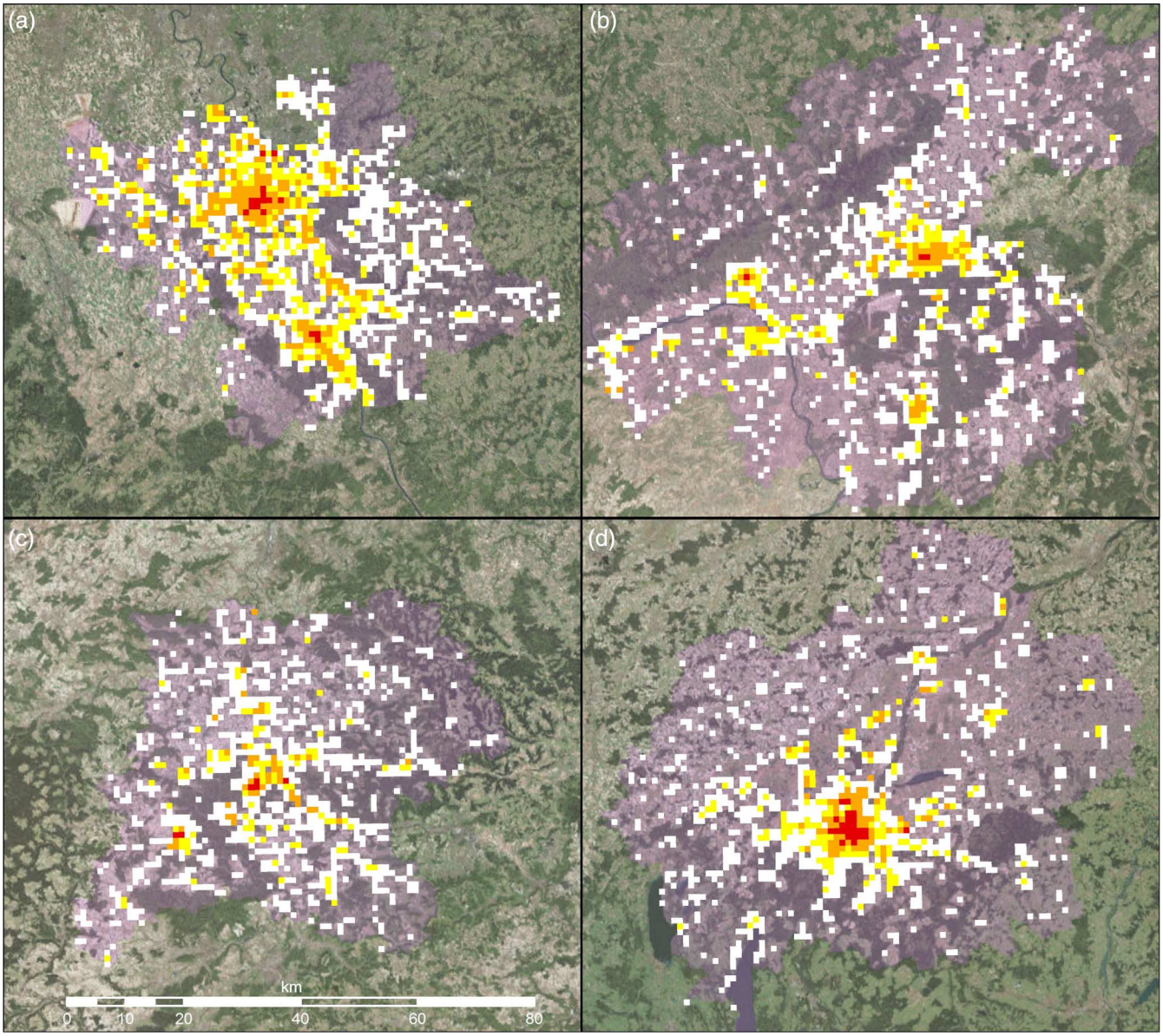

Fig. 15. Classification of UMCs based on SDs of the volume densities for the metropolitan areas of Cologne (a), Frankfurt (b), Stuttgart (c), and Munich (d) based on DTK-25 and Cartosat (red = very high UMC; orange = high UMC; yellow = medium UMC; white = low UMC; purple = very low UMC/no map or DSM data; no color/aerial image = outside the study area of Cologne). Source background image: Esri, DigitalGlobe, GeoEye, i-cubed, USDA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community.

is characterized by several independent urban centers. These are mainly Frankfurt city center, Darmstadt (in the south) and Wiesbaden/Mainz in the West.

V. Summary and Conclusion

In this paper, we describe the derivation of large-area UMCs based on the fusion of DSMs and building footprints for four metropolitan regions in Germany with a total area of more than $16,000$ km^2. As source for the DSMs, we have used spaceborne stereo data from Cartosat-1 resulting in a DSM with a point spacing of 5 m. Building footprint data has been extracted from digital topographic maps (DTK-25). Both height and footprint data were combined to derive large-area building models which are able to describe the urban morphology quantitatively. The physical characteristics such as building volume of the derived building models are used to compute volume densities for grid cells at 1×1 km. These grid cells are further classified into various types of UMCs.

To analyze the quality of the derived volume densities, we have conducted a thorough performance evaluation of the (a) derived building models, (b) the derived volume densities and the (c) classified UMCs. To assess the quality of the building models, we have applied an object-based approach where we compared the mutual overlap between the derived buildings and reference buildings. The outcome shows that due to

generalization in the map production process of DTK-25, individual building footprints are larger than the reference buildings. This effect is documented by the extremely high completeness values and the comparably low correctness values. To evaluate the performance of the building height, we have conducted two experiments. The first experiment reveals that there exists no significant correlation between the building area and the performance of the derived building heights, besides buildings with an area lower than $300\ \mathrm{m}^2$ the observed *MAEs*, *RMSE*, and *SD* are lower than for the other group. The second experiment reveals that *RMSE* and *MAE* of building height detection rise as a function of building height, but *SD* appears to be more stable. But evaluation of the relative errors shows that for buildings lower than 4 m, the relative errors are quite high, for buildings higher than 6 m, the *SD* varies between 19.4–24.2%. However, for both experiments regarding building height, we compute a mean overall height performance for all building objects of 3.21 m (*MAE*), 4.51 m (*RMSE*) and 3.67 m (*SD*). These outcomes are almost identical to the results of [23] who observe an *RMSE* of 4.14 m and a *SD* of 3.59 m and similar to [24] who performed a region-based comparison and observe a *SD* of 3.04 m and a *MAE* of 2.76 m. In comparison to these studies, we have conducted an automatic performance evaluation for a total of 30,741 building objects and investigated height errors as a function of area and height. For all those building objects we have also conducted performance evaluation of the building volumes, revealing a completeness of only 58% and a correctness of almost 70%. These errors are due to building height errors and generalization issues of the building footprints in the DTK-25.

However, our main goal in this paper is to derive and localize UMCs which are based on the volume densities. To investigate the impact of building height and building area errors on the derived volume densities, we conducted performance evaluation on the spatial level of the grid cells. Evaluation reveals that for grids with higher volume densities than $0.5\ \mathrm{km}^3/\mathrm{km}^2$, the relative *SD* ranges between 9–19.5%. The absolute errors of height and volume densities increase almost linearly with rising height or volume, respectively. Thus, the classification of UMCs into *very high, high, medium, low* and *very low UMCs* based on the log-normally distributed SDs produces similar results for the reference data and the modeled volume densities because the classifier is applied on each data set individually. But as there exists a linear correlation of reference and modeled volume densities, the achieved accuracies for the classification procedure reveals completeness values between 77 and 97%.

All in all, the main advantage of our proposed methodology is the large-area availability of both, building footprints from DTK-25 and Cartosat-1 DSMs with high accuracies. Thus, our results confirm the applicability of the data and the modeled building data can be used to support regional or even national studies.

Future research will focus on correction models of the derived building models in terms of area and height, as well as on correction models for the derived volume densities. Additionally, research should focus on alternative sources for building footprints such as from volunteered geographic information (VGI).

ACKNOWLEDGMENT

The authors would like to thank Dr. A. Schmitt and M. Weigand for their support; U. Lapschies from *Umwelt- und Verbraucherschutzamt der Stadt Köln* for providing cadastral reference data; the Federal Agency for Cartography and Geodesy (BKG) for the provision of the DTMs and topographic maps as well as Euromap GmbH; Neustrelitz for the provision of Cartosat data for scientific purpose; and the three anonymous reviewers for their valuable comments which helped to significantly improve the quality of paper.

REFERENCES

[1] S. Fina, A. Krehl, S. Siedentop, H. Taubenböck, and M. Wurm, "Getting closer! New ways of integrating geodata, statistics and remote sensing to analyze and visualize urban structures using density surfaces and profiles," *Raumforsch Raumordn*, vol. 72, no. 3, pp. 179–194, 2014, doi: 10.1007/s13147-014-0279-6.

[2] C. Acioly, Jr. and F. Davidson, "Density in urban development," *Build. Issues*, vol. 8, no. 3, pp. 3–25, 1996.

[3] H. Taubenböck *et al.*, "Delineation of central business districts in mega city regions using remotely sensed data," *Remote Sens. Environ.*, vol. 136, pp. 386–401, Sep. 2013, doi: 10.1016/j.rse.2013.05.019.

[4] F. Scholten, K. Gwinner, R. Tauch, and O. Boulgakova, "HRSC-AX—High-resolution orthoimages and digital surface models for urban regions," in *Proc. 2nd GRSS/ISPRS Joint Workshop Remote Sens. Data Fusion Urban Areas*, 2003, pp. 225–229, doi: 10.1109/DFUA.2003.1219992.

[5] F. Rottensteiner *et al.*, "Results of the ISPRS benchmark on urban object detection and 3D building reconstruction," *ISPRS J. Photogramm. Remote Sens.*, vol. 93, pp. 256–271, Jul. 2014, doi: 10.1016/j.isprsjprs.2013.10.004.

[6] B. Sirmacek, H. Taubenböck, and P. Reinartz, "Performance evaluation for 3-D city model generation of six different DSMs from air- and space-borne sensors," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 59–70, Feb. 2012, doi: 10.1109/JSTARS.2011.2178399.

[7] T. Krauss, P. Reinartz, M. Lehner, and U. Stilla, "Coarse and fast modelling of urban areas from high resolution stereo satellite images," in *Proc. Urban Remote Sens. Joint Event*, Apr. 11–13, 2007, pp. 1–12, doi: 10.1109/URS.2007.371777.

[8] T. Krauss, M. Lehner, and P. Reinartz, "Generation of coarse 3D models of urban areas from high resolution stereo satellite images," in *Proc. Int. Arch. Photogramm. Remote Sens.*, Beijing, China, 2008, pp. 1091–1098.

[9] S. Eckert and T. Hollands, "Comparison of automatic DSM generation modules by processing IKONOS stereo data of an urban area," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 3, no. 2, pp. 162–167, Jun. 2010, doi: 10.1109/JSTARS.2010.2047096.

[10] A. Thiele, E. Cadario, K. Schulz, U. Thonnessen, and U. Soergel, "Building recognition from multi-aspect high-resolution InSAR data in urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3583–3593, Nov. 2007, doi: 10.1109/TGRS.2007.898440.

[11] U. Soergel, E. Michaelsen, A. Thiele, E. Cadario, and U. Thoennessen, "Stereo analysis of high-resolution SAR images for building height estimation in cases of orthogonal aspect directions," *ISPRS J. Photogramm. Remote Sens.*, vol. 64, no. 5, pp. 490–500, Sep. 2009, doi: 10.1016/j.isprsjprs.2008.10.007.

[12] F. Ackermann, "Airborne laser scanning-present status and future expectations," *ISPRS J. Photogramm. Remote Sens.*, vol. 54, no. 2-3, pp. 64–67, Jul. 1999, doi: 10.1016/S0924-2716(99)00009-X.

[13] H.-G. Maas and G. Vosselman, "Two algorithms for extracting building models from raw laser altimetry data," *ISPRS J. Photogramm. Remote Sens.*, vol. 54, no. 2/3, pp. 153–163, Jul. 1999.

[14] F. Rottensteiner, J. Trinder, S. Clode, and K. Kubik, "Building detection by fusion of airborne laser scanner data and multi-spectral images: Performance evaluation and sensitivity analysis," *ISPRS J. Photogramm. Remote Sens.*, vol. 62, no. 2/3, pp. 135–149, Jun. 2007, doi: 10.1016/j.isprsjprs.2007.03.001.

[15] G. Miliaresis and N. Kokkas, "Segmentation and object-based classification for the extraction of the building class from LiDAR DEMs," *Comput. Geosci.*, vol. 33, no. 8, pp. 1076–1087, Aug. 2007, doi: 10.1016/j.cageo.2006.11.012.

[16] N. Haala, C. Brenner, and K.-H. Anders, "3D urban gis from laser altimeter and 2D map data," *Int. Arch. Photogramm. Remote Sens.*, vol. 32, pp. 339–346, 1998.

[17] D. Gonzalez-Aguilera, E. Crespo-Matellan, and D. Hernandez-Lopez, and P. Rodriguez-Gonzalvez, "Automated urban analysis based on LiDAR-derived building models," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1844–1851, Mar. 2013, doi: 10.1109/TGRS.2012.2205931.

[18] M. Adolphson, "Estimating a polycentric urban structure. Case study: Urban changes in the stockholm region 1991–2004," *J. Urban Planning Develop.*, vol. 135, no. 1, pp. 19–30, Mar. 2009, doi: 10.1061/(ASCE)0733-9488(2009)135:1(19).

[19] M. Adolphson, "Kernel densities and mixed functionality in a multicentred urban region," *Environ. Plann. B*, vol. 37, no. 3, pp. 550–566, 2010, doi: 10.1068/b35061.

[20] NRSA. (2006). *Cartosat-1 Data User's Handbook*, NRSA [Online]. Available: http://www.euromap.de/download/P5_data_user_handbook.pdf

[21] P. d'Angelo, M. Lehner, and T. Krauss, "Towards automated DEM generation from high resolution stereo satellite images," in *Proc. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. B4*, 2008, pp. 1137–1142.

[22] Euromap. (2013). *Euromaps 3D. Digital Surface Model, Product Specification* [Online]. Available: http://www.euromap.de/pdf/EuroMaps_3D_productinfo_V2.0_20130828.pdf

[23] M. Crespi *et al.*, "Orientation, orthorectification, DSM extraction and 3D city modeling by cartosat-1 stereo imagery: First results of a test over Rome," in *Proc. ISPRS Comm. IV Symp.*, Sep. 27–30, 2006, p. 6.

[24] J. Tian, P. Reinartz, P. d'Angelo, and M. Ehlers, "Region-based automatic building and forest change detection on Cartosat-1 stereo imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 79, pp. 226–239, May 2013, doi: 10.1016/j.isprsjprs.2013.02.017.

[25] P. Gamba, F. Dell'Acqua, and B. Dasarathy, "Urban remote sensing using multiple data sets: Past, present and future," *Inf. Fusion*, vol. 6, no. 4, pp. 319–326, Dec. 2005.

[26] R. Pal and K. Pal, "A review on image segmentation techniques," *Pattern Recognit.*, vol. 26, no. 9, pp. 1277–1294, 1993.

[27] G. Meinel, R. Hecht, H. Herold, and G. Schiller, "Automatische Ableitung von stadtstrukturellen Grundlagendaten und Integration in einem Geographischen Informationssystem," *Forschungen (BBR)*, vol. 134, p. 108, 2008 [Online]. Available: http://www2.ioer.de/recherche/pdf/2008_meinel_forschungen134.pdf

[28] G. Meinel, R. Hecht, and H. Herold, "Analyzing building stock using topographic maps and GIS," *Build. Res. Inf.*, vol. 37, no. 5–6, pp. 468–482, Nov. 2009, doi: 10.1080/09613210903159833.

[29] C. Brenner, "Interactive modeling tools for 3D building reconstruction," in *Proc. Photogramm. Week*, Stuttgart, Germany, 1999, pp. 23–34.

[30] C. Brenner, "Dreidimensionale Gebäuderekonstruktion aus digitalen Oberflächenmodellen und Grundrissen," Ph.D. dissertation, Univ. Stuttgart, Stuttgart, Germany, 2000, 124p.

[31] G. Vosselman and S. Dijkman, "3D building model reconstruction from point clouds and ground plans," in *Proc. ISPRS Int. Arch. Photogramm. Remote Sens., XXXII-3/W4*, Annapolis, MD, USA, 1999, pp. 37–43.

[32] L.-C. Chen, T.-A. Teo, C.-Y. Kuo, and J.-Y. Rau, "Shaping polyhedral buildings by the fusion of vector maps and lidar point clouds," *Photogramm. Eng. Remote Sens.*, vol. 74, no. 9, pp. 1147–1157, Sep. 2008, doi: 10.14358/PERS.74.9.1147.

[33] R. Wang, "3D building modeling using images and LiDAR: A review," *Int. J. Image Data Fusion*, vol. 4, no. 4, pp. 273–292, Dec. 2013, doi: 10.1080/19479832.2013.811124.

[34] A. Krehl, S. Siedentop, H. Taubenböck, and M. Wurm, "Agglomeration economies shaping the spatial structure in German city regions?," in *Proc. AESOP Annu. Congr.*, Utrecht/Delft, The Netherlands, 2014, p. 14.

[35] H. Hirschmüller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.

[36] GeoBasisNRW. (2013). "ATKIS—Digitale Topographische Karte 1:25.000 (DTK25)," *Bezirksregierung Köln* [Online]. Available: http://www.bezreg-koeln.nrw.de/brk_internet/presse/publikationen/geobasis/faltblatt_geobasis_atkis01.pdf

[37] BKG. (2014). *Digital Topographic Map 1:25 000, Preliminary Edition, DTK25-V* [Online]. Available: http://www.geodatenzentrum.de/docpdf/dtk25-v_eng.pdf

[38] Geodatenzentrum. (2014). *Digital Basis Landscape Model (AAA Modelling)* [Online]. Available: http://www.geodatenzentrum.de/docpdf/basis-dlm-aaa_eng.pdf

[39] BKG. (2013). *Digital Terrain Models for Germany* [Online]. Available: http://www.bkg.bund.de/nn_171762/EN/FederalOffice/Products/GeoData/Digital-Terrain-Models/DGM-Germany/DGMGermany__node.html__nnn=true#doc171766bodyText2

[40] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Proc. 3rd Eur. Conf. Comput. Vis.*, London, U.K., 1994, pp. 151–158.

[41] P. d'Angelo, "Image matching and outlier removal for large scale DSM generation," in *Proc. ISPRS Symp. Comm. I. Convergence Geomatics (CGC & ISPRS)*, Calgary, Canada, 2010, pp. 1–5.

[42] J. Grodecki and G. Dial, "Block adjustment of high-resolution satellite images described by rational polynomials," *Photogramm. Eng. Remote Sens.*, vol. 69, no. 1, pp. 59–68, Jan. 2003, doi: 10.14358/PERS.69.1.59.

[43] P. K. Srivastava *et al.*, "Recent advances in CARTOSAT-1 data processing," in *Proc. ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, Hannover, Germany, 2007, p. 10.

[44] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[45] P. d'Angelo and P. Reinartz, "DSM based orientation of large stereo satellite image blocks," in *Proc. ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XXXIX-B1*, Melbourne, Australia, 2012, pp. 209–214.

[46] T. Krauß and P. d'Angelo, "Morphological filling of digital elevation models," in *Proc. ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XXXVIII-4/W19*, Hannover, Germany, 2011, pp. 165–172.

[47] G. Grohman, G. Kroenung, and J. Strebeck, "Filling SRTM voids: The delta surface fill method," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 3, pp. 213–216, 2006.

[48] Definiens. (2009). *eCognition v 8.0* [Online]. Available: http://www.ecognition.com/sites/default/files/eCognition%20v8_Datasheet.pdf

[49] G. Gröger, T. H. Kolbe, C. Nagel, and K.-H. Häfele. (2012). *OGC City Geography Markup Language (CityGML) Encoding Standard* [Online]. Available: http://www.opengis.net/spec/citygml/2.0

[50] M. Wurm, H. Taubenböck, and S. Dech, "Quantification of urban structure on building block level utilizing multisensoral remote sensing data," in *Proc. SPIE Europe*, Berlin, Germany, 2010, p. 12.

[51] M. Wurm, H. Taubenböck, M. Schardt, T. Esch, and S. Dech, "Object-based image information fusion using multisensor earth observation data over urban areas," *Int. J. Image Data Fusion*, vol. 2, no. 2, pp. 121–147, Jun. 2011, doi: 10.1080/19479832.2010.543934.

[52] A. Forsyth. (2003). "Measuring density: Working definitions for residential density and building intensity," *Design Brief* [Online]. Available: http://www.corridordevelopment.org/pdfs/from_MDC_Website/db9.pdf

[53] S. Openshaw, *The Modifiable Areal Unit Problem*. Norwich, U.K.: Geo Books, 1983.

[54] EFGS. (2014). *GRID_ETRS89_LAEA_1K* [Online]. Available: http://www.efgs.info/data/european-datasets/eurogrid

[55] G. Foody, "Status of land cover classification accuracy assessment," *Remote Sens. Environ.*, vol. 80, no. 1, pp. 185–201, Apr. 2002.

[56] R. G. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Boca Raton, FL, USA: CRC Press, 2008, 200p.

[57] M. Rutzinger, F. Rottensteiner, and N. Pfeifer, "A comparison of evaluation techniques for building extraction from airborne laser scanning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 2, no. 1, pp. 11–20, Mar. 2009, doi: 10.1109/JSTARS.2009.2012488.

[58] C. Heipke, H. Mayer, C. Wiedemann, and O. Jamet, "Evaluation of automatic road extraction," in *Proc. Int. Arch. Photogramm. Remote Sens.*, Stuttgart, Germany, Sep. 17–19, 1997, vol. 32/3–4W2, pp. 151–160.

[59] J. Cohen, "A coefficient of agreement of nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, 1960.

[60] F. Rottensteiner, J. Trinder, S. Clode, and K. Kubik, "Using the Dempster–Shafer method for the fusion of LIDAR data and multi-spectral images for building detection," *Inf. Fusion*, vol. 6, no. 4, pp. 283–300, Dec. 2005.

Michael Wurm received the Diploma degree (Mag. rer. nat.) in geography with a specialization in remote sensing, GIS, and spatial research from the University of Graz, Graz, Austria, in 2007, and the Ph.D. degree (Dr. rer. nat.) in surveying and geoinformation from Graz University of Technology, Graz, Austria, in 2013.

He was with the Institute of Digital Image Processing at Joanneum Research, Graz, Austria, in 2007. In 2008, he joined the Remote Sensing Unit, Geography Department, University of Würzburg, Würzburg, Germany, working on the integration of earth observation data into social sciences and urban remote sensing as well as urban vulnerability research. In 2011, he joined the German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Wessling Germany. Working in the research unit "Modelling and geostatistical methods," he focuses on information extraction from multisensoral remote sensing data and its fusion with ancillary data. Until today, he was contributing to several research projects in urban remote sensing.

Pablo d'Angelo received the Diploma degree in computer engineering (Dipl.Ing. FH) from the University of Applied Sciences, Ulm, Germany, in 2004, and the Ph.D. degree (Dr.Ing.) in computer science from Bielefeld University, Bielefeld, Germany, in 2007.

From 2004 to 2007, he worked on industrial computer vision with Daimler AG, Stuttgart, Germany. In 2007, he joined the "Photogrammetry and Image Analysis," Department of Remote Sensing Technology Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Munich, Germany. His research interests include 3-D reconstruction from remotely sensed stereo imagery, with a focus on operational systems for large-scale image orientation and generation of digital elevation models.

Peter Reinartz (M'09) received the Diploma degree (Dipl.Phys.) in theoretical physics from the University of Munich, Munich, Germany, in 1983, and the Ph.D. degree (Dr.Ing) in civil engineering from the University of Hannover, Hannover, Germany, in 1989.

He is the Head of the "Photogrammetry and Image Analysis," Department, German Aerospace Center (DLR), Remote Sensing Technology Institute (IMF), Munich, Germany, and holds a Professorship of Geoinformatics with the University of Osnabrück, Osnabrück, Germany. He has more than 25 years of experience in image processing and remote sensing and 200+ publications in these fields. His research interests include direct georeferencing, stereo photogrammetry, data fusion of spaceborne and airborne data, generation of digital elevation models, and interpretation of VHR data from sensors like WorldView and Pleiades a.o. He is also engaged in using remote sensing data for disaster management and using high-frequency time series of airborne image data for real-time operations in case of disasters as well as for traffic monitoring.

Hannes Taubenböck received the Diploma degree in geography from the Ludwig-Maximilians Universität München, Munich, Germany, in 2004, and the Ph.D. degree (Dr. rer. nat.) in geography from the Julius-Maximilian's University of Würzburg, Würzburg, Germany, in 2008.

In 2005, he joined the German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Munich, Germany, working on the topic "Mega cities and Natural Disasters." He is a Lecturer with the University of Würzburg. In 2010, he was a Scientific Employee with DLR-DFD and became Head of the "Modelling and geostatistical methods" team in 2013. His research activities as Postdoc with the University of Würzburg (2007–2010) concentrated on the development of algorithms for information extraction from multisensoral remotely sensed data and subsequent value adding for transforming scientific results into practical value.