

This is the accepted manuscript version of the contribution published as:

Zhou, T., Geng, Y., Chen, J., Pan, J., **Haase, D., Lausch, A.** (2020):
High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM
derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms
Sci. Total Environ. **729** , art. 138244

The publisher's version is available at:

<http://dx.doi.org/10.1016/j.scitotenv.2020.138244>

High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms

Tao Zhou, Yajun Geng, Jie Chen, Jianjun Pan, Dagmar Haase, Angela Lausch



PII: S0048-9697(20)31757-5

DOI: <https://doi.org/10.1016/j.scitotenv.2020.138244>

Reference: STOTEN 138244

To appear in: *Science of the Total Environment*

Received date: 12 January 2020

Revised date: 7 March 2020

Accepted date: 25 March 2020

Please cite this article as: T. Zhou, Y. Geng, J. Chen, et al., High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms, *Science of the Total Environment* (2020), <https://doi.org/10.1016/j.scitotenv.2020.138244>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms

Tao Zhou^{a, b, *}, Yajun Geng^c, Jie Chen^c, Jianjun Pan^c, Dagmar Haase^{a, b}, Angela Lausch^{a, b}

^a Humboldt-Universität zu Berlin, Department of Geography, Unter den Linden 6, 10099 Berlin, Germany

^b Helmholtz Centre for Environmental Research – UFZ, Department of Computational Landscape Ecology, Permoserstraße 15, 04318 Leipzig, Germany

^c Nanjing Agricultural University, College of Resources and Environmental Sciences, Weigang 1, 210095 Nanjing, China

* Corresponding author.

E-mail address: tao.zhou@ufz.de (T. Zhou).

Abstract

Soil organic carbon (SOC) and soil total nitrogen (STN) are important indicators of soil health and play a key role in the global carbon and nitrogen cycles. High-resolution radar Sentinel-1 and multispectral Sentinel-2 images have the potential to investigate soil spatial distribution information over a large area, although Sentinel-1 and Sentinel-2 data have rarely been combined to map either SOC or STN content. In this study, we applied machine learning techniques to map both SOC and STN content in the southern part of Central Europe using digital elevation model (DEM) derivatives, multi-temporal Sentinel-1 and Sentinel-2 data, and evaluated the potential of different remote sensing sensors (Sentinel-1 and Sentinel-2) to predict SOC and STN content. Four machine-learners including random forest (RF), boosted regression trees (BRT), support vector machine (SVM) and Bagged CART were used to construct predictive models of SOC and STN contents based on 179 soil samples and different combinations of environmental covariates. The performance of these models was evaluated based on a 10-fold cross-validation method by three statistical indicators. Overall, the BRT model performed better than RF, SVM and Bagged CART, and these models yielded similar spatial distribution patterns of SOC and STN. Our results showed that multi-source sensor methods provided more accurate predictions of SOC and STN contents than individual sensors. The application of radar Sentinel-1 and multispectral Sentinel-2 images proved useful for predicting SOC and STN. A combination of Sentinel-1/2-derived predictors and DEM derivatives yielded the highest prediction accuracy. The prediction accuracy changed with and without the Sentinel-1/2-derived predictors, with the R^2 for estimating both SOC and STN content using the BRT model increasing by 12.8% and 18.8%, respectively. Topographic variables were the main explanatory variables for SOC and STN predictions, where

elevation was assigned as the variable with the most importance by the models. The results of this study illustrate the potential of free high-resolution radar Sentinel-1 and multispectral Sentinel-2 data as input when developing SOC and STN prediction models.

Keywords: soil organic carbon, soil total nitrogen, Sentinel-1, Sentinel-2, digital soil mapping, machine learning

1. Introduction

Soil is one of the most important carbon and nitrogen pools in terrestrial ecosystems, playing a key role in the global carbon and nitrogen cycles (Lal, 2008; Lausch et al., 2019). Soil organic carbon (SOC) and soil total nitrogen (STN) are important attributes of soil quality and fertility, and understanding their spatial variability is necessary to maintain food security and improve environmental quality in the context of global environmental change (Gholizadeh et al., 2018; Zeraatpisheh et al., 2019). Unfortunately, traditional soil mapping methods that rely on ground-based surveys are time consuming and expensive (Forkuor et al., 2017). Therefore, an accurate prediction of SOC and STN based on robust and cost-effective approaches is essential (Bou Kheir et al., 2010).

Digital soil mapping provides an efficient and convenient technique for obtaining reliable predictions of soil properties. Based on the numerical relationship between soil properties and predictor variables, digital soil mapping uses spatial analysis and mathematical methods to understand the spatial patterns of soil properties (Wadoux, 2019). Based on the digital soil mapping framework proposed by McBratney et al. (2003), numerous machine learning algorithms have been successfully used to predict soil properties. Machine learning methods learn the

relationship between soil properties and predictor variables, and then apply the learned relationships to areas where there is no soil data (Heung et al., 2016). Among these methods, tree-based models are probably the most commonly used learners and have been reported to have good performance in predicting SOC and STN, including classification and regression trees (CART) (Wiesmeier et al., 2011), boosted regression trees (BRT) (Bian et al., 2019), and random forest (RF) (Zhang et al., 2019). Other machine learning algorithms that have been applied to map soil properties include k-nearest neighbor (kNN) (Suominen et al., 2013), artificial neural networks (ANN) (Behrens et al., 2005), and support vector machines (SVM) (Morellos et al., 2016; Wu et al., 2018). However, choosing the best modeling technique for a given landscape has always been a challenge for soil property mapping.

Predictor variables from various sources can be combined with predictive models to map soil properties, including digital elevation models (DEM) and their derivatives, remote sensing images, and other sources of environmental variables (such as climate variables). A large number of existing DEM data sets (e.g., SRTM DEM and Aster GDEM) (Hu et al., 2017; Patel et al., 2016) can be used to extract a variety of readily available terrain parameters (e.g., elevation, slope, and the wetness index) as predictors for predicting soil properties. Remote sensing images provide an attractive source of data for the qualitative and quantitative study of soil properties and have been successfully applied to map various properties such as soil moisture (e.g., Peng et al., 2017; Sadeghi et al., 2017), pH (e.g., Roelofsen et al., 2015), texture (e.g., Gomez et al., 2019), bulk density (e.g., Ramcharan et al., 2018), cation exchange capacity (e.g., Keshavarzi et al., 2017), soil total phosphorus (e.g., Shen et al., 2019), available potassium (e.g., Dong et al., 2019), SOC (e.g., Ceddia et al., 2017), and STN (e.g., Kalambukattu et al., 2018). Previous studies of predicting

SOC and STN primarily utilized multispectral optical sensors, including Landsat (e.g., Yang et al., 2015), MODIS (e.g., Sreenivas et al., 2016), SPOT (e.g., Vaudour et al., 2013), RapidEye (e.g., Forkuor et al., 2017), and Pleiades-1A (e.g., Xu et al., 2018). However, optical sensors used to predict soil properties are limited due to their vulnerability to cloudy and rainy weather (Lausch et al., 2016; Zhou et al., 2018a).

Synthetic aperture radar (SAR) technology has great prospects for soil mapping due to its all-weather, day and night imaging advantages. SAR images have been widely used in the mapping of soil properties, especially physical properties such as soil moisture (e.g., Hosseini and McNairn, 2017) and texture (e.g., Gorraeb et al., 2015). We noticed that recent studies (e.g., Ceddia et al., 2017; Poggio and Gimona, 2017; Yang and Guo, 2019a; Yang and Guo, 2019b) of soil mapping have explored the potential of radar sensors to predict soil chemical properties through SAR techniques, suggesting that SAR data might be useful in predicting soil chemical properties. However, the application of SAR images has not been well developed in digital soil mapping compared to optical images, in part because of the complexity, diversity and availability of SAR data (Ma et al., 2017; Veloso et al., 2017). In addition, the limited availability of free and open high-resolution multispectral and SAR images limits the development of remote sensing-based soil mapping. The newly released Sentinel satellites developed by the European Space Agency (ESA) provide a large number of free remote sensing data with high spatial resolution for soil mapping. Sentinel-1 (S1) (6-day revisit) and Sentinel-2 (S2) (5-day revisit) are equipped with high spatial resolution SAR (5-20 m) and multispectral (10-60 m) sensors, respectively (Byrd et al., 2018; Wang et al., 2019b), providing new opportunities for the quantitative estimation of soil properties. The relatively high spatial resolution of S1 and S2 images helps us to understand the

local scale variability of soil properties. Some soil mapping studies have explored the potential of S2 data in predicting SOC and STN (e.g., Castaldi et al., 2019b; Gholizadeh et al., 2018; Vaudour et al., 2019a; Zhang et al., 2019). The use of S1 data for mapping SOC and STN has also been carried out by several scholars (e.g., Ma et al., 2017; Yang and Guo, 2019b; Yang et al., 2019). Some authors have predicted soil properties based on combinations of remotely sensed images with different characteristics and found that multi-sensor methods can improve soil mapping (Alexakis et al., 2017; Bousbih et al., 2019; Zeng et al., 2019). Although some progress has been made in mapping SOC and STN using S1 and S2 data, few studies have combined S1 and S2 data to predict both SOC and STN content.

The main purpose of this study was to map the SOC and STN contents in the southern part of Central Europe using DEM derivatives, S1 and S2 data by comparing four machine learning methods, and in particular to assess the capability of different remote sensing sensors to map SOC and STN content. For this purpose, DEM, multi-temporal S1 and S2 images were obtained for generating environmental variables, including DEM derivatives and Sentinel-1/2-derived predictors. We used RF, BRT, SVM, and Bagged CART models to compare the prediction accuracy of different combinations of these predictor variables in predicting SOC and STN. At the same time, we compared and evaluated the potential of different remote sensing sensors (i.e. the S1 sensor, the S2 sensor, and multi-source sensors (i.e. S1 and S2) to map SOC and STN. We then investigated the importance of the generated predictor variables and the spatial variability of the SOC and STN maps of the study area.

2. Materials and methods

2.1. Study area

The study area is located in the southern part of Central Europe (latitude: 45.61°-47.08° N, longitude: 12.93°-16.27° E) mostly covering Slovenia and a small part of Austria and Italy (Fig. 1).

Considering the availability of Sentinel-1 data, we designed the extent of the study area based on the coverage of the Sentinel-1 image covering the study area with a path of 22 and frames from 437 to 438. Forests and cultivated land are the main types of land use, accounting for 55.3% and 34.5% of the study area, respectively. Cultivated land is mainly distributed in the southwest and northeast of the study area, where the main crops are cereals (e.g., corn and wheat), potatoes, sugar beets and fruits. The average altitude of the study area is 644 m with very complex terrain. The northwestern part of the study area is an Alpine region with high altitudes, while the east and northeast are fertile Subpannonia. Diverse climates are found in the study area: the coastal and northeastern regions have Mediterranean and continental climates, respectively, while the high mountain regions have an alpine climate (Dolšak et al., 2016; Gosar et al., 2016). The average annual precipitation and average annual temperature of the study area are 8.3 °C and 1380 mm, respectively (<https://www.worldclim.org/>). The main soil types in this study area are Cambisols and Leptosols (Vrščaj et al., 2017; Wrb, 2015).

2.2. Soil data source

A total of 179 topsoil samples were collected from the Land Use and Coverage Area Frame Survey (LUCAS) topsoil dataset provided by the “European Soil Data Centre (ESDAC)” (Panagos

et al., 2012). These 179 samples with SOC and STN data were all available soil data covering the entire study area in LUCAS topsoil dataset, all used to map both SOC and STN content. LUCAS is an in-situ survey project initiated by the Statistical Office of the European Union (EUROSTAT) (Panagos et al., 2013). In the 2009 LUCAS survey, approximately 20,000 topsoil (0-20 cm) samples were collected from 25 member states of the European Union (Ward et al., 2019). LUCAS soil data is one of the largest soil databases in the world and is available from ESDAC (Orgiazzi et al., 2018). The LUCAS topsoil samples were taken from all land use types and were mainly concentrated in agricultural areas (Stevens et al., 2013). The LUCAS topsoil sample has a density of about 1 per 199 km² (Ballabio et al., 2014) and its physical and chemical properties are analyzed by international standard methods (Tóth et al., 2013). The samples collected at each sampling point included 5 topsoil (0-20 cm) subsamples which were then mixed to form a single 500 g final sample (Silva et al., 2019). The final soil samples were air-dried and then analyzed in an ISO-certified laboratory.

2.3. Predictor variables

The predictor variables used in this paper included Sentinel-1/2-derived predictors and DEM derivatives. These variables were all continuous predictor variables obtained from various sources and converted to raster data (25 m resolution) in ArcGIS 10.4. The observed SOC and STN content and all predictor variables were transferred to a geographic information system in the UTM/WGS84 projection for future analysis.

2.3.1. Topographic variables

The 25-m resolution DEM data (EU-DEM v1.1) covering the entire study area was obtained from

the Copernicus land portal. This DEM is an upgrade of EU-DEM v1.0 generated by SRTM and ASTER GDEM data. Seven DEM derivatives were calculated, including topographic wetness index (TWI), slope, elevation, valley depth (VD), length-slope factor (LSF), channel network base level (CNBL), and catchment slope (CS). These terrain variables derived from DEM data were calculated using ArcGIS 10.4 and SAGA GIS software.

2.3.2. Remote sensing variables and processing

The remote sensing data used for modeling included S1 and S2 images downloaded from ESA. S1 consists of Sentinel-1A and Sentinel-1B (Zhou et al., 2018b), where Sentinel-1A has four imaging modes with different resolutions and provides C-band images. This paper acquired five Sentinel-1A images covering the entire study area, and the specific parameter information is shown in Table 1. The multispectral high-resolution instrument on Sentinel-2A has 13 spectral bands with different spatial resolutions: 10 m (B2-496.6 nm, B3-560.0 nm, B4-664.5 nm and B8-835.1 nm), 20 m (B5-703.9 nm, B6-740.2 nm, B7-782.5 nm, B8A-864.8 nm, B11-1613.7 nm and B12-2202.4 nm) and 60 m (B1-443.9 nm, B9-945.0 nm and B10-1373.5 nm). Nine cloudless Sentinel-2A (Level-1C level product) images captured on July 11, 2015 were collected.

The following pre-processing of SAR data was performed using SARscape 5.2: multi-look, coregistration, speckle filtering (a Lee filter with a 13×13 window (Lee, 1986)), geocoding, and radiometric calibration (Zhou et al., 2018a). The S1 data was geocoded using GMTED2010 (Danielson and Gesch, 2011); digital number (DN) values were transformed to a dB scale backscatter coefficient with a resolution of 25 m. The processing of Level-1C S2 data was based on Level-1B products by applying radiometric and geometric corrections with sub-pixel accuracy.

We used ENVI 5.5 software to pre-process S2 images through the FLAASH atmospheric model, including radiometric calibration and atmospheric correction (Cai et al., 2018; Stratoulas et al., 2015). These S2 images were then mosaiced and clipped to obtain an optical image covering the study area.

The backscatter coefficients of the VH and VV polarizations from the S1 images were calculated as environmental variables. The nine bands of S2 data were extracted for subsequent studies, including B2, B3, B4, B5, B6, B7, B8A, B11, and B12. In addition, three spectral indices were calculated as predictors, which were reported to be strongly correlated with SOC (Gholizadeh et al., 2018). These spectral indices were NDVI (Normalized Difference Vegetation Index), Enhanced Vegetation Index (EVI) (Huete et al., 1997) and Soil Adjusted Total Vegetation Index (SATVI) (Marsett et al., 2006), and their formulas are as follows:

$$NDVI = \frac{NIR-RED}{NIR+RED} \quad (1)$$

$$EVI = 2.5 \times \frac{NIR-RED}{NIR+6 \times RED - 7.5 \times BLUE + 1} \quad (2)$$

$$SATVI = \frac{SWIR_1-RED}{SWIR_1+RED+1} \times 2 - \frac{SWIR_2}{2} \quad (3)$$

where *BLUE*, *RED*, *NIR*, *SWIR₁* and *SWIR₂* correspond to the B2, B4, B8, B11, and B12 bands of the S2 image, respectively.

2.4. Modelling techniques

In this section four machine learning techniques for SOC and STN mapping were described. The attribute values (corresponding to soil data points) of the predictor variables in the raster format generated in Section 2.3 were extracted in ArcGIS 10.4 to build these models. We used R software (R Development Core Team, 2011) to optimize the parameters that need to be defined for the

prediction model. Then, the performance of prediction models with the best parameters based on different combinations of predictor variables was evaluated and compared.

2.4.1. Random forest

Developed from CART, the RF model is a tree-based ensemble machine learning technique applied to classification and regression (Breiman, 2001). A large number of trees are produced in the RF algorithm (Kim and Grunwald, 2016). The unique bootstrap sample (with replacement) from the original training data independently builds each tree in the forest. The use of bootstrap sampling makes RF less sensitive to overfitting and allows general errors to be estimated based on the remaining test sets (Out-Of-Bag (OOB) sample) (Wiesmeier et al., 2014).

There are two parameters in RF modeling that need to be defined by the user: (i) the number of regression trees generated in the forest (ntree), and (ii) the number of predictors randomly selected at each node (mtry). We performed a grid search through the “caret” package (Kuhn, 2015) in the R software to select the optimal model. The parameters that produced the lowest prediction error were set in the final model (Gholizadeh et al., 2018).

2.4.2. Boosted regression trees

The BRT model combines two powerful statistical techniques (boosting techniques and decision tree algorithms) to optimize predictive performance. It constructs regression trees to predict target variables and uses boosting techniques to improve predictive power by minimizing the risk of overfitting (Akpa et al., 2016; Lawrence et al., 2004). The purpose of the boosting technique is to randomly select a subset of data to iteratively fit the new tree model to minimize the loss function (Elith et al., 2008).

There are three key parameters that need to be defined in the BRT model: tree complexity (TC), the number of trees (NT) and the learning rate (LR) (Ottoy et al., 2017; Wang et al., 2018a). LR determines the contribution of each tree to the final model. TC controls the order of interactions that can be fitted. The combination of LR and TC determines the TN (Mosleh et al., 2016). In a similar way to the RF model, these three parameters were optimized by performing a grid search through the “caret” package (Ottoy et al., 2017). The combination of parameters yielding the lowest root mean square error was set for modeling.

2.4.3. Bagged classification and regression trees

CART is a nonparametric data mining technique for regression or classification problems that has continuously been improved (Hamze-Ziabari and Bakhshpoori, 2018). This modeling technique has been widely used for the prediction of soil properties (Guo et al., 2015; Heung et al., 2016). Based on the binary partitioning algorithm, CART recursively splits data to explore the relationship between the response variables and the predictors (Aertsen et al., 2010; Wiesmeier et al., 2011). Bagged CART is an improved CART algorithm that combines CART with bagging techniques to enhance the performance of predictive models and reduce overfitting (Deng et al., 2019; Lee et al., 2010; Ranaie et al., 2018). We used the “ipred”, “plyr” and “e1071” packages to fit the Bagged CART algorithm through the “treebag” method of the “caret” package. Parameters did not have to be defined when using the “treebag” method.

2.4.4. Support vector machine

Proposed by Cortes and Vapnik (1995), SVM is a popular supervised learning technique for classification and regression that can be generalized to nonlinear models using kernel functions.

The SVM model projects the data into the new hyperspace with the help of kernel functions. In the new hyperspace, the SVM searches for a hyperplane, which separates the classes and leaves the largest possible margin between the classes (Were et al., 2015). There are four kernel function types: polynomial, sigmoid, linear and radial basis function (RBF) (Pradhan, 2013). The choice of kernel functions and their parameters affects the accuracy of the analysis results of the SVM model. The RBF kernel, which has been widely used in soil mapping research (Ahmad et al., 2010; Taghizadeh-Mehrjardi et al., 2017; Wang et al., 2018a), was selected as the kernel model of the SVM algorithm. There are two parameters that need to be defined for the RBF kernel, including penalty (cost) and kernel width (sigma) (Jebur et al., 2015). These parameters were optimized using the grid search method in the "caret" package of the R software.

2.5. Statistical analyses

We used SPSS 24 software to perform descriptive statistical analysis of SOC and STN and Pearson correlation analysis to detect collinearity between predictor variables and their correlation with SOC and STN (Kempen et al., 2019). Highly correlated predictor variables ($r \geq 0.8$) with high variance inflation factors ($VIF \geq 10$) were removed from modeling (Lombardo et al., 2018; Were et al., 2015; Zhang et al., 2019). In this study, the "gbm" (Elith et al., 2008), "randomForest" (Liaw and Wiener, 2002), and "kernlab" (Karatzoglou et al., 2018) packages of the R software were used to develop the BRT, RF, and SVM models, respectively.

2.6. Methods for evaluating model performance

We constructed SOC and STN content models based on four machine learning techniques using different combinations of predictor variables (Table 2). Model A, Model B, and Model D included

only S2 images, S1 images, and DEM derivatives, respectively. Model B was a combination of S1 and S2 data, while Model E included all predictor variables. This study used a ten-fold cross-validation method to evaluate the predictive performance of these models. The cross-validation method builds some train/test splits in the observed data and ensures that each data point occurs at least once in the test set (Taghizadeh-Mehrjardi et al., 2016). For the ten-fold cross-validation, the observed data sets were randomly divided into 10 groups (Aitkenhead, 2017). In each of the 10 folds, one group was selected as the test data set and the other nine groups were used as the training set (Ottoy et al., 2017). When compared to using data splitting methods for validation, repeated splitting in cross-validation makes it more efficient than data splitting (Brus et al., 2011). Three validation criteria were calculated to evaluate the performance of the model: the root mean square error (RMSE), the mean absolute error (MAE) and the coefficient of determination (R^2). These validation criteria are calculated from (Wang et al., 2018b):

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - O_i| \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \quad (5)$$

$$R^2 = \frac{\sum_{i=1}^n (P_i - \bar{O})^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (6)$$

where n represents the number of samples; P_i and O_i represent the predicted and observed SOC and STN at site i , respectively.

In order to evaluate the prediction uncertainty, we selected the model that performed best in the four machine learning algorithms from the results of the prediction accuracy to produce one

hundred SOC and STN maps, respectively. Based on these SOC and STN maps obtained by running the model a hundred times, we calculated the mean and standard deviation (SD) of each pixel as their final map and prediction uncertainty, respectively. Similarly, SD maps have been used as indicators of uncertainty associated with model predictions in many previous soil mapping studies (Jeong et al., 2017; Song et al., 2018; Wang et al., 2019c).

3. Results

3.1. Descriptive analysis of SOC

Descriptive statistics of both SOC and STN content are shown in Table 3. Both SOC and STN content showed a strongly skewed distribution; their skewness coefficients were 3.75 and 2.63, respectively. For all prediction models, we converted the SOC content and the STN content by using a natural logarithm; the skewness coefficients of both SOC and STN content were reduced to 0.67 and 0.12, respectively. The mean SOC content and STN content were 49.99 g/kg (ranging from 4.70 to 439.10 g/kg) and 3.50 g/kg (ranging from 0.30 to 19.70 g/kg), respectively. The SD values of the STN content (2.60 g/kg) and the SOC (55.60 g/kg) content were lower and higher than their mean values, respectively; the SOC content was more variable than the STN content.

We found some collinearity of predictor variables through collinearity analysis. To avoid collinearity, the predictor variables used for SOC and STN prediction were reduced from 29 to 19. Pearson correlation coefficients of these selected predictor variables (Figure S1) with SOC and STN are shown in Figure S2. No collinearity problems were observed during further modeling because the VIFs of all selected predictor variables were less than 10 (Table S1).

3.2. Evaluation and comparison of different models

This study built the following five models using Sentinel-1/2 images and DEM derivatives: Model A, Model B and Model D represent Sentinel-2-derived predictors, Sentinel-1-derived predictors and topographic variables, respectively; Model C and Model E represent remote sensing variables (Sentinel-1/2-derived predictors) and all environmental variables, respectively. The performances of Bagged CART, RF, BRT and SVM based on these models in predicting both SOC and STN content are shown in Table 4. Our comparative analysis based on prediction accuracy showed that the choice of modeling techniques, the type of predictor variables (different combinations constructed by environmental variables) significantly affected the predictive performance of both SOC content and STN content. For example, for predicting SOC using SVM and RF, Model D ($R^2 = 0.39$ vs. $R^2 = 0.35$, respectively) and Model E ($R^2 = 0.43$ vs. $R^2 = 0.40$, respectively) were better predicted by the former, while SVM and RF methods performed similarly in Model A ($R^2 = 0.19$ vs. $R^2 = 0.20$, respectively), Model B ($R^2 = 0.16$), and Model C ($R^2 = 0.25$). On the other hand, in terms of STN prediction, the RF method performed better than the SVM in Model A ($R^2 = 0.15$ vs. $R^2 = 0.10$, respectively), Model B ($R^2 = 0.15$ vs. $R^2 = 0.11$, respectively), Model C ($R^2 = 0.18$ vs. $R^2 = 0.12$, respectively) and Model E ($R^2 = 0.38$ vs. $R^2 = 0.35$, respectively), while the latter showed better prediction performance in Model D ($R^2 = 0.28$ vs. $R^2 = 0.33$, respectively). For SOC prediction, overall, the BRT method showed the best performance; the three machine learning techniques (Bagged CART, RF and SVM) had similar prediction performance when using Model A and Model B; BRT and SVM obtained similar prediction accuracy in Model D and Model E. Among the four modeling techniques, BRT performed best in STN mapping from Model A to Model E, whereas RF in Model E showed similar levels of prediction accuracy as the BRT

method. This suggests that different predictive models might suit different environmental variables and soil properties.

Among the three types of predictors, DEM derivatives obtained higher prediction accuracy than S1 and S2 data. For the BRT model, S1 data achieved slightly higher prediction accuracy than S2. However, for the remaining prediction models (Bagged CART, RF and SVM), overall, S2 data yielded better predictions than S1. Although there was a strong difference in predictive performance across all predictive models and predictor types, the combination of S1 and S2 improved prediction accuracy compared to using a single type of remote sensing data. For example, the addition of S1 data increased the R^2 of the BRT model from 0.19 to 0.27 in predicting SOC and from 0.16 to 0.21 in predicting STN. This result indicates that S1 data contains valuable information that is not captured and that the inclusion of S1 improves overall predictive performance. This improvement was also verified using RF, Bagged CART and SVM models.

As we expected, the highest prediction accuracy was achieved when DEM derivatives, S1 and S2 images were applied together. The R^2 of the BRT model increased by 12.8% in predicting SOC (from 0.39 to 0.44) and by 18.8% in predicting STN (from 0.32 to 0.38). Similar improvements can be observed for other machine learning methods. This result further demonstrates that remote sensing variables extracted from multi-source sensors are critical for the efficient modeling and prediction of soil properties. The combination of all predictors had the best predictive performance, with the STN prediction obtained from the BRT ($R^2 = 0.38$, RMSE = 0.50, and MAE = 0.39) and RF ($R^2 = 0.38$, RMSE = 0.51, and MAE = 0.39) models and the SOC mapping generated by the BRT ($R^2 = 0.44$, RMSE = 0.57, and MAE = 0.45) and SVM ($R^2 = 0.43$,

RMSE = 0.57, and MAE = 0.45). The R^2 values indicated that these models could explain approximately 44% and 38% of the variability of SOC and STN, respectively.

3.3. Relative importance of predictor variables

For SOC and STN mapping using Model E, the rankings of predictor variables ordered by relative importance are shown in Figure 2 (the importance was converted to a percentage). The importance of variables in the RF and BRT models was slightly different, revealing different dominating environmental features in these models. For both RF and BRT models, DEM derivatives were the main explanatory variables for SOC and STN predictions (more than 50% of total relative importance), followed by S1 and S2 images. Although the three models exhibited different ranking characteristics of importance, among all predictors, elevation was the most important in SOC and STN predictions. The importance of elevation in all models ranked first, with a relative importance of more than 25%. In addition, S1 images in the RF and BRT models explained 39% and 29% of SOC variation and 40% and 37% of STN variation, respectively. This result reveals the potential application of S1 images for predicting SOC and STN in this study area.

3.4. Spatial characteristics of SOC and STN maps

Based on Model E, the maps of SOC and STN predicted by the four methods are presented in Figure 3. The mean and SD values of the predicted SOC content were 51.40 and 27.70 g/kg for RF, 55.07 and 34.22 g/kg for BRT, 51.93 and 33.02 g/kg for SVM, and 50.35 and 25.78 g/kg for Bagged CART, respectively. For the predicted STN content, the mean and SD values were 3.61 and 1.43 g/kg for RF, 3.64 and 1.52 g/kg for BRT, 3.46 and 1.45 g/kg for SVM, and 3.61 and 1.37

g/kg for Bagged CART, respectively.

Based on the analysis of modeling accuracy, the BRT method (in Model E) was selected to run 100 times to evaluate prediction uncertainty (Jeong et al., 2017). We calculated the average (final map) and SD (prediction uncertainty) maps from these 100 runs (Fig. 4). The average concentrations of SOC and STN from 100 runs were 54.79 and 3.62 g/kg, respectively. Both the SOC and STN prediction models exhibited low levels of uncertainty, with average SDs of 1.29 and 0.04 g/kg for the 100 predicted SOC and STN maps, respectively. The low SD values from Model E also indicate that the BRT has stable prediction capabilities.

4. Discussion

4.1. Performance of predictive models using DEM derivatives, Sentinel-1 and Sentinel-2 data.

The comparison of prediction accuracy in this study showed that the choice of machine learning algorithms and the types and combinations of environmental variables have a large impact on the prediction performance of SOC and STN (Table 4). Overall, the BRT model outperformed the other three prediction models. Our results are consistent with the results of Beguin et al. (2017) who reported that the choice of predictive models and the construction of predictors have an impact on the prediction of soil properties. At three research sites (Germany, Belgium, and Luxembourg), Castaldi et al. (2019b) used different types of remote sensing images for SOC prediction, and found that the type of remote sensing data, the choice of predictive models, and the study area had an impact on prediction accuracy. In addition, there are inconsistencies in various previous comparative studies. For example, Siewert (2018) reported that RF performed better than

SVM in mapping SOC in Sweden, while Were et al. (2015) found that the latter was better than the former in SOC mapping in Kenya. Nussbaum et al. (2018) and Wang et al. (2018a) also obtained the opposite comparison results in predicting soil properties using BRT and RF models. Based on this, no single predictive model performs best under all circumstances (Lamichhane et al., 2019), so we recommend using a specific experimental data set to calibrate competitive predictive models.

Our results revealed that DEM derivatives, S1 and S2 data are important for effective modeling of SOC and STN. Moreover, the combination of S1 and S2 data allowed us to improve prediction performance. Some studies have highlighted the importance of DEM derivatives (Tziachris et al., 2019) and S2 (Castaldi et al., 2019b; Wang et al., 2019a) images for predicting soil properties. However, previous studies of predicting soil properties primarily used only one type of sensor, mainly including Landsat and MODIS (Poggio and Gimona, 2017). For instance, Vaudour et al. (2019a) and Gholizadeh et al. (2018) explored the capability of S2 optical data to predict soil properties, but they did not consider the feasibility of radar sensors. In this study, the better prediction accuracy obtained from the combined S1 and S2 images demonstrates the usefulness of S1 radar data in predicting SOC and STN. The combination of optical and radar sensors has great potential for predicting soil properties, especially in areas that are susceptible to cloud cover.

In this study, the combined DEM derivatives and Sentinel-1/2-derived predictors achieved the highest modeling accuracy, explaining 44% and 38% of the SOC and STN variability, respectively (Table 4). Compared with previous studies that used LUCAS data to predict soil properties, the results of this study were inferior. For example, Wadoux (2019) developed an RF

model using LUCAS data in France that explain only 12% and 20% of SOC and STN variability, respectively. de Brogniez et al. (2015) used the LUCAS data to develop a generalized additive model (GAM) to explain 28% of SOC variability. In a SOC mapping study conducted in Europe using LUCAS data, Aksoy et al. (2016) could explain 40% of the SOC variability of topsoil. Yigini and Panagos (2016) used regression-kriging method to explain 40% in in Europe. Our prediction results were better than previous studies using LUCAS data, but high prediction accuracy has not been obtained. These differences in prediction accuracy may result from the study area, the type and quality of predictor variables, prediction models, and the variability of SOC and STN (Mosleh et al., 2016; Were et al., 2015).

Although the combined S1 and S2 data proved to be more powerful than the single sensor approach, further improvements may be needed. The successful application of remote sensing technology in predicting soil properties depends largely on the availability and quality of remote sensing images (Forkuor et al., 2017). Unfortunately, the soil sampling time and remote sensing data acquisition time in this study were inconsistent due to the unavailability of S1 and S2 data, which often occurred in previous digital soil mapping studies (Castaldi et al., 2019a; Poggio and Gimona, 2017; Vaudour et al., 2019b). Land surface characteristics change dynamically, so the impact of remote sensing data acquisition time on SOC and STN mapping needs further investigation.

4.2. Variable importance

DEM derivatives were identified as the most influential predictor variables for both SOC and STN predictions in our study, followed by S1 and S2 data (Fig. 2). Topography is one of the most

important factors in soil formation and DEM-based topographical variables have been commonly used as key predictors for digitizing soil mapping. The terrain controls the flow of solutes, water and sediments, which in turn affects soil development and the spatial distribution of soil properties (Li et al., 2013). Some topographical variables have been reported to have a highly significant correlation (e.g., elevation, slope, and TWI) with SOC and STN contents (Obu et al., 2017; Wang et al., 2017) and have a significant impact on a great range of soil properties (Gerrard, 1981). Similar to our results, Wang et al. (2018b) and Grimm et al. (2008) found that topographic variables were the most important predictors for SOC and STN mapping. Among all DEM derivatives, elevation played the most important role in our SOC and STN predictions (Fig. 2). Altitude plays an important role in the development of microclimates (Griffiths et al., 2009), which in turn affects the distribution of plant communities and soil processes (Bochet, 2015; Lozano-García et al., 2016). In previous studies of soil property predictions (Hinge et al., 2018; Ramifehiarivo et al., 2017), elevation was also found to be the most effective topographic parameter. Other DEM derivatives were also identified as important predictors (e.g., slope, TWI, and LSF) affecting the spatial distribution of STN and SOC in this study. Slope controls the movement and accumulation of water and matter in the landscape, and then contributes to the spatial variation of SOC and STN (Tsui et al., 2004). LSF describes soil transport in overland flow (Moore and Burch, 1986) and has been found to be an important variable explaining the SOC variation in central Chile (Reyes Rojas et al., 2018). The contribution of TWI in predicting soil properties can be explained by its identification of soil moisture gradients (Siewert, 2018). The remaining DEM derivatives were identified as key predictors of soil properties in previous studies (Adhikari et al., 2014; Li et al., 2017).

Remote sensing data has been successfully applied to digital soil mapping studies at the regional and the global level (Hengl et al., 2017; Kalambukattu et al., 2018; Zhang et al., 2019). However, most of these previous studies only focused on optical imagery. Our results revealed that both optical and SAR images were effective predictors for determining SOC and STN in Model E. This result is similar to previous studies, which reported that Sentinel-2-derived predictors are important factors in predicting SOC and STN distribution (Gholizadeh et al., 2018; Vaudour et al., 2019a). This is not surprising due to the close relationship between soil properties and vegetation cover, and the vegetation index captures changes in soil properties (Mahmoudabadi et al., 2017; Maynard and Levi, 2017). This was supported by Yang et al. (2016) and Wang et al. (2018c), who reported that vegetation index and remote sensing reflectance are important indicators for predicting soil properties. For example, as the most commonly used vegetation index for predicting soil properties, the NDVI has been reported to be a satisfactory predictor of SOC and STN (Page et al., 2013; Sumfleth and Duttman, 2008). The contribution of SAR data to the model depends on the sensitivity of the backscatter coefficient to changes in land surface conditions and soil moisture (Kasischke et al., 1997). Although there was a weak and fluctuating correlation, the significant correlation between the backscatter coefficient of S1 images and soil properties (including SOC and STN) was observed by Yang and Guo (2019b). Yang et al. (2019) reported that S1 images can be used to predict soil properties through their ability to capture characteristics of the short-term variation of vegetation.

4.3. Spatial characteristics of SOC and STN maps

The maps of predicted SOC content had a similar spatial pattern to the SoilGrid product (Hengl et

al., 2017). Broadly speaking, the four predictive models produced similar spatial distribution patterns of SOC and STN (Fig. 3). High SOC and STN concentrations were concentrated in the central and northwestern mountainous areas of the study area mainly covered by forests where the altitude was high. Lower SOC content and STN content were mostly found in farmland and other low-altitude land in the northeast and southwest of the study area. This result reflected the effect of elevation and DEM derivatives on the variability of SOC and STN. The important role of elevation in predicting soil properties has been confirmed by recent studies (Lamichhane et al., 2019; Song et al., 2016). Tsui et al. (2013) reported that elevation is an effective variable for predicting soil properties and that SOC stock increases with elevation. Differences in altitude gradients may indirectly affect soil carbon and nitrogen input and loss through its close relationship with vegetation cover, temperature and precipitation. In the mountain areas of this study area, relatively high rainfall and low temperatures are conducive to the accumulation of carbon in the soil (see de Luis et al. (2014) for rainfall and temperature distribution). Low-altitude land with relatively low rainfall and high temperatures in this study area had lower SOC content and STN content and was mainly covered by farmland, which was similar to the results of Were et al. (2015) who found higher SOC content in forest areas and lower SOC content in farmland areas. Wang et al. (2018b) found that the SOC content of farmland was lower than that of forests, which can be explained by the increase of organic matter decomposition and the loss by erosion and cultivation.

Accurate predictions of SOC and STN are important factors in mitigating climate change, running various environmental models, and improving agricultural farming measures. High-resolution free S1 and S2 data provide unique opportunities for prediction of soil properties

and environmental monitoring. Therefore, our results not only provide a theoretical basis for the development of high-resolution digital SOC and STN products based on a large amount of available S1 and S2 data, but are also crucial for managing climate change and optimizing land use.

5. Conclusions

We used DEM derivatives, S1 and S2 data to investigate the spatial distribution of both SOC and STN content in the southern part of Central Europe by comparing four machine learning techniques. Our main conclusions can be summarized as follows:

- Overall, the BRT model was superior to RF, SVM and Bagged CART in predicting SOC content and STN content and exhibited the best performance. The SOC and STN maps predicted by the four machine learning techniques had similar spatial distribution patterns. These maps showed significant spatial variability, with both high SOC and STN content in high altitude mountainous areas.
- The application of radar S1 and multispectral S2 images proved useful for predicting SOC and STN. Multi-source sensor methods showed significantly better predictive performance than single sensors. A large number of free and easily accessible S1 and S2 images may provide more opportunities to obtain soil spatial distribution information.
- The combination of Sentinel-1/2-derived predictors and DEM derivatives had the highest prediction accuracy. When comparing the prediction accuracy changes with and without remote sensing variables, the R^2 for predicting both SOC and STN content using the BRT model was improved by 12.8% and 18.8%, respectively.

- DEM derivatives were the main explanatory variables for SOC and STN predictions, followed by S1 and S2 data. Elevation was the most important environmental variable that affects the spatial distribution of SOC and STN.

Acknowledgments

Tao Zhou thanks the China Scholarship Council for its financial support. We are very grateful for the data support of the ESDAC.

References

- Adhikari, K., Hartemink, A.E., Minasny, B., Bou Kheir, R., Greve, M.B., Greve, M.H., 2014. Digital Mapping of Soil Organic Carbon Contents and Stocks in Denmark. *PLOS ONE* 9(8), e105519.
- Aertsen, W., Kint, V., van Orshoven, J., Özkan, K., Muys, B., 2010. Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. *Ecological Modelling* 221(8), 1119-1130.
- Ahmad, S., Kalra, A., Stephen, H., 2010. Estimating soil moisture using remote sensing data: A machine learning approach. *Advances in Water Resources* 33(1), 69-80.
- Aitkenhead, M.J., 2017. Mapping peat in Scotland with remote sensing and site characteristics. *Eur. J. Soil Sci.* 68(1), 28-38.
- Akpa, S.I.C., Odeh, I.O.A., Bishop, T.F.A., Hartemink, A.E., Amapu, I.Y., 2016. Total soil organic carbon and carbon sequestration potential in Nigeria. *Geoderma* 271, 202-215.
- Aksoy, E., Yigini, Y., Montanarella, L., 2016. Combining Soil Databases for Topsoil Organic Carbon Mapping in Europe. *PLOS ONE* 11(3), e0152098.
- Alexakis, D.D., Mexis, F.-D.K., Vozinaki, A.-E.K., Daliakopoulos, I.N., Tsanis, I.K., 2017. Soil Moisture

- Content Estimation Based on Sentinel-1 and Auxiliary Earth Observation Products. A Hydrological Approach. *Sensors* 17(6), 1455.
- Ballabio, C., Panagos, P., Montanarella, L., 2014. Predicting soil organic carbon content in Cyprus using remote sensing and Earth observation data. Second International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2014), 9229. SPIE.
- Beguín, J., Fuglstad, G.-A., Mansuy, N., Paré, D., 2017. Predicting soil properties in the Canadian boreal forest with limited data: Comparison of spatial and non-spatial statistical approaches. *Geoderma* 306, 195-205.
- Behrens, T., Förster, H., Scholten, T., Steinrücken, U., Spies, E.-D., Goldschmitt, M., 2005. Digital soil mapping using artificial neural networks. *Journal of Plant Nutrition and Soil Science* 168(1), 21-33.
- Bian, Z., Guo, X., Wang, S., Zhuang, Q., Jin, X., Wang, Q., Jia, S., 2019. Applying statistical methods to map soil organic carbon of agricultural lands in northeastern coastal areas of China. *Archives of Agronomy and Soil Science*, 1-13.
- Bochet, E., 2015. The fate of seeds in the soil: a review of the influence of overland flow on seed removal and its consequences for the vegetation of arid and semiarid patchy ecosystems. *SOIL* 1(1), 131-146.
- Bou Kheir, R., Greve, M.H., Bøcher, P.K., Greve, M.B., Larsen, R., McCloy, K., 2010. Predictive mapping of soil organic carbon in wet cultivated lands using classification-tree based models: The case study of Denmark. *Journal of Environmental Management* 91(5), 1150-1160.
- Bousbih, S., Zribi, M., Pelletier, C., Gorraab, A., Lili-Chabaane, Z., Baghdadi, N., Ben Aissa, N., Mougenot, B., 2019. Soil Texture Estimation Using Radar and Optical Data from Sentinel-1 and Sentinel-2.

- Remote Sens. 11(13), 1520.
- Breiman, L., 2001. Random Forests. Machine Learning 45(1), 5-32.
- Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. Eur. J. Soil Sci. 62(3), 394-407.
- Byrd, K.B., Ballanti, L., Thomas, N., Nguyen, D., Holmquist, J.R., Simard, M., Windham-Myers, L., 2018. A remote sensing-based model of tidal marsh aboveground carbon stocks for the conterminous United States. ISPRS Journal of Photogrammetry and Remote Sensing 139, 255-271.
- Cai, W., Zhao, S., Zhang, Z., Peng, F., Xu, J., 2018. Comparison of Different Crop Residue Indices for Estimating Crop Residue Cover Using Field Observation Data, 2018 7th International Conference on Agro-geoinformatics (Agro-geoinformatics), pp. 1-4.
- Castaldi, F., Chabrillat, S., Don, A., van Wesemael, B., 2019a. Soil Organic Carbon Mapping Using LUCAS Topsoil Database and Sentinel-2 Data: An Approach to Reduce Soil Moisture and Crop Residue Effects. Remote Sens. 11(18), 2121.
- Castaldi, F., Hueni, A., Chabrillat, S., Ward, K., Buttafuoco, G., Bomans, B., Vreys, K., Brell, M., van Wesemael, B., 2019b. Evaluating the capability of the Sentinel 2 data for soil organic carbon prediction in croplands. ISPRS Journal of Photogrammetry and Remote Sensing 147, 267-282.
- Ceddia, M.B., Gomes, A.S., Vasques, G.M., Pinheiro, É.F.M., 2017. Soil Carbon Stock and Particle Size Fractions in the Central Amazon Predicted from Remotely Sensed Relief, Multispectral and Radar Data. Remote Sens. 9(2), 124.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine Learning 20(3), 273-297.
- Danielson, J.J., Gesch, D.B., 2011. Global multi-resolution terrain elevation data 2010 (GMTED2010).

2331-1258, US Geological Survey.

de Brogniez, D., Ballabio, C., Stevens, A., Jones, R.J.A., Montanarella, L., van Wesemael, B., 2015. A

map of the topsoil organic carbon content of Europe generated by a generalized additive model. *Eur. J. Soil Sci.* 66(1), 121-134.

de Luis, M., Čufar, K., Saz, M.A., Longares, L.A., Ceglar, A., Kajfež-Bogataj, L., 2014. Trends in seasonal precipitation and temperature in Slovenia during 1951–2007. *Reg. Envir. Chang.* 14(5), 1801-1810.

Deng, H., Diao, Y., Wu, W., Zhang, J., Ma, M., Zhong, X., 2019. A high-speed D-CART online fault diagnosis algorithm for rotor systems. *Applied Intelligence*.

Dolšak, D., Bezak, N., Šraj, M., 2016. Temporal characteristics of rainfall events under three climate types in Slovenia. *Journal of Hydrology* 541, 1395-1405.

Dong, W., Wu, T., Luo, J., Sun, Y., Xia, L., 2019. Land parcel-based digital soil mapping of soil nutrient properties in an alluvial-diluvia plain agricultural area in China. *Geoderma* 340, 234-248.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77(4), 802-813.

Forkuor, G., Hounkpatin, O.K.L., Welp, G., Thiel, M., 2017. High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models. *PLOS ONE* 12(1), e0170478.

Gerrard, A.J., 1981. *Soils and landforms. An integration of geomorphology and pedology.* George Allen & Unwin (Publishers) Ltd.

Gholizadeh, A., Žižala, D., Saberioon, M., Borůvka, L., 2018. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sens Environ*

218, 89-103.

- Gomez, C., Dharumarajan, S., Féret, J.-B., Lagacherie, P., Ruiz, L., Sekhar, M., 2019. Use of Sentinel-2 Time-Series Images for Classification and Uncertainty Analysis of Inherent Biophysical Property: Case of Soil Texture Mapping. *Remote Sens.* 11(5), 565.
- Gorrab, A., Zribi, M., Baghdadi, N., Mougenot, B., Fanise, P., Chabaane, Z.L., 2015. Retrieval of Both Soil Moisture and Texture Using TerraSAR-X Images. *Remote Sens.* 7(8), 10098-10116.
- Gosar, M., Šajn, R., Teršič, T., 2016. Distribution pattern of mercury in the Slovenian soil: Geochemical mapping based on multiple geochemical datasets. *Journal of Geochemical Exploration* 167, 38-48.
- Griffiths, R.P., Madritch, M.D., Swanson, A.K., 2009. The effects of topography on forest soil characteristics in the Oregon Cascade Mountains (USA): Implications for the effects of climate change on soil properties. *Forest Ecology and Management* 257(1), 1-7.
- Grimm, R., Behrens, T., Märker, M., Elsenbeer, H., 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island — Digital soil mapping using Random Forests analysis. *Geoderma* 146(1), 102-113.
- Guo, P.-T., Li, M.-F., Luo, W., Tang, Q.-F., Liu, Z.-W., Lin, Z.-M., 2015. Digital mapping of soil organic matter for rubber plantation at regional scale: An application of random forest plus residuals kriging approach. *Geoderma* 237-238, 49-59.
- Hamze-Ziabari, S.M., Bakhshpoori, T., 2018. Improving the prediction of ground motion parameters based on an efficient bagging ensemble model of M5' and CART algorithms. *Applied Soft Computing* 68, 147-161.
- Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A.,

- Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE* 12(2), e0169748.
- Heung, B., Ho, H.C., Zhang, J., Knudby, A., Bulmer, C.E., Schmidt, M.G., 2016. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma* 265, 62-77.
- Hinge, G., Surampalli, R.Y., Goyal, M.K., 2018. Prediction of soil organic carbon stock using digital mapping approach in humid India. *Environ Earth Sci* 77(5), 172.
- Hosseini, M., McNairn, H., 2017. Using multi-polarization C- and L-band synthetic aperture radar to estimate biomass and soil moisture of wheat fields. *International Journal of Applied Earth Observation and Geoinformation* 58, 50-64.
- Hu, Z., Peng, J., Hou, Y., Shan, J., 2017. Evaluation of Recently Released Open Global Digital Elevation Models of Hubei, China. *Remote Sens.* 9(3), 262.
- Huete, A.R., Liu, H.Q., Batchily, K., van Leeuwen, W., 1997. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sens Environ* 59(3), 440-451.
- Jebur, M.N., Pradhan, B., Tehrany, M.S., 2015. Manifestation of LiDAR-Derived Parameters in the Spatial Prediction of Landslides Using Novel Ensemble Evidential Belief Functions and Support Vector Machine Models in GIS. *Ieee J-Stars* 8(2), 674-690.
- Jeong, G., Oeverdieck, H., Park, S.J., Huwe, B., Ließ, M., 2017. Spatial soil nutrients prediction using three supervised learning methods for assessment of land potentials in complex terrain. *CATENA* 154, 73-84.

- Kalambukattu, J.G., Kumar, S., Arya Raj, R., 2018. Digital soil mapping in a Himalayan watershed using remote sensing and terrain parameters employing artificial neural network model. *Environ Earth Sci* 77(5), 203.
- Karatzoglou, A., Smola, A., Hornik, K., Karatzoglou, M.A., 2018. Package 'kernlab', Technical report, CRAN, 03 2016.
- Kasischke, E.S., Melack, J.M., Craig Dobson, M., 1997. The use of imaging radars for ecological applications—A review. *Remote Sens Environ* 59(2), 141-156.
- Kempen, B., Dalsgaard, S., Kaaya, A.K., Chamuya, N., Ruipérez-González, M., Pekkarinen, A., Walsh, M.G., 2019. Mapping topsoil organic carbon concentrations and stocks for Tanzania. *Geoderma* 337, 164-180.
- Keshavarzi, A., Sarmadian, F., Shiri, J., Iqbal, M., Tirado-Corbalá, R., Omran, E.-S.E., 2017. Application of ANFIS-based subtractive clustering algorithm in soil Cation Exchange Capacity estimation using soil and remotely sensed data. *Measurement* 95, 173-180.
- Kim, J., Grunwald, S., 2016. Assessment of Carbon Stocks in the Topsoil Using Random Forest and Remote Sensing Images. *Journal of Environmental Quality* 45, 1910-1918.
- Kuhn, M., 2015. Caret: classification and regression training. *Astrophysics Source Code Library*.
- Lal, R., 2008. Sequestration of atmospheric CO₂ in global carbon pools. *Energy & Environmental Science* 1(1), 86-100.
- Lamichhane, S., Kumar, L., Wilson, B., 2019. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. *Geoderma* 352, 395-413.
- Lausch, A., Baade, J., Bannehr, L., Borg, E., Bumberger, J., Chabrilliat, S., Dietrich, P., Gerighausen, H., Glässer, C., Hacker, J.M., Haase, D., Jagdhuber, T., Jany, S., Jung, A., Karnieli, A., Kraemer, R.,

- Makki, M., Mielke, C., Möller, M., Mollenhauer, H., Montzka, C., Pause, M., Rogass, C., Rozenstein, O., Schmulius, C., Schrod, F., Schrön, M., Schulz, K., Schütze, C., Schweitzer, C., Selsam, P., Skidmore, A.K., Spengler, D., Thiel, C., Truckenbrodt, S.C., Vohland, M., Wagner, R., Weber, U., Werban, U., Wollschläger, U., Zacharias, S., Schaepman, M.E., 2019. Linking Remote Sensing and Geodiversity and Their Traits Relevant to Biodiversity—Part I: Soil Characteristics. *Remote Sens.* 11(20), 2356.
- Lausch, A., Bannehr, L., Beckmann, M., Boehm, C., Feilhauer, H., Hacker, J.M., Heurich, M., Jung, A., Klenke, R., Neumann, C., Pause, M., Rocchini, D., Schaepman, M.E., Schmidtlein, S., Schulz, K., Selsam, P., Settele, J., Skidmore, A.K., Cord, A.F., 2016. Linking Earth Observation and taxonomic, structural and functional biodiversity: Local to ecosystem perspectives. *Ecological Indicators* 70, 317-339.
- Lawrence, R., Bunn, A., Powell, S., Zambon, M., 2004. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote Sens Environ* 90(3), 331-336.
- Lee, B.K., Lessler, J., Stuart, E.A., 2010. Improving propensity score weighting using machine learning. *Stat. Med.* 29(3), 337-346.
- Lee, J.-S., 1986. Speckle suppression and analysis for synthetic aperture radar images. *Opt Eng* 25(5), 255636.
- Li, A., Tan, X., Wu, W., Liu, H., Zhu, J., 2017. Predicting active-layer soil thickness using topographic variables at a small watershed scale. *PLOS ONE* 12(9), e0183742.
- Li, Q.-q., Yue, T.-x., Wang, C.-q., Zhang, W.-j., Yu, Y., Li, B., Yang, J., Bai, G.-c., 2013. Spatially distributed modeling of soil organic matter across China: An application of artificial neural network

- approach. CATENA 104, 210-218.
- Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R news 2(3), 18-22.
- Lombardo, L., Saia, S., Schillaci, C., Mai, P.M., Huser, R., 2018. Modeling soil organic carbon with Quantile Regression: Dissecting predictors' effects on carbon stocks. Geoderma 318, 148-159.
- Lozano-García, B., Parras-Alcántara, L., Brevik, E.C., 2016. Impact of topographic aspect and vegetation (native and reforested areas) on soil organic carbon and nitrogen budgets in Mediterranean natural areas. Science of The Total Environment 544, 963-970.
- Ma, Y., Minasny, B., Wu, C., 2017. Mapping key soil properties to support agricultural production in Eastern China. Geoderma Regional 10, 144-153.
- Mahmoudabadi, E., Karimi, A., Haghnia, G.H., Sepehr, A., 2017. Digital soil mapping using remote sensing indices, terrain attributes, and vegetation features in the rangelands of northeastern Iran. Environ Monit Assess 189(10), 500.
- Marsett, R.C., Qi, J., Heilman, P., Biedenbender, S.H., Carolyn Watson, M., Amer, S., Weltz, M., Goodrich, D., Marsett, R., 2006. Remote Sensing for Grassland Management in the Arid Southwest. Rangeland Ecology & Management 59(5), 530-540.
- Maynard, J.J., Levi, M.R., 2017. Hyper-temporal remote sensing for digital soil mapping: Characterizing soil-vegetation response to climatic variability. Geoderma 285, 94-109.
- McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117(1), 3-52.
- Moore, I.D., Burch, G.J., 1986. Physical Basis of the Length-slope Factor in the Universal Soil Loss Equation1. Soil Science Society of America Journal 50, 1294-1298.
- Morellos, A., Pantazi, X.-E., Moshou, D., Alexandridis, T., Whetton, R., Tziotziou, G., Wiebensohn, J., Bill,

- R., Mouazen, A.M., 2016. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. *Biosystems Engineering* 152, 104-116.
- Mosleh, Z., Salehi, M.H., Jafari, A., Borujeni, I.E., Mehnatkesh, A., 2016. The effectiveness of digital soil mapping to predict soil properties over low-relief areas. *Environ Monit Assess* 188(3), 195.
- Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., Schaepman, M.E., Papritz, A., 2018. Evaluation of digital soil mapping approaches with large sets of environmental covariates. *SOIL* 4(1), 1-22.
- Obu, J., Lantuit, H., Myers-Smith, I., Heim, B., Wolter, J., Fritz, M., 2017. Effect of Terrain Characteristics on Soil Organic Carbon and Total Nitrogen Stocks in Soils of Herschel Island, Western Canadian Arctic. *Permafrost and Periglacial Processes* 28(1), 92-107.
- Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., Fernández-Ugalde, O., 2018. LUCAS Soil, the largest expandable soil dataset for Europe: a review. *Eur. J. Soil Sci.* 69(1), 140-153.
- Ottoy, S., De Vos, B., Sindayihebura, A., Hermy, M., Van Orshoven, J., 2017. Assessing soil organic carbon stocks under current and potential forest cover using digital soil mapping and spatial generalisation. *Ecological Indicators* 77, 139-150.
- Page, K.L., Dalal, R.C., Pringle, M.J., Bell, M., Dang, Y.P., Radford, B., Bailey, K., 2013. Organic carbon stocks in cropping soils of Queensland, Australia, as affected by tillage management, climate, and soil characteristics. *Soil Research* 51(8), 596-607.
- Panagos, P., Ballabio, C., Yigini, Y., Dunbar, M.B., 2013. Estimating the soil organic carbon content for European NUTS2 regions based on LUCAS data collection. *Science of The Total Environment* 442, 235-246.

- Panagos, P., Van Liedekerke, M., Jones, A., Montanarella, L., 2012. European Soil Data Centre: Response to European policy support and public data requirements. *Land Use Policy* 29(2), 329-338.
- Patel, A., Katiyar, S.K., Prasad, V., 2016. Performances evaluation of different open source DEM using Differential Global Positioning System (DGPS). *The Egyptian Journal of Remote Sensing and Space Science* 19(1), 7-16.
- Peng, J., Loew, A., Merlin, O., Verhoest, N.E.C., 2017. A review of spatial downscaling of satellite remotely sensed soil moisture. *Reviews of Geophysics* 55(2), 341-366.
- Poggio, L., Gimona, A., 2017. Assimilation of optical and radar remote sensing data in 3D mapping of soil properties over large areas. *Science of The Total Environment* 579, 1094-1110.
- Pradhan, B., 2013. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Comput. Geosci.* 51, 350-365.
- R Development Core Team, R., 2011. R: A language and environment for statistical computing. R foundation for statistical computing Vienna, Austria.
- Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., Thompson, J., 2018. Soil Property and Class Maps of the Conterminous United States at 100-Meter Spatial Resolution. *Soil Science Society of America Journal* 82, 186-201.
- Ramifehiarivo, N., Brossard, M., Grinand, C., Andriamananjara, A., Razafimbelo, T., Rasolohery, A., Razafimahatratra, H., Seyler, F., Ranaivoson, N., Rabenarivo, M., Albrecht, A., Razafindrabe, F., Razakamanarivo, H., 2017. Mapping soil organic carbon on a national scale: Towards an improved and updated map of Madagascar. *Geoderma Regional* 9, 29-38.

- Ranaie, M., Soffianian, A., Pourmanafi, S., Mirghaffari, N., Tarkesh, M., 2018. Evaluating the statistical performance of less applied algorithms in classification of worldview-3 imagery data in an urbanized landscape. *Advances in Space Research* 61(6), 1558-1572.
- Reyes Rojas, L.A., Adhikari, K., Ventura, S.J., 2018. Projecting Soil Organic Carbon Distribution in Central Chile under Future Climate Scenarios. *Journal of Environmental Quality* 47, 735-745.
- Roelofsen, H.D., van Bodegom, P.M., Kooistra, L., van Amerongen, J.J., Witte, J.-P.M., 2015. An evaluation of remote sensing derived soil pH and average spring groundwater table for ecological assessments. *International Journal of Applied Earth Observation and Geoinformation* 43, 149-159.
- Sadeghi, M., Babaeian, E., Tuller, M., Jones, S.B., 2017. The optical trapezoid model: A novel approach to remote sensing of soil moisture applied to Sentinel-2 and Landsat-8 observations. *Remote Sens Environ* 198, 52-68.
- Shen, Q., Wang, Y., Wang, X., Liu, X., Zhang, X., Zhang, S., 2019. Comparing interpolation methods to predict soil total phosphorus in the Mollisol area of Northeast China. *CATENA* 174, 59-72.
- Siewert, M.B., 2018. High-resolution digital mapping of soil organic carbon in permafrost terrain using machine learning: a case study in a sub-Arctic peatland environment. *Biogeosciences* 15(6), 1663-1682.
- Silva, V., Mol, H.G.J., Zomer, P., Tienstra, M., Ritsema, C.J., Geissen, V., 2019. Pesticide residues in European agricultural soils – A hidden reality unfolded. *Science of The Total Environment* 653, 1532-1545.
- Song, X.-D., Brus, D.J., Liu, F., Li, D.-C., Zhao, Y.-G., Yang, J.-L., Zhang, G.-L., 2016. Mapping soil organic carbon content by geographically weighted regression: A case study in the Heihe River Basin,

- China. *Geoderma* 261, 11-22.
- Song, X.-D., Yang, F., Ju, B., Li, D.-C., Zhao, Y.-G., Yang, J.-L., Zhang, G.-L., 2018. The influence of the conversion of grassland to cropland on changes in soil organic carbon and total nitrogen stocks in the Songnen Plain of Northeast China. *CATENA* 171, 588-601.
- Sreenivas, K., Dadhwal, V.K., Kumar, S., Harsha, G.S., Mitran, T., Sujatha, G., Suresh, G.J.R., Fyzee, M.A., Ravisankar, T., 2016. Digital mapping of soil organic and inorganic carbon status in India. *Geoderma* 269, 160-173.
- Stevens, A., Nocita, M., Tóth, G., Montanarella, L., van Wesemael, B., 2013. Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. *PLOS ONE* 8(6), e66409.
- Stratoulas, D., Balzter, H., Sykioti, O., Zlinszky, A., Tóth, V.R., 2015. Evaluating Sentinel-2 for Lakeshore Habitat Mapping Based on Airborne Hyperspectral Data. *Sensors* 15(9), 22956-22969.
- Sumfleth, K., Duttman, R., 2008. Prediction of soil property distribution in paddy soil landscapes using terrain data and satellite information as indicators. *Ecological Indicators* 8(5), 485-501.
- Suominen, L., Ruokolainen, K., Tuomisto, H., Llerena, N., Higgins, M.A., 2013. Predicting soil properties from floristic composition in western Amazonian rain forests: performance of k-nearest neighbour estimation and weighted averaging calibration. *Journal of Applied Ecology* 50(6), 1441-1449.
- Taghizadeh-Mehrjardi, R., Nabiollahi, K., Kerry, R., 2016. Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. *Geoderma* 266, 98-110.
- Taghizadeh-Mehrjardi, R., Neupane, R., Sood, K., Kumar, S., 2017. Artificial bee colony feature

- selection algorithm combined with machine learning algorithms to predict vertical and lateral distribution of soil organic matter in South Dakota, USA. *Carbon Management* 8(3), 277-291.
- Tóth, G., Jones, A., Montanarella, L., 2013. The LUCAS topsoil database and derived information on the regional variability of cropland topsoil properties in the European Union. *Environ Monit Assess* 185(9), 7409-7425.
- Tsui, C.-C., Chen, Z.-S., Hsieh, C.-F., 2004. Relationships between soil properties and slope position in a lowland rain forest of southern Taiwan. *Geoderma* 123(1), 131-142.
- Tsui, C.-C., Tsai, C.-C., Chen, Z.-S., 2013. Soil organic carbon stocks in relation to elevation gradients in volcanic ash soils of Taiwan. *Geoderma* 209-210, 119-127.
- Tziachris, P., Aschonitis, V., Chatzistathis, T., Papadopoulou, M., 2019. Assessment of spatial hybrid methods for predicting soil organic matter using DEM derivatives and soil parameters. *CATENA* 174, 206-216.
- Vaudour, E., Bel, L., Gilliot, J.M., Coquet, Y., Hadjar, D., Cambier, P., Michelin, J., Houot, S., 2013. Potential of SPOT Multispectral Satellite Images for Mapping Topsoil Organic Carbon Content over Peri-Urban Croplands. *Soil Science Society of America Journal* 77, 2122-2139.
- Vaudour, E., Gomez, C., Fouad, Y., Lagacherie, P., 2019a. Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems. *Remote Sens Environ* 223, 21-33.
- Vaudour, E., Gomez, C., Loiseau, T., Baghdadi, N., Loubet, B., Arrouays, D., Ali, L., Lagacherie, P., 2019b. The Impact of Acquisition Date on the Prediction Performance of Topsoil Organic Carbon from Sentinel-2 for Croplands. *Remote Sens.* 11(18), 2143.
- Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J.-F., Ceschia, E., 2017.

- Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. *Remote Sens Environ* 199, 415-426.
- Vrščaj, B., Repe, B., Simončič, P., 2017. *The soils of Slovenia*. Springer.
- Wadoux, A.M.J.C., 2019. Using deep learning for multivariate mapping of soil with quantified uncertainty. *Geoderma* 351, 59-70.
- Wang, B., Waters, C., Orgill, S., Gray, J., Cowie, A., Clark, A., Liu, D.L., 2018a. High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern Australia. *Science of The Total Environment* 630, 367-378.
- Wang, J., Ding, J., Yu, D., Ma, X., Zhang, Z., Ge, X., Teng, D., Li, X., Liang, J., Lizaga, I., Chen, X., Yuan, L., Guo, Y., 2019a. Capability of Sentinel-2 MSI data for monitoring and mapping of soil salinity in dry and wet seasons in the Ebinur Lake region, Xinjiang, China. *Geoderma* 353, 172-187.
- Wang, J., Xiao, X., Bajgain, R., Starks, P., Steiner, J., Doughty, R.B., Chang, Q., 2019b. Estimating leaf area index and aboveground biomass of grazing pastures using Sentinel-1, Sentinel-2 and Landsat images. *ISPRS Journal of Photogrammetry and Remote Sensing* 154, 189-201.
- Wang, S., Adhikari, K., Wang, Q., Jin, X., Li, H., 2018b. Role of environmental variables in the spatial distribution of soil carbon (C), nitrogen (N), and C:N ratio from the northeastern coastal agroecosystems in China. *Ecological Indicators* 84, 263-272.
- Wang, S., Jin, X., Adhikari, K., Li, W., Yu, M., Bian, Z., Wang, Q., 2018c. Mapping total soil nitrogen from a site in northeastern China. *CATENA* 166, 134-146.
- Wang, S., Zhuang, Q., Wang, Q., Jin, X., Han, C., 2017. Mapping stocks of soil organic carbon and soil total nitrogen in Liaoning Province of China. *Geoderma* 305, 250-263.
- Wang, Y., Wang, S., Adhikari, K., Wang, Q., Sui, Y., Xin, G., 2019c. Effect of cultivation history on soil

- organic carbon status of arable land in northeastern China. *Geoderma* 342, 55-64.
- Ward, K.J., Chabrillat, S., Neumann, C., Foerster, S., 2019. A remote sensing adapted approach for soil organic carbon prediction based on the spectrally clustered LUCAS soil database. *Geoderma* 353, 297-307.
- Were, K., Bui, D.T., Dick, O.B., Singh, B.R., 2015. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. *Ecological Indicators* 52, 394-403.
- Wiesmeier, M., Barthold, F., Blank, B., Kogel-Knabner, I., 2011. Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. *Plant and Soil* 340(1-2), 7-24.
- Wiesmeier, M., Barthold, F., Spörlein, P., Geuß, U., Hangen, E., Reischl, A., Schilling, B., Angst, G., von Lützw, M., Kögel-Knabner, I., 2014. Estimation of total organic carbon storage and its driving factors in soils of Bavaria (southeast Germany). *Geoderma Regional* 1, 67-78.
- Wrb, I.W.G., 2015. World reference base for soil resources 2014, update 2015: International soil classification system for naming soils and creating legends for soil maps. Fao Rome, pp. 192.
- Wu, W., Li, A.-D., He, X.-H., Ma, R., Liu, H.-B., Lv, J.-K., 2018. A comparison of support vector machines, artificial neural network and classification tree for identifying soil texture classes in southwest China. *Computers and Electronics in Agriculture* 144, 86-93.
- Xu, Y., Smith, S.E., Grunwald, S., Abd-Elrahman, A., Wani, S.P., 2018. Effects of image pansharpening on soil total nitrogen prediction models in South India. *Geoderma* 320, 52-66.
- Yang, R.-M., Guo, W.-W., 2019a. Modelling of soil organic carbon and bulk density in invaded coastal wetlands using Sentinel-1 imagery. *International Journal of Applied Earth Observation and*

- Geoinformation 82, 101906.
- Yang, R.-M., Guo, W.-W., 2019b. Using time-series Sentinel-1 data for soil prediction on invaded coastal wetlands. *Environ Monit Assess* 191(7), 462.
- Yang, R.-M., Guo, W.-W., Zheng, J.-B., 2019. Soil prediction for coastal wetlands following *Spartina alterniflora* invasion using Sentinel-1 imagery and structural equation modeling. *CATENA* 173, 465-470.
- Yang, R.-M., Zhang, G.-L., Liu, F., Lu, Y.-Y., Yang, F., Yang, F., Yang, M., Zhao, Y.-G., Li, D.-C., 2016. Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. *Ecological Indicators* 60, 870-878.
- Yang, R., Rossiter, D.G., Liu, F., Lu, Y., Yang, F., Yang, F., Zhao, Y., Li, D., Zhang, G., 2015. Predictive Mapping of Topsoil Organic Carbon in an Alpine Environment Aided by Landsat TM. *PLOS ONE* 10(10), e0139042.
- Yigini, Y., Panagos, P., 2016. Assessment of soil organic carbon stocks under future climate and land cover changes in Europe. *Science of The Total Environment* 557-558, 838-850.
- Zeng, C.Y., Zhu, A.X., Qi, F., Liu, J.Z., Yang, L., Liu, F., Li, F.L., 2019. Construction of land surface dynamic feedbacks for digital soil mapping with fusion of multisource remote sensing data. *Eur. J. Soil Sci.* 70(1), 174-184.
- Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., Finke, P., 2019. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. *Geoderma* 338, 445-452.
- Zhang, Y., Sui, B., Shen, H., Ouyang, L., 2019. Mapping stocks of soil total nitrogen using remote sensing data: A comparison of random forest models with different predictors. *Computers and Electronics in Agriculture* 160, 23-30.

Zhou, T., Li, Z., Pan, J., 2018a. Multi-Feature Classification of Multi-Sensor Satellite Imagery Based on Dual-Polarimetric Sentinel-1A, Landsat-8 OLI, and Hyperion Images for Urban Land-Cover Classification. *Sensors* 18(2), 373.

Zhou, T., Zhao, M., Sun, C., Pan, J., 2018b. Exploring the Impact of Seasonality on Urban Land-Cover Mapping Using Multi-Season Sentinel-1A and GF-1 WFV Images in a Subtropical Monsoon-Climate Region. *ISPRS International Journal of Geo-Information* 7(1), 3.

Credit Author Statement

Data curation, Tao Zhou, Yajun Geng and Jie Chen; writing—original draft preparation, Tao Zhou; supervision, Jianjun Pan, Dagmar Haase and Angela Lausch; Methodology, Tao Zhou and Yajun Geng.

Journal Pre-proof

Declaration of interests

- ☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
- ☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

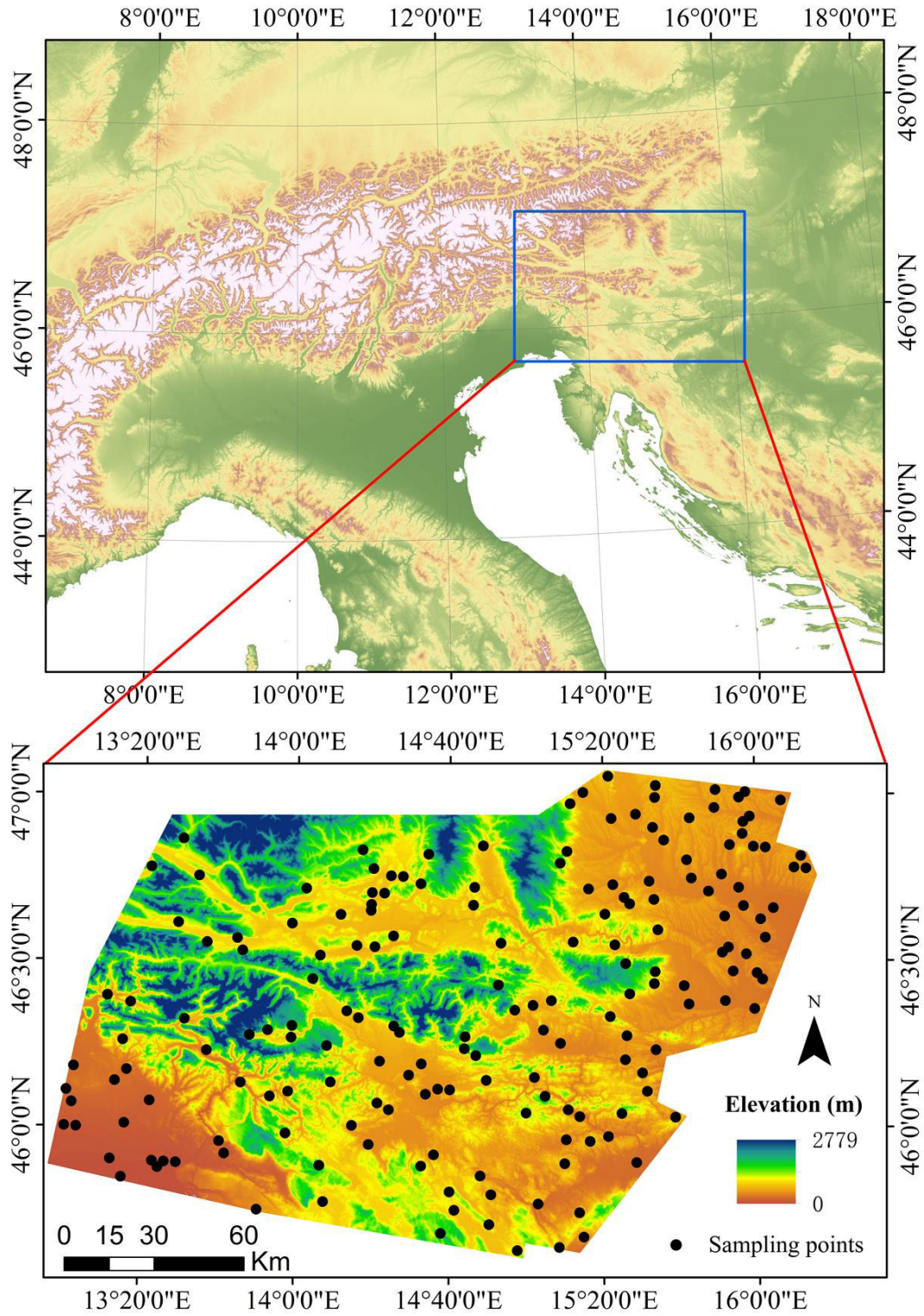
Figure Legends

Fig. 1. The location of the study area and 179 soil samples.

Fig. 2. The relative importance of predictor variables in Model E for predicting SOC and STN. Model E, DEM derivatives + Sentinel-1/2-derived predictors; TWI, topographic wetness index; CS, catchment slope; VD, valley depth; LSF, length-slope factor; VV_1 to VV_5 are the backscatter coefficients of VV polarization of five Sentinel-1 images, respectively (the meaning of the notations 1, 2, 3, 4, and 5 can be found in Table 1); VH_1 to VH_5 are the backscatter coefficients of VH polarization of five Sentinel-1 images, respectively (the meaning of the notations 1, 2, 3, 4, and 5 can be found in Table 1).

Fig. 3. Maps of SOC and STN predicted in Model E using RF, BRT, SVM, and Bagged CART (CART) (Model E: Sentinel-1/2-derived predictors + DEM derivatives).

Fig. 4. Mean SOC and STN maps obtained from Model E based on 100 runs of the BRT model and their corresponding standard deviation maps (Model E: Sentinel-1/2-derived predictors + DEM derivatives).

**Fig. 1**

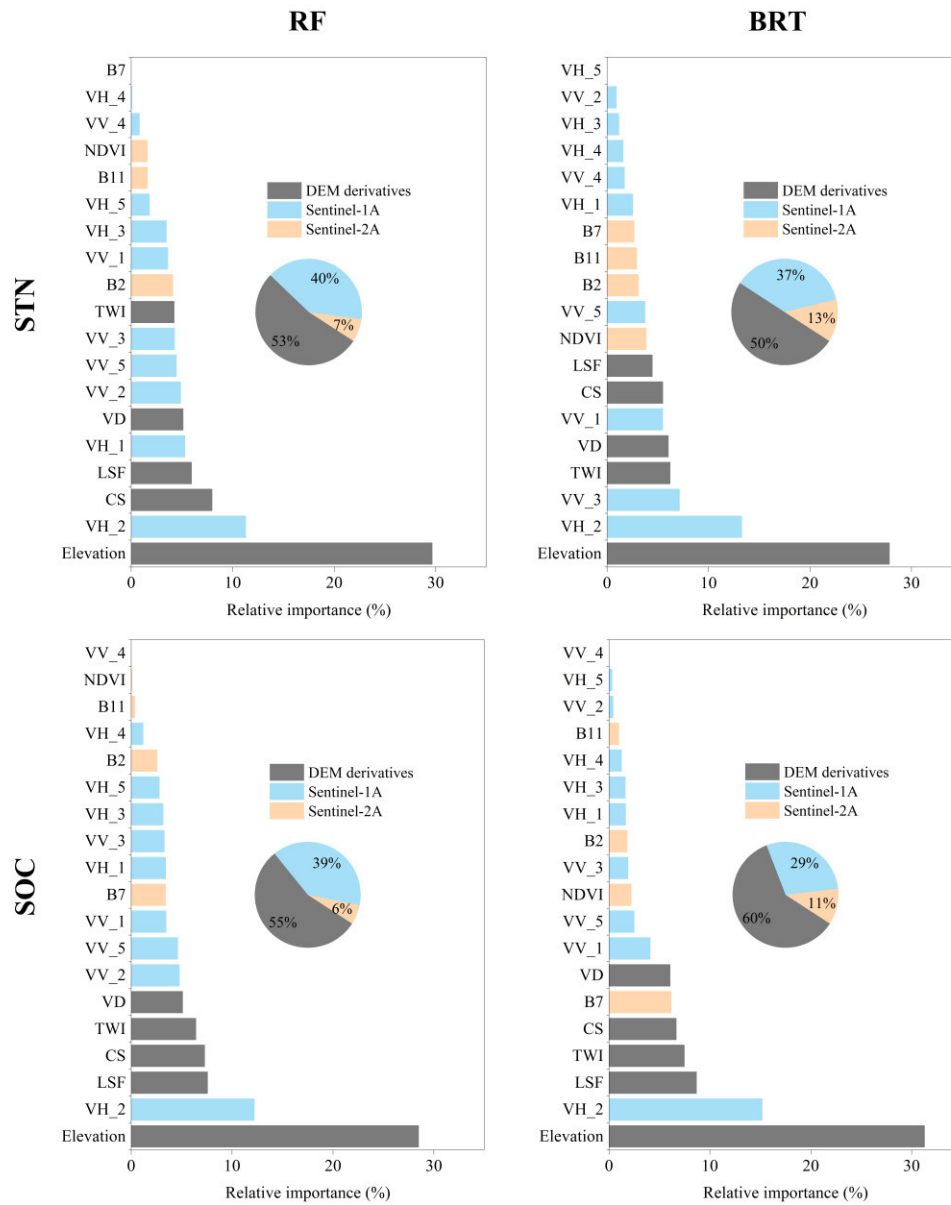


Fig. 2

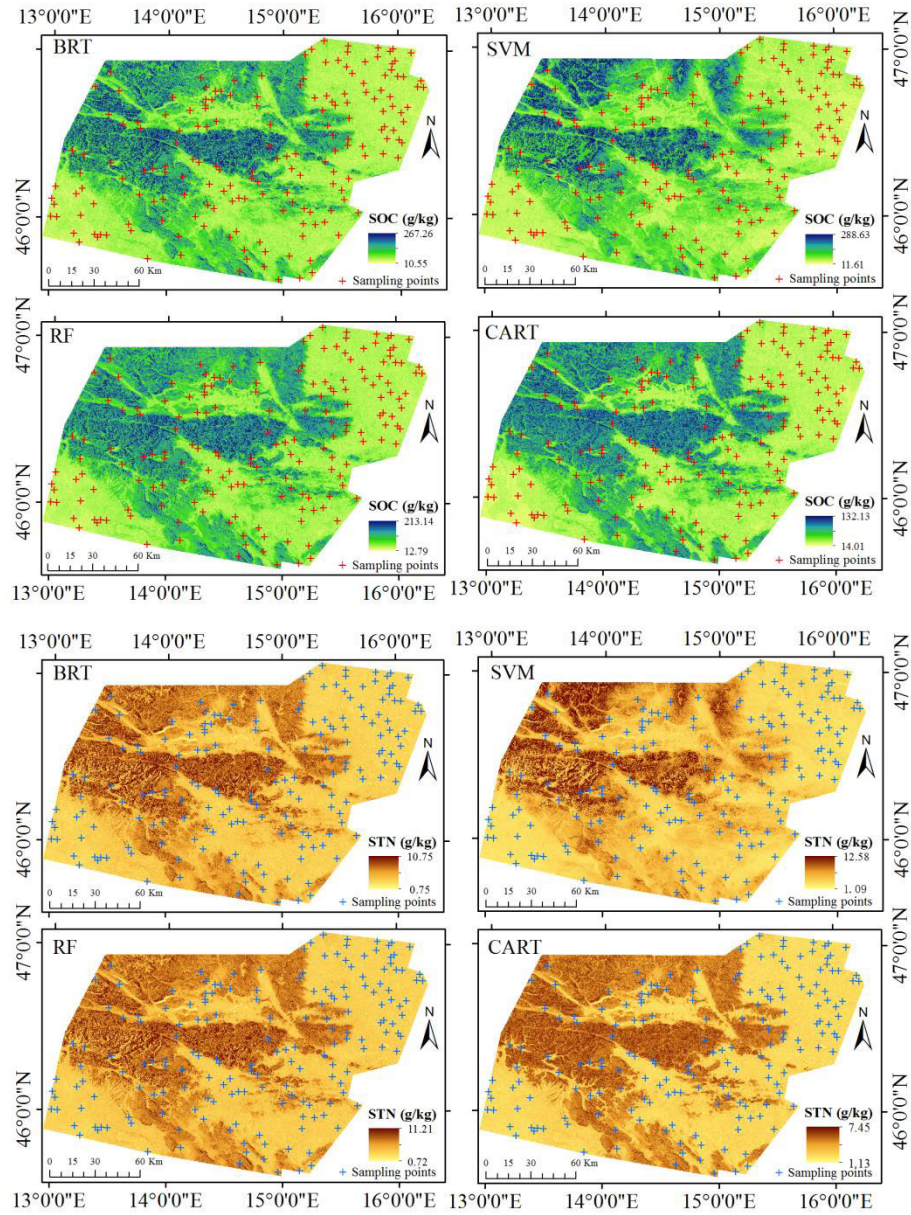


Fig. 3

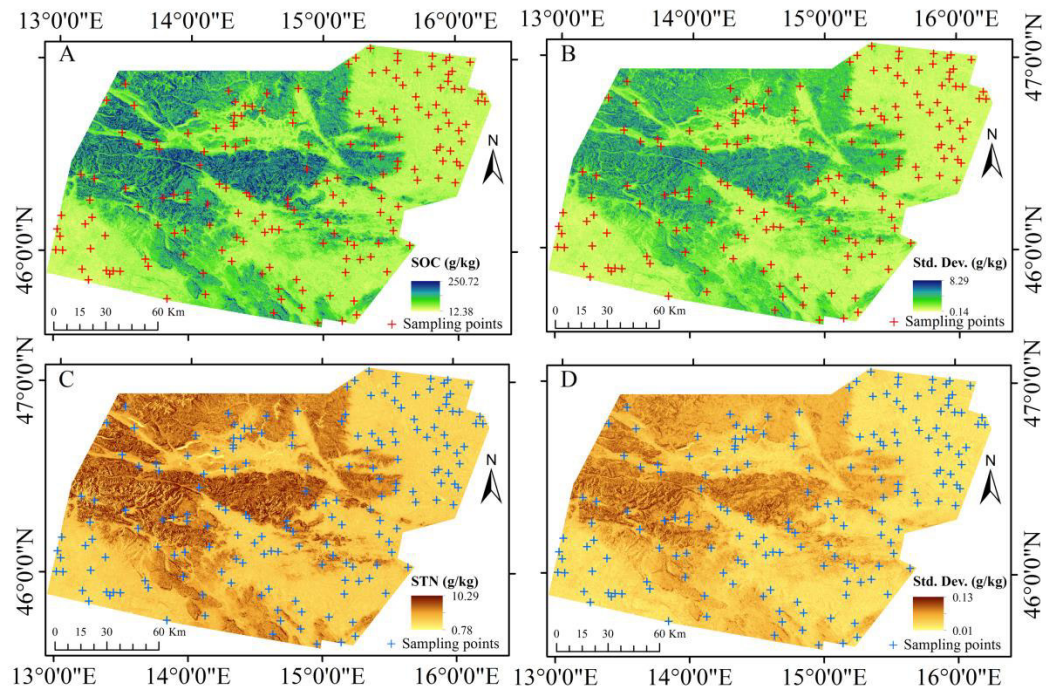


Fig. 4

Journal Pre-proof

Table 1

Sentinel-1A images obtained for predicting SOC content and STN content.

ID	Date	Beam mode	Polarization	Incident Angle (°)	Direction
1	23 rd May 2015	IW	VV/VH	38.90	Descending
2	4 th June 2015	IW	VV/VH	38.96	Descending
3	15 th August 2015	IW	VV/VH	38.96	Descending
4	2 nd October 2015	IW	VV/VH	38.96	Descending
5	26 th October 2015	IW	VV/VH	38.99	Descending

Table 2

Different combinations of DEM derivatives, Sentinel-1 and Sentinel-2 data.

NO.	Model	Environmental variables
i	Model A	Sentinel-2 imagery
ii	Model B	Sentinel-1 imagery
iii	Model C	Sentinel-1 and Sentinel-2 images
iv	Model D	DEM derivatives
v	Model E	Sentinel-1/2-derived predictors and DEM derivatives

Table 3

Descriptive analysis of SOC and STN.

	Minimum	Maximum	Mean	Median	SD	Skewness
SOC (g/kg)	4.70	439.10	49.99	33.40	55.60	3.75
LnSOC (g/kg)	1.54	6.08	3.58	3.50	0.74	0.67
STN (g/kg)	0.30	19.70	3.50	2.70	2.60	2.63
LnSTN (g/kg)	-1.20	2.98	1.05	0.99	0.62	0.12

Notes: LnSOC, log-transformed SOC; LnSTN, log-transformed STN; SD, standard deviation.

Table 4

Prediction accuracy of SOC content and STN content using different combinations of predictors.

The most accurate results are shown in bold.

Modeling technique	Model	SOC			STN		
		MAE	RMSE	R ²	MAE	RMSE	R ²
Bagged CART	Model A	0.53	0.69	0.17	0.45	0.58	0.16
	Model B	0.55	0.70	0.16	0.48	0.59	0.13
	Model C	0.53	0.67	0.22	0.45	0.57	0.18
	Model D	0.48	0.62	0.36	0.41	0.54	0.30
	Model E	0.46	0.59	0.39	0.39	0.51	0.34
RF	Model A	0.52	0.68	0.20	0.45	0.58	0.15
	Model B	0.55	0.70	0.16	0.47	0.58	0.15
	Model C	0.52	0.67	0.25	0.45	0.57	0.18
	Model D	0.48	0.63	0.35	0.41	0.54	0.28
	Model E	0.46	0.59	0.40	0.39	0.51	0.38
BRT	Model A	0.52	0.68	0.19	0.44	0.57	0.16
	Model B	0.53	0.67	0.22	0.46	0.57	0.18
	Model C	0.51	0.65	0.27	0.44	0.56	0.21
	Model D	0.47	0.60	0.39	0.40	0.52	0.32
	Model E	0.45	0.57	0.44	0.39	0.50	0.38
SVM	Model A	0.52	0.68	0.19	0.46	0.58	0.10
	Model B	0.53	0.70	0.16	0.47	0.59	0.11
	Model C	0.51	0.67	0.25	0.46	0.58	0.12
	Model D	0.46	0.60	0.39	0.40	0.52	0.33
	Model E	0.45	0.57	0.43	0.39	0.51	0.35

Notes: Model A, Sentinel-2 imagery alone; Model B, Sentinel-1 imagery alone; Model C,

Sentinel-1 and Sentinel-2 images; Model D, DEM derivatives; Model E, Sentinel-1/2-derived predictors and DEM derivatives.

Journal Pre-proof

Highlights

- Multi-source sensor methods achieved more accurate SOC and STN predictions than single sensors.
- The potential of Sentinel-1 and 2 data in predicting SOC and STN was explored.
- Boosted regression trees model performed best in predicting SOC and STN.