

Automatic Training Set Compilation with Multisource Geodata for DTM Generation from the TanDEM-X DSM

Christian Geiß, *Member, IEEE*, Patrick Aravena Pelizari, Stefan Bauer, Andreas Schmitt, and Hannes Taubenböck

Abstract—The TanDEM-X mission (TDM) is a spaceborne radar interferometer which delivers a global digital surface model (DSM) with a spatial resolution of 0.4 arcseconds. In this letter, we propose an automatic workflow for digital terrain model (DTM) generation from TDM DSM data through additional consideration of Sentinel-2 imagery and open-source geospatial vector data. The method includes the automatic and robust compilation of training samples by imposing dedicated criteria on the multisource geodata for subsequent learning of a classification model. The model is capable of supporting the accurate distinction of elevated objects and bare earth (BE) measurements in the TDM DSM. Finally, a DTM is interpolated from identified BE measurements. Experimental results obtained from a test site which covers a complex and heterogeneous built environment of Santiago de Chile, Chile, underline the usefulness of the proposed workflow, since it allows for substantially increased accuracies compared to a morphological filter-based method.

Index Terms— Automatic Training Sample Compilation, DTM Generation, Supervised Classification, TanDEM-X, Sentinel-2, OpenStreetMap

I. INTRODUCTION

The extraction of digital terrain models (DTMs) from digital surface models (DSMs) is a frequent task in Earth Observation data analysis. The combination of both models allows computing a normalized DSM (nDSM) which solely contains heights of objects which are elevated from the Earth’s surface such as buildings or vegetation. The extraction procedure foresees frequently a two-step approach: *First*, the elevation measurements (i.e., pixels) which represent bare earth (BE) are distinguished from elevation measurements which represent elevated objects (OBJ). *Second*, a DTM is generated by interpolating between identified BE pixels [1], [2].

Such procedures were developed primarily for very high spatial resolution (VHR) DSMs which are frequently acquired by e.g., airborne LiDAR campaigns. In comparison, the TanDEM-X mission (TDM), which is a spaceborne radar interferometer, provides a global elevation model with a consistent but considerably lower spatial resolution of 0.4 arcseconds (~12 meters) [1]. Tailored methods to extract objects such as buildings or vegetation from this model were already proposed (e.g., [1], [3], [4]). However, due to the coarser spatial resolution properties compared to DSMs derived from e.g., LiDAR, it remains challenging to accurately separate OBJ and BE pixels in the TDM elevation

measurements in an automated manner.

To enhance accuracy properties of previous approaches, here, we jointly exploit TDM DSM data, open-source geospatial vector data, and imagery from the Sentinel-2 (S-2) constellation. To this purpose, we establish a novel workflow which foresees the automatic compilation of training samples by imposing dedicated criteria on the multisource geodata for subsequent learning of a classification model which is capable of supporting the accurate distinction of OBJ and BE measurements.

Regarding the underlying methodological principle to automatically compile training samples, existing works exploited the idea of *rule-based training sample selection* in different application contexts: Huang *et al.* [5] deploy local top-of-atmosphere reflectance histograms computed from Landsat imagery to identify the so-called “forest peak” and select samples for supervised classification of forest areas thereof. With a focus on urban areas, Cao *et al.* [6] establish rules for DMSP-OLS and SPOT data to create a training set with “urban” and “non-urban” samples. Huang *et al.* [7] automatically generate labeled training samples for the same thematic classes by imposing constraints on the MODIS LandCover Type Yearly Global 500m product. Regarding mapping efforts with a higher thematic resolution, Huang *et al.* [8] use several spectral indices computed from VHR multispectral imagery and ancillary geospatial data (i.e., geospatial vector data from the OpenStreetMap (OSM) project) for automated compilation of a training set with labeled samples for six common land cover classes. In the context of DTM generation, Gevaert *et al.* [2] impose two threshold rules on the results of morphological top-hat filter operations on a VHR DSM for compilation of a training set containing samples for OBJ and BE measurements.

As can be seen from the literature, existing works either exploited the principle of rule-based training sample selection for land cover mapping using optical data or generate DTMs from VHR DSM data. To address the coarser resolution of the TDM DSM data, here, we uniquely combine optical and DSM data as well as geospatial vector data for an enhanced distinction of OBJ and BE measurements. Moreover, for a complete initial discrimination of OBJ and BE pixels in the TDM model, we built upon a tailored region-growing-based progressive morphological filtering (RPMF) procedure. In parallel, an initial training set is compiled by computing spectral indices from the S-2 imagery which are physically meaningful for six common land cover (LC) classes (i.e., “high vegetation”, “low vegetation”, “bare soil”, “building”, “water”, and “shadow”). Dedicated threshold-based rules are employed to automatically label samples thereof. In addition, a tailored road extraction (RE) module is implemented which

The work of Christian Geiß was supported by the Helmholtz Association under the grant “pre_DICT” (PD-305). Additionally, this research was funded in part by the German Federal Ministry of Education and Research (BMBF) under grant no. 03G0876 (project RIESGOS).

C. Geiß, P. Aravena Pelizari, S. Bauer, and H. Taubenböck are with the German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), 82234 Weßling-Oberpfaffenhofen, Germany (e-mail: christian.geiss@dlr.de; patrick.aravenapelizari@dlr.de; stefan.bauer@dlr.de; hannes.taubenboeck@dlr.de)

A. Schmitt is with the Munich University of Applied Sciences, Department of Geoinformatics, 80333 Munich, Germany (e-mail: andreas.schmitt@hm.edu)

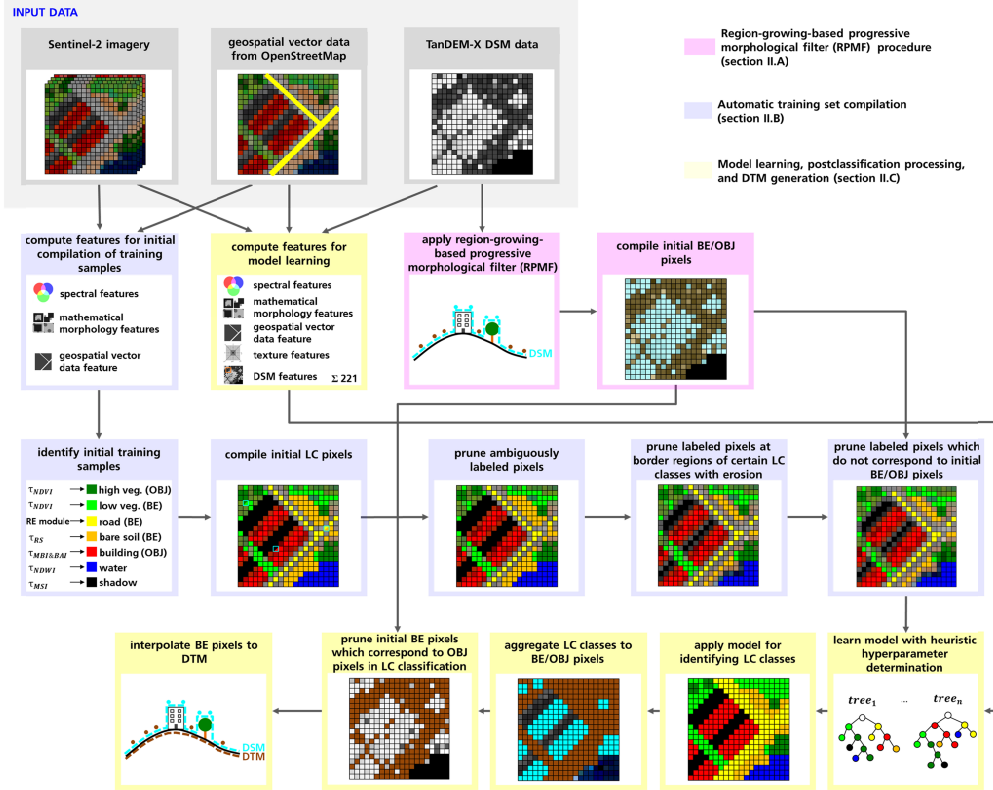


Fig. 1. Scheme of the automatic training set compilation approach for DTM generation.

additionally exploits geospatial vector data from OSM. Subsequently, labeled samples are pruned from the initial training set under consideration of consistency checks, topological constraints, and taking the initial outcomes of the RPMF procedure into account to only employ reliable samples for model learning. The remaining labeled samples are used for learning a Random Forest (RF) [9] model with heuristically determined hyperparameters. The model is learned from an extensive feature set which internalizes spectral information, transformations based on mathematical morphology, geospatial vector data, texture measures, and TDM DSM properties. Classification outcomes are aggregated to BE and OBJ pixels and are fused with the outcomes of the RPMF procedure to identify reliable BE pixels in the TDM model. Finally, a DTM is generated by interpolating between BE pixels. We evaluate our method in a comparative setup and provide experimental results from a test site which covers a complex and heterogeneous built environment of the city of Santiago de Chile, Chile.

The remainder of this letter is organized as follows. Section II documents the developed methods. Section III is used to describe study area, data sets, and parameterization of methods. The results of the actual experiments are reported in Section IV. We give concluding remarks in Section V.

II. PROPOSED METHODOLOGY

An overview of the method with individual processing steps is provided by Fig. 1.

A. Region-growing-based progressive morphological filter for initial compilation of BE/OBJ pixels

For an initial compilation of BE and OBJ pixels, which is used for both imposing constraints on the training set and identification of robust BE pixels in the TDM later on, a

RPMF procedure is followed. It is based on the idea of progressive morphological filters, which aim to discriminate ground and non-ground pixels in the DSM based on algebraic set operations. Such filters identify non-ground pixels in the DSM by gradually increasing the size of a structuring element (SE) and applying iteratively an elevation difference threshold. After the identification of initial non-ground pixels, here, potential non-ground pixels are identified within each iteration and their similarity with respect to neighboring non-ground pixels is assessed. Potential non-ground pixels are finally labeled as non-ground if they feature a high similarity to already identified non-ground pixels. Thereby, it is also made sure that water bodies are excluded from the analysis using the TDM water indication mask, which is an automatically processed information layer included in the TDM DSM product. After complete identification of OBJ pixels, residual pixels are labeled as BE to establish a complete allocation of TDM measurements according to BE and OBJ [1].

B. Automatic training set compilation from S-2 imagery

The initial compilation of the training set from the pool of unlabeled samples (i.e., the S-2 imagery) is based on six indices which are deployed for six different land cover classes (Table I). Characteristic numerical values of those indices are used to determine initial labeled samples for respective LC classes, whereby corresponding threshold values can be adapted for various image scenes [8]. To reliably identify road samples, which feature frequently similar spectral properties as other impervious surface-related LC classes, we implement a dedicated RE module. *First*, the multispectral imagery is subject to a binary conversion to impose noise removal and visibility of edges. *Second*, a texture image is computed from the binary image using Haralick's mean texture feature [11] to

TABLE I.
SPECTRAL INDICES USED FOR INITIAL TRAINING SET COMPILATION

Index	Description	Formula	LC class	thresholds
NDVI	The normalized difference vegetation index (NDVI) internalizes different reflection properties of vegetation in the red (I_{red}) and NIR (I_{NIR}) band, respectively. High numerical values indicate photosynthetically active vegetation.	$NDVI = \frac{I_{NIR} - I_{red}}{I_{NIR} + I_{red}}$	low vegetation	$NDVI > 0.55$
			high vegetation	$NDVI \leq 0.55$ $\cap NDVI > 0.3$
NDWI	The normalized difference water index (NDWI) highlights LC classes such as water, which feature a strong reflectance in the green band (I_{green}) and strong absorption in the NIR band.	$NDWI = \frac{I_{green} - I_{NIR}}{I_{green} + I_{NIR}}$	water	$NDWI \leq 0.35$ $\cap NDWI > 0.2$
R_{SI}	The saturation-intensity ratio index (R_{SI}) is computed from a conversion of the normalized false color composite image into the Hue-Saturation-Intensity space. Bare soil shows low numerical values in this domain.	$R_{SI} = \frac{\text{saturation} - \text{intensity}}{\text{saturation} + \text{intensity}}$	bare soil	$R_{SI} \leq -0.45$ $\cap R_{SI} > -0.6$
BAI	The built-up area index (BAI) considered here deploys the blue band (I_{blue}) and NIR band to highlight impervious surfaces including buildings.	$BAI = \frac{I_{blue} - I_{NIR}}{I_{blue} + I_{NIR}}$	building	$BAI \leq -0.1$ $\cap BAI > -0.25$
MBI	The morphologic building index (MBI) builds upon the circumstance that the relatively high reflectance of roofs and adjacent shadows induce high local contrast of buildings. It employs the sum of the differential morphologic profile (DMP) of the white-top-hat (WTH) transform. Thereby, γ_B is the opening-by-reconstruction of the brightness image (B) and s the parameter of scale. Huang <i>et al.</i> [10] use DMPs based on a linear SE applied with varying directions d . To reduce computation time we use a disc-shaped SE instead and vary solely its size (i.e., scale s).	$MBI = \sum_s DMP_{WTH(s)}, WTH(s)$ $= B - \gamma_B(s)$		$MBI > 5$
MSI	The morphologic shadow index (MSI) utilizes the fact that shadows feature a low reflectance and high local contrast. Contrary to the MBI, the MSI utilizes a black-top-hat (BTH) transform where φ_B is the closing-by-reconstruction of B [10].	$MSI = \sum_s DMP_{BTH(s)}, BTH(s)$ $= \varphi_B(s) - B$	shadow	$MSI \leq 13$ $\cap MSI > 12$

further reduce noise and stress edges related to roads in the image. *Third*, the resulting image is subject to an edge extraction procedure (i.e., a gradient filter) to identify candidate road segments. *Finally*, candidate road segments are combined with buffered geospatial road vector data from OSM and labeled as road samples if an intersection exists.

An initial training set is compiled based on the described procedure for seven LC classes. However, to enhance reliability of the labeled samples (i.e., reduce commission errors), a three-step pruning strategy is followed. *First*, labeled samples are pruned from the training set if they feature ambiguous labels (i.e., more than one LC class was assigned to a single pixel). *Second*, labeled samples at border regions of the LC classes “water”, “shadow”, and “bare soil” are also pruned from the training set by applying an erosion operation [12]. This is done to reduce the likelihood of including mixed pixels in the training set, which can be dominantly found at the border regions of LC classes. *Third*, it is verified that the remaining labeled samples also correspond to the initial outcome of the RPMF procedure, i.e., the samples of the LC classes “high vegetation”, and “buildings” feature the label “OBJ” in the intimal outcome of the RPMF procedure, whereas the samples of the LC classes “low vegetation”, “roads”, and “bare soil” feature the label “BE”. Initially labeled samples which correspond to one or multiple criteria of the pruning strategy are added to the pool of unlabeled samples again.

C. Model learning, postclassification processing and DTM generation

The remaining labeled samples in the training set are used for learning a RF model based on an extensive feature set. The feature set comprises S-2 spectral information, which is encoded by the spectral indices which were also used for the training set compilation, additional band ratios, as well as tasseled cap transformations. Beside MBI and MSI, further features from mathematical morphology are employed. In this

manner, we conducted opening and closing operations and computed reconstruction and top-hat profiles [12]. The OSM geospatial vector data feature which was used for identifying “road” samples is also considered here by means of a binary coding scheme. First and second order texture features are also computed, whereas the latter are based on the co-occurrence matrix [11]. Lastly, TDM DSM properties are encoded by means of measures of central tendency and spread computed from moving windows with various sizes on the DSM greyscale image. Given the different feature groups and various window sizes for the deployed spatial features, finally, each pixel is represented by a 221-dimensional feature vector.

The RF model is learned with a heuristic determination of hyperparameters, i.e., the number of classification trees to be grown n_{tree} and the number of features m_{try} used at each node is set to 500 and \sqrt{p} (with p denoting the number of the input features), respectively, since this parameterization can yield near optimum classification results [9]. The learned model is applied to identify the LC class of each pixel. Subsequent to this, the LC classes are aggregated to BE/OBJ pixels and BE pixels in the intimal outcome of the RPMF procedure are pruned if they correspond to OBJ pixels in the LC classification.

The identified BE pixels are used to interpolate a DTM. Thereby, an exact interpolation method, which predicts a value that is equal to the actual value at a sampled location and interpolates solely between values of sampled locations, is favorable to keep elevation values of BE pixels unaltered.

III. DATA AND EXPERIMENTAL SETUP

A. Data

The test site contains 342,875 pixels which cover parts of the city of Santiago de Chile, Chile. Generally, the TDM elevation model can be regarded as a DSM, especially when analyzing built environments. Only few surfaces such as ice, snow, or vegetation can be penetrated by the X-band SAR signal. Comparisons to ICESat data underline the high quality of

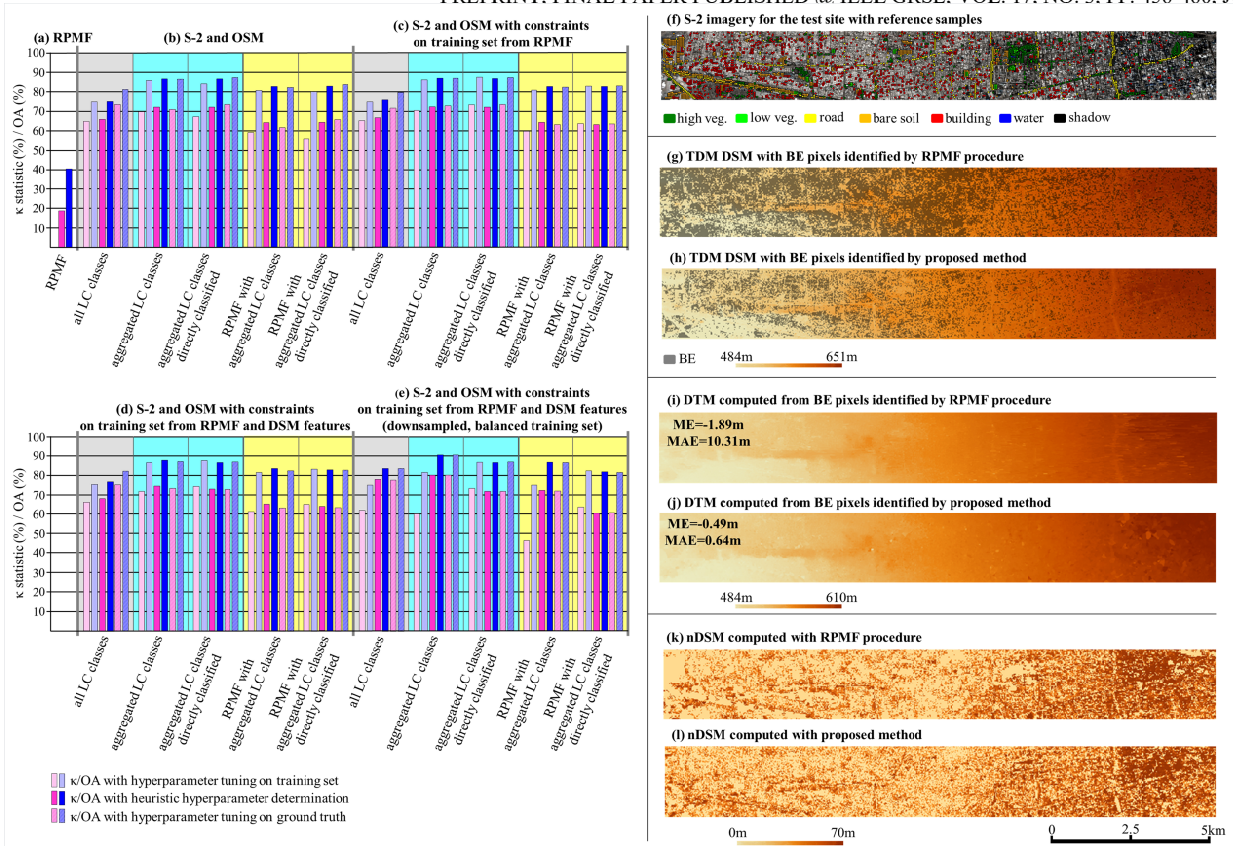


Fig. 2. Experimental results; classification accuracy measures when (a) deploying the RPFM procedure, (b) using information from S-2 and OSM to compile the training set and feature vector for classification, (c) also including the initial outcome of the RPFM procedure to compile the training set, (d)-(e) and additionally using DSM features and learning the model with a downsampled balanced training set, respectively; (f) S-2 imagery of the test site with reference samples; (g)-(h) TDM DSM data with superimposed BE pixels as identified by the RPFM procedure and proposed method, respectively; (i)-(j) DTMs interpolated from BE pixels of the RPFM procedure and proposed method, respectively; (k)-(l) nDSMs computed based on the RPFM procedure and proposed method, respectively.

elevation measurements, which feature less than one meter deviation in absolute vertical accuracy for surfaces as in this study. The high temporal resolution of S-2 allows obtaining numerous images over an area. To compile cloud-free S-2 image data, we pruned images from the stack of available images which featured significant cloud coverage. The remaining images were used to compute the median from the numerical values of each image with respect to each pixel for the blue, green, red, and NIR band, which feature a spatial resolution of 10 meters. VHR optical imagery was deployed in addition to the S-2 imagery and TDM DSM data to compile a reference data set based on manual image interpretation (Fig. 2f).

B. Experimental Setup

Regarding the RPFM procedure, the SE of the morphological filter must always exceed a building's outline to ensure extraction of all buildings present in the area under investigation. Consequently, the side length of the SE was determined empirically for the study areas according to the largest building present in the data. Additionally, an elevation difference threshold and a similarity constraint are to be determined. Both were set in accordance to previous experimental analysis, to enable a favorable trade-off between decrease of omission errors and increase of commission errors when classifying ground pixels (i.e., 2.6 meters for the elevation difference threshold and one meter for the similarity constraint as suggested in [1]). TDM DSM data were resampled and properly aligned to the S-2 imagery. Regarding

model learning, it was ensured that training and test areas are strictly spatially disjoint since the deployment of spatial features can induce a bias which leads to overoptimistic estimations of model accuracy [13]. Finally, for the DTM generation, we used inverse distance weighting as interpolation method.

IV. EXPERIMENTAL RESULTS

First, global accuracy measures (i.e., κ statistic and overall accuracy (OA)) for RPFM (Fig. 2a) and variations of the proposed method are presented (Fig. 2b-e). Regarding the latter, accuracies for different strategies for hyperparameter selection, i.e., exhaustive hyperparameter tuning based on the automatically compiled training set, heuristic hyperparameter determination, and hyperparameter tuning based on the reference data (the latter serves as a theoretical benchmark here since reference data is not available when aiming for a fully automatic workflow), are also documented. Thereby, κ statistic is consistently used as a global accuracy measure for model selection. In addition, accuracy measures for different ways to identify LC classes are presented: accuracies when classifying all seven LC classes (grey background in Fig. 2b-e); accuracies when aggregating the seven LC classes to BE/OBJ; or, as a variant of the proposed workflow, directly classify solely BE/OBJ pixels (i.e., aggregating the training samples accordingly) (cyan background in Fig. 2b-e). Lastly, accuracy measures for the final BE/OBJ discrimination in the DSM are presented when intersecting the LC classification with the initial outcome of RPFM (yellow background in Fig.

2b-e) using either aggregated LC classes (i.e., BE/OBJ) from the seven LC classes or, as a variant to the proposed workflow, using the aggregated LC classes which were directly classified. Thereby, results are presented for variations of the proposed method to quantify the contribution of individual processing steps: **Fig. 2b** shows results when solely information from S-2 and OSM is used to compile the training set and feature vector for classification (i.e., the third step of the three-step pruning strategy (cf. section II.B) was not carried out and DSM features were not included for the classification process). **Fig. 2c** shows results where initial outcomes of RPMF are also deployed to compile the training set and the results shown in **Fig. 2d** were obtained with the inclusion of DSM features. Finally, **Fig. 2e** provides a further variant where samples of certain LC classes in the training set were pruned (i.e., downsampled) to establish a balanced training set. It can be noted that **Fig. 2d-e** document the approach which internalizes the most exhaustive exploration of the given input data.

First, it can be recognized that the accuracies for the LC classification with all seven classes show consistently a substantial agreement: κ statistics and OAs increase in a strictly monotonic way from 66% to 78% and from 75% to 84%, respectively, depending on the level of input data exploration. The accuracy levels of the aggregated LC classification clearly favor the approach which foresees the aggregation of the seven LC classes as proposed throughout the description of the workflow in contrast to the direct classification of BE/OBJ pixels. This circumstance can be related to the fact that the aggregated treatment of the LC classes for model learning can impose divergence in the class patterns in feature space since LC classes with similar spectral properties do not constitute their own distinctive class pattern in feature space anymore but belong to opposite aggregated classes. The same holds true for the accuracy levels when comparing final BE/OBJ discriminations in the DSM: the accuracies obtained with the aggregated LC classification show superior accuracy measures which range from 64% to 72% and from 83% to 87% regarding κ statistic and OA, respectively. The reported numbers were obtained with a heuristic hyperparameter determination. This approach provided more favorable results here compared to exhaustive hyperparameter tuning. In this manner, it can be noted that model selection based on the automatically compiled labeled samples and exhaustive hyperparameter tuning may fit the training data well but can fail to properly generalize for unseen samples. However, the numerical results unambiguously underline the usefulness of the exhaustive exploration of the given input data which allows vastly increasing the initial accuracy of the RPMF procedure. Thereby, downsampling of the majority classes to establish a balanced training set also proved useful here in achieving the most favorable accuracy levels.

Fig. 2g-h shows the identified BE pixels with the RPMF procedure and the proposed method (based on the most exhaustive exploration of the given input data and a balanced training set), respectively. The proposed method enables a decrease of commission errors with respect to BE pixels: areas which were mistakenly labeled as BE by RPMF such as settlement areas with connected highly dense building

structures with low heights are excluded. The corresponding DTMs (**Fig. 2i-j**) underline the importance of the enhanced identification of BE pixels: the DTM computed based on the proposed method features both a decreased mean error (ME) and mean absolute error (MAE) (The measures were computed from the reference BE pixels in the DSM to enable a relative comparison with respect to the methods. An evaluation based on external DTM data from e.g., LiDAR measurements was not followed here to avoid an amalgamated assessment of both methods and data quality of the TDM DSM). The resulting nDSMs (**Fig. 2k-l**) mirror the previous findings: the nDSM computed with the proposed method contains numerous elevated objects which are not captured otherwise. As such, results unambiguously underline the beneficial joint exploitation of TDM DSM data, S-2 imagery and geospatial vector data for discriminating BE/OBJ pixels.

V. CONCLUSIONS

We proposed a novel workflow for DTM generation which foresees the automatic compilation of training samples by imposing dedicated criteria on TDM DSM data, OSM geospatial vector data and multispectral S-2 imagery. A classification model is learned from the training samples which is capable of supporting the accurate distinction of OBJ and BE measurements in TDM DSM data. Finally, identified BE measurements are deployed to interpolate a DTM. Experimental results underline the usefulness of the proposed workflow, since it allows for substantially increased accuracies compared to a morphological filter-based method.

REFERENCES

- [1] C. Geiß *et al.*, "Normalization of TanDEM-X DSM Data in Urban Environments with Morphological Filters," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4348–4362, 2015.
- [2] C. M. Gevaert *et al.*, "A deep learning approach to DTM extraction from imagery using rule-based training labels," *ISPRS J. Photogramm. Remote Sens.*, vol. 142, pp. 106–123, 2018.
- [3] C. Geiß *et al.*, "Estimation of Seismic Vulnerability Levels of Urban Structures With Multisensor Remote Sensing," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 9, no. 5, pp. 1913–1936, 2016.
- [4] J. Schreyer *et al.*, "TanDEM-X for Large-Area Modeling of Urban Vegetation Height: Evidence from Berlin, Germany," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 9, no. 5, pp. 1876–1887, 2016.
- [5] C. Huang *et al.*, "Use of dark object concept and support vector machines to automate forest cover change analysis," *Remote Sens. Environ.*, vol. 112, pp. 970–985, 2008.
- [6] X. Cao *et al.*, "A SVM-based method to extract urban areas from DMSP-OLS and SPOT VGT data," *Remote Sens. Environ.*, vol. 113, pp. 2205–2209, 2009.
- [7] X. Huang *et al.*, "Mapping Urban Areas in China Using Multisource Data With a Novel Ensemble SVM Method," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, Aug., pp. 4258–4273, 2018.
- [8] X. Huang *et al.*, "Automatic Labelling and Selection of Training samples for High-Resolution Remote Sensing Image Classification over Urban Areas," *Remote Sens.*, vol. 7, no. 8, pp. 16024–16044, 2015.
- [9] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [10] X. Huang *et al.*, "Morphological Building/Shadow Index for Building Extraction From High-Resolution Imagery Over Urban Areas," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 5, no. 1, pp. 161–172, Feb. 2012.
- [11] R. M. Haralick *et al.*, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern. SMC-3*, pp. 610–621, 1973.
- [12] P. Soille, *Morphological Image Analysis: Principles and Applications*. 2nd ed. Berlin, Germany: Springer-Verlag, 2004.
- [13] C. Geiß *et al.*, "On the Effect of Spatially Non-disjoint Training and Test Samples on Estimated Model Generalization Capabilities in Supervised Classification with Spatial Features," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2008–2012, 2017.