This is the preprint of the contribution published as:

Zhou, T., Geng, Y., Ji, C., Xu, X., Wang, H., Pan, J., **Bumberger, J.**, **Haase, D.**, **Lausch, A.** (2021):
Prediction of soil organic carbon and the C:N ratio on a national scale using machine learning and satellite data: A comparison between Sentinel-2, Sentinel-3 and Landsat-8 images
Sci. Total Environ. **755, Part 2** , art. 142661

The publisher's version is available at:

http://dx.doi.org/10.1016/j.scitotenv.2020.142661

Prediction of soil organic carbon and the C:N ratio on a national scale using machine learning and satellite data: A comparison between Sentinel-2, Sentinel-3 and Landsat-8 images

Tao Zhou[a, b, *], Yajun Geng[c, *], Cheng Ji[d], Xiangrui Xu[c], Hong Wang[e], Jianjun Pan[c], Jan Bumberger[f], 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

[d] Jiangsu Academy of Agricultural Sciences, Institute of Agricultural Resource and Environmental Sciences, Zhongling Street 50, 210014 Nanjing, China

[e] Anhui Science and Technology University, College of Resource and Environment, Donghua Road 9, 233100 Chuzhou, China

[f] Helmholtz Centre for Environmental Research – UFZ, Department Monitoring and Exploration Technology, Permoserstraße 15, 04318 Leipzig, Germany

* Corresponding author.

E-mail address: tao.zhou@ufz.de (T. Zhou), yajungeng@126.com (Y. Geng).

Abstract

Soil organic carbon (SOC) and soil carbon-to-nitrogen ratio (C:N) are the main indicators of soil quality and health and play an important role in maintaining soil quality. Together with Landsat, the improved spatial and temporal resolution Sentinel sensors provide the potential to investigate soil information on various scales. We analyzed and compared the potential of satellite sensors (Landsat-8, Sentinel-2 and Sentinel-3) with various spatial and temporal resolutions to predict SOC content and C:N ratio in Switzerland. Modeling was carried out at four spatial resolutions (800 m, 400 m, 100 m and 20 m) using three machine learning techniques: support vector machine (SVM), boosted regression tree (BRT) and random forest (RF). Soil prediction models were generated in these three machine learners in which 150 soil samples and different combinations of environmental data (topography, climate and satellite imagery) were used as inputs. The prediction results were evaluated by cross-validation. Our results revealed that the model type, modeling resolution and sensor selection greatly influenced outputs. By comparing satellite-based SOC models, the models built by Landsat-8 and Sentinel-2 performed the best and the worst, respectively. C:N ratio prediction models based on Landsat-8 and Sentinel-2 showed better results than Sentinel-3. However, the prediction models built by Sentinel-3 had competitive or better accuracy at coarse resolutions. The BRT models constructed by all available predictors at a resolution of 100 m obtained the best prediction accuracy of SOC content and C:N ratio; their relative improvements (in terms of R^2) compared to models without remote sensing data input were 29.1% and 58.4%, respectively. The results of variable importance revealed that remote sensing variables were the best predictors for our soil prediction

models. The predicted maps indicated that the higher SOC content was mainly distributed in the Alps, while the C:N ratio shared a similar distribution pattern with land use and had higher values in forest areas. This study provides useful indicators for a more effective modeling of soil properties on various scales based on satellite imagery.

Keywords: Soil organic carbon; C:N ratio; Sentinel; Landsat; Machine learning; Digital soil mapping.

1. Introduction

Soil organic carbon (SOC), as one of the main indicators of soil quality and health, is also an important and variable carbon pool in terrestrial ecosystems and thus plays an important role in regulating the global carbon cycle and in maintaining soil quality (Lausch et al., 2019). The ratio of SOC to total nitrogen (C:N ratio) is also an important index of soil quality and fertility, reflecting the interaction or coupling between SOC and total nitrogen (Lou et al., 2012; Xu et al., 2018a). Moreover, the C:N ratio is the main factor affecting soil microbial communities and thus plays a key role in the terrestrial carbon and nitrogen cycle (Wan et al., 2015; Wu, 2020; Xu et al., 2019). Quantifying the spatial distribution of SOC and the C:N ratio is essential for establishing better soil management, ecological environment monitoring and climate policy. Unfortunately, the costs and efficiency associated with ground surveys, soil sampling, and laboratory analysis limit the large-scale monitoring of soil properties (Chen et al., 2019; Xu et al., 2020). Reliable and cost-effective approaches for predicting SOC content and C:N ratio are therefore indispensable.

Digital soil mapping is an effective method to accurately predict soil properties over large areas, while reducing the cost of sampling and analysis (Jeong et al., 2017). Digital soil mapping establishes soil prediction models based on the quantitative relationship between field soil observations and environmental predictors representing soil formation factors to understand the spatial patterns of soil properties (Loiseau et al., 2019; McBratney et al., 2003). Many techniques have been developed to link soil and environmental predictors through the framework of digital soil mapping, where machine learning algorithms have become very popular due to excellent predictive performance (Padarian et al., 2020; Taghizadeh-Mehrjardi et al., 2020). Machine learners commonly used in soil mapping, as listed in the reviews by Heung et al. (2016) and Lamichhane et al. (2019), mainly include: random forest (RF), Cubist, boosted regression tree (BRT) and support vector machine (SVM). However, the results of various comparative studies based on machine learning methods were not consistent (Jeong et al., 2017; Wang et al., 2018a; Were et al., 2015).

With these advances in soil mapping, environmental variables (e.g., satellite imagery, terrain and climate data) obtained from various sources have been combined with field soil observations to predict soil properties (Kalambukattu et al., 2018; Matos-Moreira et al., 2017; Were et al., 2016). Among them, remote sensing images provide a large number of environmental variables with multiple spatial and temporal resolutions for simulating soil-landscape relationships. The prediction of soil properties has been achieved from field to global scales, taking into account the different specific characteristics of remote sensing sensors. For example, MODIS (Cui et al., 2018) satellite products with low spatial but high temporal resolution and a wide-angle field of view have been used by researchers for soil mapping on a global scale (Hengl et al., 2014). At present, numerous regional and national digital soil products have been obtained using Landsat (Bhattarai et al., 2015) sensors with medium spatial and low temporal resolution (Broderick et al., 2015; Ramifehiarivo et al., 2017; Zhi et al., 2018). The selection of appropriate sensors is important for soil mapping because each sensor has its advantages and characteristics (Cui et al., 2018). In general, sensors with a high temporal resolution produce wide-area coverage with lower spatial resolution, while sensors with a high spatial resolution are limited in their spatial coverage and temporal resolution (Zeng et al., 2019). For example, MODIS data with a repeat cycle of about 1-2 days has a coarse spatial resolution of 250 to 1000 m (Lausch et al., 2016; Xie et al., 2008). This sensor was proposed to support soil mapping in areas where data availability is limited (Minasny et al., 2008). The Landsat sensor with a spatial resolution of 30 m has a long return cycle (16 days) (Wulder et al., 2019). Such temporal resolution and the impact of the cloud reduce the availability of Landsat data (Bhattarai et al., 2015). Although these two most commonly used sensors have been widely and successfully applied, the improved spatial and temporal resolution characteristics of recently available free and open access remote sensing images have attracted great interest from scientists (Loiseau et al., 2019; Yang and Guo, 2019).

The recently released Sentinel satellite series developed by the European Space Agency (ESA) provides an unprecedented amount of free data for global environmental safety monitoring (Berger et al., 2012). In particular, Sentinel-2 (S2) and Sentinel-3 (S3) with high revisit frequency (i.e., 5 days and < 2 days respectively) based on two satellites provide near real-time operational products for land monitoring (Verrelst et al., 2012). It is expected that these frequent measurements will greatly improve the ability to detect useful information for various land applications, especially in areas prone to clouds (Clark, 2017). S2 and S3 satellite sensors with different spatial resolutions (i.e., 10–60 m and 300 m, respectively vs. 30 m) have larger swath widths (i.e., 290 km and 1270 km, respectively vs. 185 km) compared to the Landsat-8 (L8) sensor (Berger et al., 2012; Li and Roy, 2017). S2 sensors have now been successfully applied to predict different soil properties such as SOC (Castaldi et al., 2019; Gholizadeh et al., 2018; Vaudour et al., 2019), texture (Bousbih et al., 2019; Gomez et al., 2019), soil total nitrogen (Zhang et al., 2019) and soil salinity (Davis et al., 2019; Taghadosi et al., 2019) in various environments. However, so far the potential of the S3 sensor in predicting SOC content and the C:N ratio has not yet been fully exploited. Generally speaking, the prediction accuracy depends largely on the selected satellite product (Lin et al., 2020). Quantitative evaluation of the performance of soil prediction models based on multi-satellite sensors can help end users choose the most appropriate satellite imagery. Although several sensors have been compared in the literature to predict soil properties, there is no consensus on the potential of recently available satellite sensors. For example, recent studies by Wang et al. (2020a), Davis et al. (2019) and (Xu et al., 2017a) compared the effects of remote sensing images with different temporal and spatial resolutions on soil prediction models. Previous studies, such as Kim et al. (2012), Chi et al. (2019), Samuel-Rosa et al. (2015) and Taylor et al. (2013) observed the advantages or disadvantages of environmental variables with different spatial resolutions (e.g., satellite images, terrain attributes and ecological indicators) in the prediction of soil properties. Most previous studies have only compared the effects of different sensors with medium resolution on soil prediction models. The effects of satellite sensors with a medium to coarse spatial resolution on soil prediction models have rarely been compared and analyzed before, especially Sentinel sensors with a broad application potential. Comparing the performance of soil prediction models based on different sensors will improve our understanding of the capabilities and advantages of these sensors in soil mapping. Therefore, the selection of appropriate satellite sensors in digital soil mapping requires further efforts through the evaluation of different sensors to improve soil mapping.

Some scholars have focused much of their attention on national-scale SOC mapping due to the high demand for national information on soil properties e.g. in Hungary (Szatmári et al., 2019), China (Liang et al., 2019), India (Sreenivas et al., 2016), Brazil (Gomes et al., 2019), Sri Lanka (Vitharana et al., 2019) and France (Martin et al., 2011). In Switzerland, some studies have carried out spatial prediction of SOC based on digital soil mapping technology but most studies have focused on a few small areas or specific land use types (Hoffmann et al., 2014; Nussbaum et al., 2014; Nussbaum et al., 2018). Although some of the existing digital soil products have been produced on a European scale, they do not cover Switzerland (Panagos et al., 2013; Rial et al., 2017; Yigini and Panagos, 2016). Therefore, there is a lack of information on the spatial distribution of SOC and the C:N ratio at the national scale in Switzerland.

The main objective of this study was to analyze and compare the potential of satellite sensors (i.e., L8, S2, and S3 sensors) for predicting SOC content and the C:N ratio in Switzerland using three machine learning techniques. In particular, our study aimed (i) to compare and select the best model to map the spatial distribution of SOC content and the C:N ratio for the whole of Switzerland and (ii) to evaluate the effects of satellite sensors with different temporal and spatial resolutions on the SOC and C:N ratio prediction models with four distinct spatial resolutions. These objectives were achieved by using different combinations of environmental data (topography, climate and satellite imagery) to generate soil prediction models in three machine learners (i.e., BRT, RF, and SVM algorithms). The soil prediction models were constructed with four spatial resolutions (800 m, 400 m, 100 m and 20 m). We compared the accuracy of soil prediction models and also evaluated the spatial pattern of soil properties and the importance of predictors.

2. Materials and methods

2.1. Study area

Switzerland is located in Central Europe and covers an area of 41,000 km^2, ranging from 196 m to 4634 m above sea level (Stumpf et al., 2018) (Fig. 1). It is located in a temperate climate zone with a mean annual temperature (MAT) of 8.6 ° C and a mean annual precipitation (MAP) of 500-2000 mm. The main soil types in the area are Haplic Podzols and Haplic Cambisols (https://soilgrids.org/) (Hengl et al., 2017). Land use in Switzerland is dominated by agricultural and forest areas (Price et al., 2015). Agricultural land accounts for 37% of the total area, mainly including arable land and permanent grassland (Leifeld et al., 2005). Cereals, fruits and vegetables are the main agricultural products.

2.2. Soil dataset

The soil data we used was obtained from the European Soil Data Centre (ESDAC) that included 150 soil samples from Switzerland (Fig. 1) (Fernández-Ugalde et al., 2020; Panagos et al., 2012). Soil sampling (0–20 cm) was conducted in 2015 as part of the European-scale LUCAS 2015 Topsoil Survey. The LUCAS survey has been conducted every three years since 2009, of which the LUCAS 2015 Survey is the latest (Ballabio et al., 2019). The LUCAS sampling density is 14 km × 14 km corresponding to one sample (Panagos et al., 2014). Five sub-samples were collected at each location to prepare a composite sample of approximately 500 g. The air-dried samples were sent to the laboratory for analysis by ISO standard methods. The LUCAS data set recorded the sample locations and corresponding main soil physicochemical properties, including SOC and the C:N ratio used for modeling. More details about sampling strategies and analysis methods are provided by Fernández-Ugalde et al. (2020).

2.3. Environmental data for modeling

Based on soil formation factors, we collected the following types of environmental variables from public sources for modeling analysis: remote sensing images, terrain attributes, and climate data. These environmental variables were converted into raster layers (UTM WGS84 Zone 32N projection system) with spatial resolutions of 20, 100, 400 and 800 m using ArcGIS 10.4 software. For all environmental variables, the attribute values corresponding to each soil sample were extracted as input for the modeling (Chen et al., 2019). The source and processing of environment variables were as follows:

2.3.1. Terrain attributes

EU-DEM v1.1 products covering the study area with a resolution of 25 m were used to extract various terrain attributes. From this DEM data, the terrain variables generated in this study using SAGA GIS software were as follows: elevation, slope, valley depth (VD), SAGA topographic wetness index (TWI), channel network base level (CNBL), vertical distance to channel network (VDCN), catchment slope (CS) and slope length (SL). Details of the calculations for these variables can be referenced here: http://www.saga-gis.org/.

2.3.2. Satellite imagery and processing

The satellite images used in this study included L8 OLI data downloaded from the Earth Explorer website and S2 and S3 images obtained from ESA. L8 and S2 data were mosaiced using 6 and 12 images captured from August to September 2015, respectively (Chen et al., 2014; Li et al., 2019). The S3 data covering the study area was trimmed from the S3 OLCI (Ocean and Land Colour Instrument) full resolution (FR) image on August 23rd, 2016. The cloud cover on all of these remote sensing images was less than 10%. Both S2 and S3 are constellations with two satellites, of which S2A and S3A were launched in 2015 and 2016 respectively. The S2 MultiSpectral Instrument (MSI) and S3 OLCI sensors have 13 (from 443 nm to 2190 nm) and 21 (from 400 nm to 1020 nm) spectral bands, respectively (Kokhanovsky et al., 2019). The former has a swath width of 290 km and a 5-day revisit cycle, while the latter has a wider swath width (1270 km) and shorter revisit time (less than two days) (Clevers and Gitelson, 2013; Defourny et al., 2019). We used ENVI 5.5.3 software for geometric correction of the S3 data. All remote sensing data sets were then atmospherically corrected using the FLAASH atmospheric model, including radiometric calibration and atmospheric correction (Ke et al., 2015; Lin et al., 2020; Misra et al., 2018). Twenty-one S3 bands, nine bands of S2 (B2, B3, B4, B5, B6, B7, B8A, B11 and B12) (Vaudour et al., 2019; Wang et al., 2019), and bands 1 to 7 of L8 OLI were used as candidate predictors for modeling.

2.3.3. Climate data

Climate variables downloaded from Worldclim (Hijmans et al., 2005) included MAP and MAT data with a spatial resolution of 1 km as environmental variables for soil mapping in this study. Worldclim provides interpolated climate data for global land areas and has many applications in ecological modeling. These data were developed using thin-plate splines to interpolate weather station data. The methods used to produce and interpolate Worldclim data are described in detail by Hijmans et al. (2005) and Fick and Hijmans (2017).

2.4. Predictive models

2.4.1. Support vector machine

SVM is a machine learning technique based on the statistical learning theory. The SVM model uses kernel functions to project data into a high-dimensional space where separation is performed (Forkuor et al., 2017). In this study, the radial basis function (RBF) (Eq. (1)) was selected as a kernel function due to its good performance in soil mapping (Keskin et al., 2019).

$$\mathrm{k}(x_i, x_j) = \exp\left(-\sigma\left\|x_i + x_j\right\|^2\right) \quad (1)$$

where k is the user-defined kernel function, x is the input vector, and σ represents the width of the RBF (Jeong et al., 2017).

We used the "kernlab" package of R software to develop the SVM model. In SVM modeling, there are two parameters that need to be adjusted, including kernel width (sigma) and penalty (cost). Using the grid search approach, the best parameters were obtained with the "caret" package in the R software (Forkuor et al., 2017). More specific information about the SVM model is provided by Were et al. (2015).

2.4.2. Random forest

RF is a tree-based method for modeling the relationship between target variables and potential predictors (Rasaei and Bogaert, 2019). The RF model takes decision trees as the basic unit and averages all tree results to obtain its predicted results. A large number of decision trees are constructed in RF to ensure the stability of the model, where each tree is independently planted by a unique bootstrap sample of the training dataset (Khanal et al., 2018). RF estimates error and variable importance by using out-of-bag (OOB) samples, which are samples omitted from the bootstrap samples (Were et al., 2015). The OOB mean square error (MSE_{OOB}) is calculated by aggregating the predictions of all trees (Eq. (2)).

$$MSE_{OBB} = \frac{1}{n}\sum_{i=1}^{n}\left(z_i - \hat{z}\,_i^{OBB}\right)^2 \quad (2)$$

where *n* is the number of observations and $\hat{z}\,_i^{OBB}$ is the OOB prediction for observation z_i.

This modeling technique is generally preferred in soil mapping studies because it can estimate the importance of variables, it is insensitive to overfitting and has stable and accurate predictions (Wiesmeier et al., 2011; Yang et al., 2020). The RF model was implemented through the "randomForest" package in R . The user needs to define two main parameters in RF modeling: the number of input variables (mtry) in each tree and the number of trees (ntree). The grid search method of the "caret" package in R was used to optimize these parameters. The combination of parameters with the lowest prediction error was used for the final modeling.

2.4.3. Boosted regression trees

Developed by Friedman et al. (2000), BRT combines the advantages of two algorithms (i.e., regression trees and boosting) to improve the performance of a single model. Boosting is a numerical optimization algorithm that minimizes the loss function by adding a new tree to the first regression tree model at each step (Arabameri et al., 2019; Elith et al., 2008). We developed BRT models using the "gbm" package in R. Three main parameters need to be optimized to run BRT: the number of trees (NT), the learning rate (LR) and the tree complexity (TC) (Wang et al., 2018a). In a similar way to the SVM and RF models, we optimized these three parameters using the grid search approach through the "caret" package. The optimal combination of NT, TC and LR parameters that provide the minimum predictive deviance was set in the BRT model.

2.5. Statistical analyses

A descriptive statistical analysis of the target soil properties was performed using SPSS 21.0 software. Some environmental variables may not provide information to predict target soil properties and may be redundant or highly correlated. Boruta is an all-relevant variable selection algorithm that can cope with redundancy and collinearity between environmental variables (Xiong et al., 2014; Xu et al., 2020). To extract useful information from a large set of variables and reduce multicollinearity, the Boruta algorithm was used to identify the environmental variables that were relevant for each soil property. After identifying the relevant variables for each soil property, these selected environmental variables were then used for modeling analysis of each soil property. In previous soil mapping studies (Keskin et al., 2019; Xu et al., 2017b), the Boruta algorithm was adopted and reported as an effective method to reduce the multicollinearity of predictors. In this study, the “Boruta” package was used to run the Boruta algorithm.

2.6. Accuracy assessment and uncertainty

To evaluate and compare the capability of freely and globally available multispectral sensors with different temporal and spatial resolutions to predict the C:N ratio and SOC at four spatial resolutions, we used three machine learning techniques to construct the following five experimental models: Model I, Model II, and Model III were constructed from L8, S2, and S3 images, respectively; Model IV was a combination of climate and terrain variables, while Model V included all available predictors (Table 1). Fig. 2 shows an overview of the flowchart for SOC and C:N ratio mapping using these experimental models in Switzerland. Ten-fold cross-validation was used to evaluate the performance of these models. This technique divides the data set into ten equal-sized subsets. After that, one of the subsets is used to evaluate the model, while the other nine subsets are used to train the model. This method is repeated ten times to ensure that each of the ten subsamples evaluates the model once (Amirian-Chakan et al., 2019). The following three evaluation indices were calculated: the coefficient of determination (R^2), the root mean square error (RMSE) and the mean absolute error (MAE) (Eqs. (3)–(5)).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|P_i - O_i\right| \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{4}$$

$$R^2 = \frac{\sum_{i=1}^{n}(P_i - \overline{O}_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O}_i)^2} \tag{5}$$

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

For every soil property, each model was run a hundred times and their average was used as the final prediction. We calculated the standard deviation (SD) of each raster cell based on the 100 soil maps generated and used the spatial variation of these SDs to represent the prediction uncertainty (Hamzehpour et al., 2019; Wang et al., 2020b).

3. Results

3.1. Descriptive statistics of soil properties

Descriptive statistics of soil properties are exhibited in Table 2. The observed SOC content ranged from 8.90 to 151.50 g/kg with an average of 43.93 g/kg and a median of 37.45 g/kg. The average value of the C:N ratio was 10.60 (median: 9.16), ranging from 3.33 to 22.26. The distribution of SOC and C:N ratio data was strongly skewed, with skewness values of 1.57 and 1.06, respectively. Therefore, we applied the natural log transformation to these soil data. After conversion, the skewness values of SOC and the C:N ratio data were reduced to 0.13 and 0.16 respectively. The SD values of the raw SOC and the C:N ratio were 27.65 g/kg and 3.24, respectively, while the corresponding SD values after conversion were 0.58 g/kg and 0.29, respectively.

3.2. Model evaluation and comparison

The performance results of RF, BRT and SVM in predicting SOC and the C:N ratio based on five experimental models at four different spatial resolutions are shown in Table 3. The comparative analysis of model performance obviously demonstrated that the choice of sensors, modeling resolution and model type significantly affected the prediction accuracy of SOC and the C:N ratio. For instance, in terms of the best prediction of SOC obtained by each experimental model at four resolutions, RF showed a higher accuracy than SVM when using Model I and Model V, while the latter performed better in Model II, Model III and Model IV. At the same time, for the best prediction of the C:N ratio using RF and SVM, Model I, Model II, Model III and Model IV were all better predicted by SVM, while RF and SVM had similar prediction quality in Model V. Among the SOC predictions of the five experimental models, BRT had the highest prediction accuracy in Model I and Model V, while the best predictions of the remaining three experimental models were obtained by SVM. For C:N ratio mapping, SVM achieved the best prediction from Model I to Model IV, while the highest accuracy of Model V came from BRT prediction. When evaluating the models that most accurately predicted SOC and the C:N ratio, it was found that although SVM performed best in some experimental models, BRT had the lowest RMSE and MAE values and the highest R^2 values in both SOC and the C:N ratio predictions.

Compared with Model IV that was constructed from climate and terrain variables, the prediction accuracy using only one satellite sensor was competitive. For example, it can be observed that L8-based (R^2 = 0.363 and R^2 = 0.353 for SOC and C:N ratio predictions, respectively) prediction accuracy using BRT was not inferior to Model IV (R^2 = 0.364 and R^2 = 0.255 for SOC and C:N ratio predictions, respectively). Competitive prediction accuracy has also been observed in models related to S2 (R^2 = 0.253 and R^2 = 0.334 for SOC and C:N ratio predictions, respectively) and S3 (R^2 = 0.290 and R^2 = 0.247 for SOC and C:N ratio predictions, respectively) sensors. This result indicates that these three satellite sensors with different characteristics are very important auxiliary variables for the effective modeling of SOC and the C:N ratio.

Among the three satellite-based experimental models with four resolutions, the best performing models in predicting the C:N ratio and SOC were all derived from L8 (Model I), followed by S2 (Model II) and S3 (Model III). Specifically, for SOC prediction at the same resolution, the L8 model performed best, followed by the S3 and S2 models, where the S2 model had the worst overall prediction accuracy. However, the S3 model performed better at a coarser spatial resolution, especially at 800 m. For C:N ratio prediction at the same resolution, the overall performance of the L8 model was the best and the S3 model was the worst, but the latter had a higher accuracy at 800 m. At the same time, the three satellite-based experimental models performed best at different resolutions: the L8 models achieved the best predictions for SOC and the C:N ratio at 100 and 20 m, respectively; S2 models were all implemented at 100 m, while S3 models performed best at 400 or 800 m. The prediction accuracy of the L8 and S2 models both decreased significantly when the resolution moved from 100 m to 800 m. For example, along the resolution from 100 m to 800 m, the R^2 of the L8 model based on BRT in predicting SOC and the C:N ratio dropped from 0.363 to 0.226 and from 0.330 to 0.142, respectively. The results revealed that these sensors have different capabilities to predict soil properties at distinct modeling resolutions. In addition, the prediction accuracy of sensors with a coarser spatial resolution can provide competitive and even higher accuracy of soil properties compared to sensors with a higher spatial resolution. On the other hand, Model V (all available predictors) also produced the highest prediction accuracy at 100 m. Moreover, the prediction accuracy of Model V at 20 m was lower and higher than its accuracy in predicting SOC and the C:N ratio at coarser spatial resolutions (400 and 800 m), respectively. When moving from 100 m to 800 m, the R^2 of using Model V to predict SOC and the C:N ratio with BRT decreased by 16.0% (from 0.470 to 0.395) and 47.5% (from 0.404 to 0.212), respectively.

For all machine learners and modeling resolutions, soil prediction models always displayed higher accuracy when remote sensing variables and other variables (climate and terrain variables) were applied together. Compared with the use of climate and terrain variables (Model IV) alone, the R^2 of the BRT model at 100 m improved by 29.1% (from 0.364 to 0.470) in predicting SOC and by 58.4% (from 0.255 to 0.404) in predicting the C:N ratio due to the addition of remote sensing variables. We were able to observe this improvement for other prediction models and modeling resolutions. These results further indicate that the variables derived from these three sensors contain valuable information that can improve the overall prediction accuracy. Although the prediction performance was very different on all machine learning algorithms and modeling resolutions, the Model V (all available predictors) consistently performed best, with the BRT model achieving the highest accuracy of SOC (R^2 = 0.470, RMSE = 0.437, and MAE = 0.336) and the C:N ratio (R^2 = 0.404, RMSE = 0.223, and MAE = 0.167) predictions at 100 m. The R^2 values suggested that these models could explain about 47% and 40% of the SOC and the C:N ratio variability, respectively.

3.3. Relative importance of environmental variables

The relative importance of the top twenty most important environmental variables used for SOC and the C:N ratio mapping in Model V at 100 m based on BRT and RF is shown in Fig. 3. There were slight differences in the ranking of environmental variables in these two predictive models. For example, the five most important environmental variables in the BRT model were L8_B1, CS, VD, MAP, and S2_B2 when predicting SOC, while the top five variables in the RF model were VD, CS, L8_B1, L8_B7, and S3_B6; L8_B1, CS and VD were all ranked in the top three in both models. For C:N ratio prediction, S2_B3, slope and L8_B1 followed by L8_B3 and TWI were the most important environmental variables in the BRT model, where S2_B3 and slope were also in the top three in the RF model. The results of the BRT model also showed that topography (relative importance of 31%) was the main explanatory variable for SOC prediction, followed by L8 (26%), S3 (17%), S2 (13%) and climate (13%). In contrast, S2, L8, topography, and S3 explained 41%, 26%, 22%, and 11% of the C:N ratio variability in the BRT model, respectively. Moreover, remote sensing imagery (L8, S2 and S3) accounted for 56% and 78% of the relative importance of SOC and the C:N ratio prediction in the BRT model, respectively, which indicates that these remote sensing data have the most important impact in predicting SOC and the C:N ratio in Switzerland.

3.4. Spatial prediction

As shown in Section 3.2, Model V based on BRT at 100 m had the highest prediction accuracy, which was adopted to predict the spatial distribution of SOC content and the C:N ratio in Switzerland. The spatial prediction results are the two maps for SOC content and the C:N ratio, which display the average and SD values of one hundred predictions (Fig. 4). The average SOC content (SD: 18.26 g/kg) and the C:N ratio (SD: 1.98) were 44.60 g/kg and 10.73, respectively. Both the SOC and C:N ratio prediction models based on Model V showed low uncertainty. The average SD from 100 predicted outputs was 1.89 g/kg for SOC content and 0.08 for the C:N ratio. The low SD value of BRT indicated that it was stable in predicting SOC content and the C:N ratio. The spatial details of the soil properties were lost when moving from high resolution to coarse resolution, especially at 800 m (Figs. 4–6).

4. Discussion

4.1. Performance of soil prediction models using different combinations of environmental variables

In this study, comparative analysis revealed that the selection of prediction models, satellite sensors, and modeling resolution significantly affected the accuracy of soil prediction models (Table 3). We found that the BRT model achieved the highest prediction accuracy of SOC content and C:N ratio, although SVM performed better in some experimental models. This is consistent with the results of Wang et al. (2018a) who compared the performance of RF, SVM, and BRT in predicting SOC and reported similar results. These comparable results were also supported by Ottoy et al. (2017). However, opposing results were also observed in previous soil mapping studies (Paul et al., 2020; Wang et al., 2020c), which found that the RF model performed better than BRT. These differences may be caused by the location and the spatial extent of the study area, the density and number of soil samples, and the type and resolution of environmental variables. Similarly, no model has been found in this study to consistently outperform other models in predicting soil properties using different experimental models at four resolutions. Therefore, it is necessary to calibrate and evaluate competitive prediction models based on specific experimental data sets at different spatial resolutions.

Our results demonstrated that three satellite sensors (i.e., L8, S2, and S3 sensors) are essential for effective mapping of SOC and C:N ratio. Various satellite sensors have been successfully applied to digital soil mapping of different scales, among which the most commonly used are Landsat and MODIS sensors. For example, Landsat has been widely used to model soil properties at local (Xu et al., 2017a), regional (Scudiero et al., 2014) and national (Wadoux, 2019) scales. Some scholars have applied MODIS data to conduct soil mapping studies at European (Ballabio et al., 2018; Panagos et al., 2014), African (Hengl et al., 2015; Vågen et al., 2016) and global (Hengl et al., 2017) scales. Previous studies have reported that sensors with a coarser resolution are ideal for capturing the general characteristics of the landscape, while sensors with a higher resolution are suitable for capturing small spatial variations in soil properties (Kim et al., 2012; Schmid et al., 2008). Remote sensing based soil mapping is subject to the availability and quality of the remote sensing imagery (Li et al., 2014). Although Landsat has a higher spatial resolution, its lower overpass increases the difficulty of selecting cloudless scenes (Poggio and Gimona, 2017). The S3 sensor has a coarser spatial resolution, but its higher overpass can easily meet the needs of soil mapping applications for remote sensing data, especially in areas susceptible to cloud cover and rain. However, so far, the application of S3 products in soil mapping has been limited and it's potential for SOC and C:N ratio prediction has not yet been fully developed. In this study, the competitive accuracy obtained by the S3 sensor demonstrates the feasibility of globally available S3 data in predicting C:N ratio and SOC. Such products are expected to improve the current data availability of soil mapping based on remote sensing.

The prediction results showed different accuracies using different satellite sensors at different modeling resolutions (Table 3). Prediction models built by sensors with coarse spatial resolution can provide competitive or even better accuracy than models based on higher resolution sensors. This is consistent with the research by Xu et al. (2017a), who used different images to investigate soil property prediction in a small farmer environment and found that the soil prediction model with a coarser spatial resolution demonstrated competitive accuracy compared to the model with a higher spatial resolution. Similar results were reported by Kim et al. (2012) and Steinberg et al. (2016). The quantitative evaluation of prediction accuracy also showed that the construction of multi-scale prediction models can better predict soil properties. Some previous studies have highlighted the ability of multiple-scale methods to improve soil mapping (Chi et al., 2019; Taylor et al., 2013). Although it is well known that the spatial scale of input variables may have a significant impact on prediction performance, most previous digital soil mapping studies have only performed a single analytical scale (Forkuor et al., 2017). Therefore, we recommend building multi-scale prediction models for soil mapping to investigate the optimization of the spatial resolution of input variables, which may be beneficial for some soil properties.

The BRT models built by all available predictors at a resolution of 100 m had the highest accuracy, explaining about 47% and 40% of the SOC and the C:N ratio variability, respectively (Table 3). Compared with other soil mapping studies carried out in Switzerland, our model performance results were comparable. Nussbaum et al. (2014) used the robust external-drift kriging method to perform SOC mapping in the Swiss forest area, explaining 34% (0–30 cm) and 40% (0–100 cm) of the SOC variability at different depths, respectively. Blanchet et al. (2017) developed an RF model that was able to explain 29% in the Canton of Fribourg in Switzerland.

4.2. Environmental variables controlling the distribution of SOC content and C:N ratio in Switzerland

Terrain variables were identified as important predictors of our soil prediction models, especially SOC prediction models (Fig. 3). As a key factor in controlling the landscape scale hydrology and soil processes, topography has an important influence on soil formation, which in turn affects the spatial distribution of soil properties (Xu et al., 2018b). Among all terrain variables, VD and CS were the most important variables for SOC prediction and slope had the highest importance for the C:N ratio. This was also found by Schillaci et al. (2017) who reported that VD was the most important variable for SOC prediction in Sicily (Italy) in 1993 from all terrain variables. Previous studies have shown that CS is an effective auxiliary variable for soil property modeling (Adhikari et al., 2019; Amirian-Chakan et al., 2019). Slope controls the hydrological conditions in the landscape and produces different soil moisture conditions and flow patterns (Seibert et al., 2007). Indeed, some scholars have observed a strong relationship between soil properties (soil carbon and nitrogen) and slope at the field and landscape scales (Fissore et al., 2017; Jendoubi et al., 2019; Senthilkumar et al., 2009). Other terrain variables, such as TWI and VDCN also played an important role in our predictions. Considering the ability of TWI to capture soil moisture distribution, it is frequently used as a key predictor for mapping soil properties (Pei et al., 2010; Raduła et al., 2018).

In addition to topography, climate is also one of the five basic elements affecting the process of soil formation and its impact on soil carbon and nitrogen has been fully demonstrated (Dash et al., 2019; Ma and Chang, 2019). In the BRT model, MAP was identified as the fourth most important variable of SOC, revealing the moderate importance of rainfall for SOC mapping in the region (Fig. 3). Similar to our results, Deng et al. (2018) found that precipitation is a fairly important variable affecting SOC prediction in eastern China. The contribution of climate variables can be explained by their strong correlation with soil carbon and nitrogen pools. Temperature and rainfall are the most important climatic variables controlling soil carbon and nitrogen cycles (Geng et al., 2017; Lupon et al., 2015). They affect soil carbon and nitrogen pools through biotic or abiotic pathways (Lin et al., 2016). For example, temperature and rainfall could affect soil carbon and nitrogen dynamics by influencing net primary productivity (NPP) and related carbon and nitrogen input into the soil, as well as biological activity and litter accumulation and decomposition rates.

Besides topographic and climatic variables, remote sensing variables explain other sources of variation in SOC content and C:N ratio estimates, with a sum of corresponding importance of 56% and 78%, respectively (Fig. 3). Similar results were observed in previous studies, which reported that remote sensing variables most importantly contributed to explaining the variability of SOC (Wang et al., 2018a; Yang et al., 2016). Among the three satellite sensors, L8 had the highest importance of SOC prediction, indicating that it exerts a greater influence on the SOC distribution than S2 and S3 in the study area. In contrast, S2 exhibited a stronger influence than

S3 and L8 in the C:N ratio prediction. Different studies have highlighted the importance of L8 and S2 in predicting SOC content and the C:N ratio at various scales (Gholizadeh et al., 2018; Kumar et al., 2018; Rahman et al., 2020; Žížala et al., 2019). Remote sensing data can provide biophysical properties related to vegetation growth and soil conditions (Marshall and Thenkabail, 2015; Xu et al., 2017b). Vegetation, which is an important source of organic carbon and total nitrogen in the soil, is highly related to the spatial pattern of soil carbon and nitrogen in the topsoil (DeLuca et al., 2008; Jobbágy and Jackson, 2000). Many researchers have found that the relationship between soil and vegetation helps to understand the spatial distribution of soil properties through remote sensing technology (Maynard and Levi, 2017; Yang et al., 2019). This is supported by Anne et al. (2014) and Demattê et al. (2017) who explored the relationship between soil characteristics and vegetation with satellite remote sensing.

4.3. Spatial distribution of SOC content and C:N ratio in Switzerland

The digital soil maps obtained in this study exhibited similar patterns to previous soil information products, such as SoilGrid products (Hengl et al., 2017) and digital maps of SOC stock predicted by Nussbaum et al. (2014). However, the former had a relatively low spatial resolution and the latter only focused on the SOC of Swiss forest soils. The predicted map showed strong spatial variation of the topsoil SOC between the three main biogeographic regions (the Jura Mountains, the Central Plateau and the Alps) of Switzerland (Fig. 4). Specifically, higher SOC concentrations were mainly concentrated in the Alps, where high-altitude mountain areas usually have a cool climate and high forest cover. Most of the lower SOC concentrations were located in the Central Plateau at low altitudes, while the Jura Mountains with middle altitudes had relatively higher SOC values than the Central Plateau. These different SOC contents may be due to obvious differences in climatic conditions, vegetation types and the topography in these three biogeographic regions. In the Alps, SOC is promoted by abundant plant litter under dense forest cover, and the cold environment leads to a slow decomposition of organic matter, which contributes to the accumulation of SOC. The low-altitude Central Plateau with low SOC content was dominated by farmland and urban areas, which were often disturbed by human activities. Leifeld et al. (2005) reported that the Swiss SOC stock has been greatly reduced due to urbanization, deforestation and peatland cultivation. Land use has also been confirmed by other relevant studies as an important factor in determining the SOC content in Switzerland (Bolliger et al., 2008; Stumpf et al., 2018). The spatial pattern of the C:N ratio was closely related to the land use distribution pattern (see Price et al. (2015) for Swiss land use distribution), with higher values in the forest areas (see the map of forest cover in Waser et al. (2015)), especially the Alps (Fig. 4). A soil mapping study by Wang et al. (2018b) in Northeast China, also found that the spatial distribution of the C:N ratio corresponded to the land use pattern, where the C:N ratio for the forest area was higher than for other land use types (e.g., grassland and cultivated land). The relatively low C:N ratio of farmland might be due to less carbon input in soil and a high organic carbon mineralization rate during cultivation, while forest land has an obvious SOC accumulation and a low SOC decomposition rate (Chen et al., 2016; Yimer et al., 2007). Ballabio et al. (2019) and Beguin et al. (2017) reported that vegetation distribution significantly affected the C:N ratio distribution, with higher values observed under coniferous trees in Europe and Canada, respectively.

5. Conclusions

This work combined satellite sensors (L8, S2 and S3) with different spatial and temporal resolutions and three machine learning techniques to map the national distribution of SOC content and the C:N ratio in Switzerland at four spatial resolutions. Our conclusions can be summarized as follows:

- Comparative analysis showed that better predictions of soil properties can be achieved through quantitative evaluation when selecting prediction models, satellite sensors and the modeling resolution.
- Overall, the L8 and S2 sensors performed best and worst among satellite-based SOC models, respectively. These two sensors showed a better accuracy than S3 for C:N ratio mapping. However, the accuracy of the S3 sensor at a coarse resolution was either comparable or better.
- The best predictions for SOC content (R^2 = 0.470) and the C:N ratio (R^2 = 0.404) were achieved by BRT models constructed by all available predictors at a resolution of 100 m. In these models, the addition of remote sensing variables improved the prediction accuracy of SOC content and the C:N ratio by about 29.1% and 58.4%, respectively (in terms of R^2).
- The high relative importance of remote sensing images in the BRT model suggests their powerful ability to model national scale SOC content and the C:N ratio.
- The predicted maps of SOC content and the C:N ratio displayed significant spatial heterogeneity. In general, higher SOC concentrations were mainly concentrated in the Alps at high altitudes, while the C:N ratio shared a similar distribution pattern with land use and showed higher values for forest areas.

Acknowledgments

We thank the European Soil Data Centre, the European Space Agency and the U.S. Geological Survey for their open data policies. First author acknowledges the financial support provided by the China Scholarship Council.

References

Adhikari, K., Owens, P.R., Libohova, Z., Miller, D.M., Wills, S.A., Nemecek, J., 2019. Assessing soil organic carbon stock of Wisconsin, USA and its fate under future land use and climate change. Science of The Total Environment 667, 833-845.

Amirian-Chakan, A., Minasny, B., Taghizadeh-Mehrjardi, R., Akbarifazli, R., Darvishpasand, Z., Khordehbin, S., 2019. Some practical aspects of predicting texture data in digital soil mapping. Soil and Tillage Research 194, 104289.

Anne, N.J.P., Abd-Elrahman, A.H., Lewis, D.B., Hewitt, N.A., 2014. Modeling soil parameters using hyperspectral image reflectance in subtropical coastal wetlands. International Journal of Applied Earth Observation and Geoinformation 33, 47-56.

Arabameri, A., Yamani, M., Pradhan, B., Melesse, A., Shirani, K., Tien Bui, D., 2019. Novel ensembles of COPRAS multi-criteria decision-making with logistic regression, boosted regression tree, and random forest for spatial prediction of gully erosion susceptibility. Science of The Total Environment 688, 903-916.

Ballabio, C., Lugato, E., Fernández-Ugalde, O., Orgiazzi, A., Jones, A., Borrelli, P., Montanarella, L., Panagos, P., 2019. Mapping LUCAS topsoil chemical properties at European scale using Gaussian process regression. Geoderma 355, 113912.

Ballabio, C., Panagos, P., Lugato, E., Huang, J.-H., Orgiazzi, A., Jones, A., Fernández-Ugalde, O., Borrelli, P., Montanarella, L., 2018. Copper distribution in European topsoils: An assessment based on LUCAS soil survey. Science of The Total Environment 636, 282-298.

Beguin, 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.

Berger, M., Moreno, J., Johannessen, J.A., Levelt, P.F., Hanssen, R.F., 2012. ESA's sentinel missions in support of Earth system science. Remote Sensing of Environment 120, 84-90.

Bhattarai, N., Quackenbush, L.J., Dougherty, M., Marzen, L.J., 2015. A simple Landsat–MODIS fusion approach for monitoring seasonal evapotranspiration at 30 m spatial resolution. International Journal of Remote Sensing 36(1), 115-143.

Blanchet, G., Libohova, Z., Joost, S., Rossier, N., Schneider, A., Jeangros, B., Sinaj, S., 2017. Spatial variability of potassium in agricultural soils of the canton of Fribourg, Switzerland. Geoderma 290, 107-121.

Bolliger, J., Hagedorn, F., Leifeld, J., Böhl, J., Zimmermann, S., Soliva, R., Kienast, F., 2008. Effects of Land-Use Change on Carbon Stocks in Switzerland. Ecosystems 11(6), 895-907.

Bousbih, S., Zribi, M., Pelletier, C., Gorrab, 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 Sensing 11(13), 1520.

Broderick, D.E., Frey, K.E., Rogan, J., Alexander, H.D., Zimov, N.S., 2015. Estimating upper soil horizon carbon stocks in a permafrost watershed of Northeast Siberia by integrating field measurements with Landsat-5 TM and WorldView-2 satellite data. GIScience & Remote Sensing 52(2), 131-157.

Castaldi, F., Hueni, A., Chabrillat, S., Ward, K., Buttafuoco, G., Bomans, B., Vreys, K., Brell, M., van Wesemael, B., 2019. 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.

Chen, C.-P., Juang, K.-W., Cheng, C.-H., Pai, C.-W., 2016. Effects of adjacent land-use types on the distribution of soil organic carbon stocks in the montane area of central Taiwan. Botanical Studies 57(1), 32.

Chen, C., Chen, Z., Li, M., Liu, Y., Cheng, L., Ren, Y., 2014. Parallel relative radiometric normalisation for remote sensing image mosaics. Comput. Geosci. 73, 28-36.

Chen, D., Chang, N., Xiao, J., Zhou, Q., Wu, W., 2019. Mapping dynamics of soil organic matter in croplands with MODIS data and machine learning algorithms. Science of The Total Environment 669, 844-855.

Chi, Y., Zhao, M., Sun, J., Xie, Z., Wang, E., 2019. Mapping soil total nitrogen in an estuarine area with high landscape fragmentation using a multiple-scale approach. Geoderma 339, 70-84.

Clark, M.L., 2017. Comparison of simulated hyperspectral HyspIRI and multispectral Landsat 8 and Sentinel-2 imagery for multi-seasonal, regional land-cover mapping. Remote Sensing of Environment 200, 311-325.

Clevers, J.G.P.W., Gitelson, A.A., 2013. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. International Journal of Applied Earth Observation and Geoinformation 23, 344-351.

Cui, J., Zhang, X., Luo, M., 2018. Combining Linear Pixel Unmixing and STARFM for Spatiotemporal Fusion of Gaofen-1 Wide Field of View Imagery and MODIS Imagery. Remote Sensing 10(7), 1047.

Dash, P.K., Bhattacharyya, P., Roy, K.S., Neogi, S., Nayak, A.K., 2019. Environmental constraints' sensitivity of soil organic carbon decomposition to temperature, management practices and climate change. Ecol Indicators 107, 105644.

Davis, E., Wang, C., Dow, K., 2019. Comparing Sentinel-2 MSI and Landsat 8 OLI in soil salinity detection: a case study of agricultural lands in coastal North Carolina. International Journal of Remote Sensing 40(16), 6134-6153.

Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Nicola, L., Rabaute, T., Savinaud, M., Udroiu, C., Valero, S., Bégué, A., Dejoux, J.-F., El Harti, A., Ezzahar, J., Kussul, N., Labbassi, K., Lebourgeois, V., Miao, Z., Newby, T., Nyamugama, A., Salh, N., Shelestov, A., Simonneaux, V., Traore, P.S., Traore, S.S., Koetz, B., 2019. Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world. Remote Sensing of Environment 221, 551-568.

DeLuca, T.H., Zackrisson, O., Gundale, M.J., Nilsson, M.-C., 2008. Ecosystem Feedbacks and Nitrogen Fixation in Boreal Forests. Science 320(5880), 1181-1181.

Demattê, J.A.M., Sayão, V.M., Rizzo, R., Fongaro, C.T., 2017. Soil class and attribute dynamics and their relationship with natural vegetation based on satellite remote sensing. Geoderma 302, 39-51.

Deng, X., Chen, X., Ma, W., Ren, Z., Zhang, M., Grieneisen, M.L., Long, W., Ni, Z., Zhan, Y., Lv, X., 2018. Baseline map of organic carbon stock in farmland topsoil in East China. Agric, Ecosyst Environ 254, 213-223.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J Anim Ecol 77(4), 802-813.

Fernández-Ugalde, O., Jones, A., Meuli, R.G., 2020. Comparison of sampling with a spade and gouge auger for topsoil monitoring at the continental scale. Eur J Soil Sci 71(2), 137-150.

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37(12), 4302-4315.

Fissore, C., Dalzell, B.J., Berhe, A.A., Voegtle, M., Evans, M., Wu, A., 2017. Influence of topography on soil organic carbon dynamics in a Southern California grassland. Catena 149, 140-149.

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.

Friedman, J., Hastie, T., Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). Ann. Statist. 28(2), 337-407.

Geng, Y., Baumann, F., Song, C., Zhang, M., Shi, Y., Kühn, P., Scholten, T., He, J.-S., 2017. Increasing temperature reduces the coupling between available nitrogen and phosphorus in soils of Chinese grasslands. Sci Rep 7(1), 43524.

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 Sensing of Environment 218, 89-103.

Gomes, L.C., Faria, R.M., de Souza, E., Veloso, G.V., Schaefer, C.E.G.R., Filho, E.I.F., 2019. Modelling and mapping soil organic carbon stocks in Brazil. Geoderma 340, 337-350.

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 Sensing 11(5), 565.

Hamzehpour, N., Shafizadeh-Moghadam, H., Valavi, R., 2019. Exploring the driving forces and digital mapping of soil organic carbon using remote sensing and soil texture. Catena 182, 104141.

Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Gonzalez, M.R., 2014. SoilGrids1km — Global Soil Information Based on Automated Mapping. PLoS One 9(8), e105992.

Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., Mendes de Jesus, J., Tamene, L., Tondoh, J.E., 2015. Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions. PLoS One 10(6), e0125814.

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.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25(15), 1965-1978.

Hoffmann, U., Hoffmann, T., Jurasinski, G., Glatzel, S., Kuhn, N.J., 2014. Assessing the spatial variability of soil organic carbon stocks in an alpine setting (Grindelwald, Swiss Alps). Geoderma 232-234, 270-283.

Jendoubi, D., Liniger, H., Ifejika Speranza, C., 2019. Impacts of land use and topography on soil organic carbon in a Mediterranean landscape (north-western Tunisia). SOIL 5(2), 239-251.

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.

Jobbágy, E.G., Jackson, R.B., 2000. THE VERTICAL DISTRIBUTION OF SOIL ORGANIC CARBON AND ITS RELATION TO CLIMATE AND VEGETATION. Ecol Appl 10(2), 423-436.

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. Environmental Earth Sciences 77(5), 203.

Ke, Y., Im, J., Lee, J., Gong, H., Ryu, Y., 2015. Characteristics of Landsat 8 OLI-derived NDVI by comparison with multiple satellite sensors and in-situ observations. Remote Sensing of Environment 164, 298-313.

Keskin, H., Grunwald, S., Harris, W.G., 2019. Digital mapping of soil carbon fractions with machine learning. Geoderma 339, 40-58.

Khanal, S., Fulton, J., Klopfenstein, A., Douridas, N., Shearer, S., 2018. Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. Comput Electron Agric 153, 213-225.

Kim, J., Grunwald, S., Rivero, R.G., Robbins, R., 2012. Multi-scale Modeling of Soil Series Using Remote Sensing in a Wetland Ecosystem. Soil Science Society of America Journal 76(6), 2327-2341.

Kokhanovsky, A., Lamare, M., Danne, O., Brockmann, C., Dumont, M., Picard, G., Arnaud, L., Favier, V., Jourdain, B., Le Meur, E., Di Mauro, B., Aoki, T., Niwano, M., Rozanov, V., Korkin, S., Kipfstuhl, S., Freitag, J., Hoerhold, M., Zuhr, A., Vladimirova, D., Faber, A.-K., Steen-Larsen, C.H., Wahl, S., Andersen, K.J., Vandecrux, B., van As, D., Mankoff, D.K., Kern, M., Zege, E., Box, E.J., 2019. Retrieval of Snow Properties from the Sentinel-3 Ocean and Land Colour Instrument. Remote Sensing 11(19).

Kumar, P., Sajjad, H., Tripathy, B.R., Ahmed, R., Mandal, V.P., 2018. Prediction of spatial soil organic carbon distribution using Sentinel-2A and field inventory data in Sariska Tiger Reserve. Nat Hazards 90(2), 693-704.

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., Schmullius, C., Schrodt, 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 Sensing 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. Ecol Indicators 70, 317-339.

Leifeld, J., Bassin, S., Fuhrer, J., 2005. Carbon stocks in Swiss agricultural soils predicted by land-use, soil characteristics, and altitude. Agric, Ecosyst Environ 105(1), 255-266.

Li, J., Roy, D.P., 2017. A Global Analysis of Sentinel-2A, Sentinel-2B and Landsat-8 Data Revisit Intervals and Implications for Terrestrial Monitoring. Remote Sensing 9(9), 902.

Li, X., Feng, R., Guan, X., Shen, H., Zhang, L., 2019. Remote Sensing Image Mosaicking: Achievements and Challenges. IEEE Geoscience and Remote Sensing Magazine 7(4), 8-22.

Li, X., Shen, H., Zhang, L., Zhang, H., Yuan, Q., Yang, G., 2014. Recovering Quantitative Remote Sensing Products

Contaminated by Thick Clouds and Shadows Using Multitemporal Dictionary Learning. IEEE Transactions on Geoscience and Remote Sensing 52(11), 7086-7098.

Liang, Z., Chen, S., Yang, Y., Zhou, Y., Shi, Z., 2019. High-resolution three-dimensional mapping of soil organic carbon in China: Effects of SoilGrids products on national modeling. Science of The Total Environment 685, 480-489.

Lin, C., Zhu, A.X., Wang, Z., Wang, X., Ma, R., 2020. The refined spatiotemporal representation of soil organic matter based on remote images fusion of Sentinel-2 and Sentinel-3. International Journal of Applied Earth Observation and Geoinformation 89, 102094.

Lin, L., Zhu, B., Chen, C., Zhang, Z., Wang, Q.-B., He, J.-S., 2016. Precipitation overrides warming in mediating soil nitrogen pools in an alpine grassland ecosystem on the Tibetan Plateau. Sci Rep 6(1), 31438.

Loiseau, T., Chen, S., Mulder, V.L., Román Dobarco, M., Richer-de-Forges, A.C., Lehmann, S., Bourennane, H., Saby, N.P.A., Martin, M.P., Vaudour, E., Gomez, C., Lagacherie, P., Arrouays, D., 2019. Satellite data integration for soil clay content modelling at a national scale. International Journal of Applied Earth Observation and Geoinformation 82, 101905.

Lou, Y., Xu, M., Chen, X., He, X., Zhao, K., 2012. Stratification of soil organic C, N and C:N ratio as affected by conservation tillage in two maize fields of China. Catena 95, 124-130.

Lupon, A., Gerber, S., Sabater, F., Bernal, S., 2015. Climate response of the soil nitrogen cycle in three forest types of a headwater Mediterranean catchment. Journal of Geophysical Research: Biogeosciences 120(5), 859-875.

Ma, M., Chang, R., 2019. Temperature drive the altitudinal change in soil carbon and nitrogen of montane forests: Implication for global warming. Catena 182, 104126.

Marshall, M., Thenkabail, P., 2015. Advantage of hyperspectral EO-1 Hyperion over multispectral IKONOS, GeoEye-1, WorldView-2, Landsat ETM+, and MODIS vegetation indices in crop biomass estimation. ISPRS Journal of Photogrammetry and Remote Sensing 108, 205-218.

Martin, M.P., Wattenbach, M., Smith, P., Meersmans, J., Jolivet, C., Boulonne, L., Arrouays, D., 2011. Spatial distribution of soil organic carbon stocks in France. Biogeosciences 8(5), 1053-1065.

Matos-Moreira, M., Lemercier, B., Dupas, R., Michot, D., Viaud, V., Akkal-Corfini, N., Louis, B., Gascuel-Odoux, C., 2017. High-resolution mapping of soil phosphorus concentration in agricultural landscapes with readily

available or detailed survey data. Eur J Soil Sci 68(3), 281-294.
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.
Minasny, B., McBratney, A.B., Lark, R.M., 2008. Digital Soil Mapping Technologies for Countries with Sparse Data Infrastructures. In: A.E. Hartemink, A. McBratney, M.d.L. Mendonça-Santos (Eds.), Digital Soil Mapping with Limited Data. Springer Netherlands, Dordrecht, pp. 15-30.
Misra, A., Vojinovic, Z., Ramakrishnan, B., Luijendijk, A., Ranasinghe, R., 2018. Shallow water bathymetry mapping using Support Vector Machine (SVM) technique and multispectral imagery. International Journal of Remote Sensing 39(13), 4431-4450.
Nussbaum, M., Papritz, A., Baltensweiler, A., Walthert, L., 2014. Estimating soil organic carbon stocks of Swiss forest soils by robust external-drift kriging. Geosci. Model Dev. 7(3), 1197-1210.
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.
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. Ecol Indicators 77, 139-150.
Padarian, J., Minasny, B., McBratney, A.B., 2020. Machine learning and soil sciences: a review aided by machine learning tools. SOIL 6(1), 35-52.
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., Meusburger, K., Ballabio, C., Borrelli, P., Alewell, C., 2014. Soil erodibility in Europe: A high-resolution dataset based on LUCAS. Science of The Total Environment 479-480, 189-200.
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.
Paul, S.S., Coops, N.C., Johnson, M.S., Krzic, M., Chandna, A., Smukler, S.M., 2020. Mapping soil organic carbon and clay using remote sensing to predict soil workability for enhanced climate change adaptation. Geoderma 363, 114177.

Pei, T., Qin, C.-Z., Zhu, A.X., Yang, L., Luo, M., Li, B., Zhou, C., 2010. Mapping soil organic matter using the topographic wetness index: A comparative study based on different flow-direction algorithms and kriging methods. Ecol Indicators 10(3), 610-619.

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.

Price, B., Kienast, F., Seidl, I., Ginzler, C., Verburg, P.H., Bolliger, J., 2015. Future landscapes of Switzerland: Risk areas for urbanisation and land abandonment. Applied Geography 57, 32-41.

Raduła, M.W., Szymura, T.H., Szymura, M., 2018. Topographic wetness index explains soil moisture better than bioindication with Ellenberg's indicator values. Ecol Indicators 85, 172-179.

Rahman, M.M., Zhang, X., Ahmed, I., Iqbal, Z., Zeraatpisheh, M., Kanzaki, M., Xu, M., 2020. Remote Sensing-Based Mapping of Senescent Leaf C:N Ratio in the Sundarbans Reserved Forest Using Machine Learning Techniques. Remote Sensing 12(9), 1375.

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.

Rasaei, Z., Bogaert, P., 2019. Spatial filtering and Bayesian data fusion for mapping soil properties: A case study combining legacy and remotely sensed data in Iran. Geoderma 344, 50-62.

Rial, M., Martínez Cortizas, A., Rodríguez-Lado, L., 2017. Understanding the spatial distribution of factors controlling topsoil organic carbon content in European soils. Science of The Total Environment 609, 1411-1422.

Samuel-Rosa, A., Heuvelink, G.B.M., Vasques, G.M., Anjos, L.H.C., 2015. Do more detailed environmental covariates deliver more accurate soil maps? Geoderma 243-244, 214-227.

Schillaci, C., Acutis, M., Lombardo, L., Lipani, A., Fantappiè, M., Märker, M., Saia, S., 2017. Spatio-temporal topsoil organic carbon mapping of a semi-arid Mediterranean region: The role of land use, soil texture, topographic indices and the influence of remote sensing data to modelling. Science of The Total Environment 601-602, 821-832.

Schmid, T., Koch, M., DiBlasi, M., Hagos, M., 2008. Spatial and spectral analysis of soil surface properties for an archaeological area in Aksum, Ethiopia, applying high and medium resolution data. Catena 75(1), 93-101.

Scudiero, E., Skaggs, T.H., Corwin, D.L., 2014. Regional scale soil salinity evaluation using Landsat 7, western San Joaquin Valley, California, USA. Geoderma Regional 2-3, 82-90.

Seibert, J., Stendahl, J., Sørensen, R., 2007. Topographical influences on soil properties in boreal forests. Geoderma 141(1), 139-148.

Senthilkumar, S., Kravchenko, A.N., Robertson, G.P., 2009. Topography Influences Management System Effects on Total Soil Carbon and Nitrogen. Soil Science Society of America Journal 73(6), 2059-2067.

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.

Steinberg, A., Chabrillat, S., Stevens, A., Segl, K., Foerster, S., 2016. Prediction of Common Surface Soil Properties Based on Vis-NIR Airborne and Simulated EnMAP Imaging Spectroscopy Data: Prediction Accuracy and Influence of Spatial Resolution. Remote Sensing 8(7), 613.

Stumpf, F., Keller, A., Schmidt, K., Mayr, A., Gubler, A., Schaepman, M., 2018. Spatio-temporal land use dynamics and soil organic carbon in Swiss agroecosystems. Agric, Ecosyst Environ 258, 129-142.

Szatmári, G., Pirkó, B., Koós, S., Laborczi, A., Bakacsi, Z., Szabó, J., Pásztor, L., 2019. Spatio-temporal assessment of topsoil organic carbon stock change in Hungary. Soil and Tillage Research 195, 104410.

Taghadosi, M.M., Hasanlou, M., Eftekhari, K., 2019. Retrieval of soil salinity from Sentinel-2 multispectral imagery. European Journal of Remote Sensing 52(1), 138-154.

Taghizadeh-Mehrjardi, R., Schmidt, K., Amirian-Chakan, A., Rentschler, T., Zeraatpisheh, M., Sarmadian, F., Valavi, R., Davatgar, N., Behrens, T., Scholten, T., 2020. Improving the Spatial Prediction of Soil Organic Carbon Content in Two Contrasting Climatic Regions by Stacking Machine Learning Models and Rescanning Covariate Space. Remote Sensing 12(7), 1095.

Taylor, J.A., Jacob, F., Galleguillos, M., Prévot, L., Guix, N., Lagacherie, P., 2013. The utility of remotely-sensed vegetative and terrain covariates at different spatial resolutions in modelling soil and watertable depth (for digital soil mapping). Geoderma 193-194, 83-93.

Vågen, T.-G., Winowiecki, L.A., Tondoh, J.E., Desta, L.T., Gumbricht, T., 2016. Mapping of soil properties and land degradation risk in Africa using MODIS reflectance. Geoderma 263, 216-225.

Vaudour, E., Gomez, C., Fouad, Y., Lagacherie, P., 2019. Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems. Remote Sensing of Environment 223, 21-33.

Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J.P., Camps-Valls, G., Moreno, J., 2012. Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3. Remote Sensing of Environment 118, 127-139.

Vitharana, U.W.A., Mishra, U., Mapa, R.B., 2019. National soil organic carbon estimates can improve global estimates. Geoderma 337, 55-64.

Wadoux, A.M.J.C., 2019. Using deep learning for multivariate mapping of soil with quantified uncertainty. Geoderma 351, 59-70.

Wan, X., Huang, Z., He, Z., Yu, Z., Wang, M., Davis, M.R., Yang, Y., 2015. Soil C:N ratio is the major determinant of soil microbial community structure in subtropical coniferous and broadleaf forest plantations. Plant Soil 387(1), 103-116.

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., 2019. 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., Ding, J., Yu, D., Teng, D., He, B., Chen, X., Ge, X., Zhang, Z., Wang, Y., Yang, X., Shi, T., Su, F., 2020a. Machine learning-based detection of soil salinity in an arid desert region, Northwest China: A comparison between Landsat-8 OLI and Sentinel-2 MSI. Science of The Total Environment 707, 136092.

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. Ecol Indicators 84, 263-272.

Wang, S., Adhikari, K., Zhuang, Q., Gu, H., Jin, X., 2020b. Impacts of urbanization on soil organic carbon stocks in the northeast coastal agricultural areas of China. Science of The Total Environment 721, 137814.

Wang, Z., Shi, W., Zhou, W., Li, X., Yue, T., 2020c. Comparison of additive and isometric log-ratio transformations combined with machine learning and regression kriging models for mapping soil particle size fractions.

Geoderma 365, 114214.

Waser, L.T., Fischer, C., Wang, Z., Ginzler, C., 2015. Wall-to-Wall Forest Mapping Based on Digital Surface Models from Image-Based Point Clouds and a NFI Forest Definition. Forests 6(12), 4510-4528.

Were, K., Bui, D.T., Dick, Ø.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. Ecol Indicators 52, 394-403.

Were, K., Singh, B.R., Dick, Ø.B., 2016. Spatially distributed modelling and mapping of soil organic carbon and total nitrogen stocks in the Eastern Mau Forest Reserve, Kenya. Journal of Geographical Sciences 26(1), 102-124.

Wiesmeier, M., Barthold, F., Blank, B., Kögel-Knabner, I., 2011. Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. Plant Soil 340(1), 7-24.

Wu, J., 2020. Change in soil microbial biomass and regulating factors in an alpine meadow site on the Qinghai-Tibetan Plateau. Soil Sci Plant Nutr 66(1), 177-194.

Wulder, M.A., Loveland, T.R., Roy, D.P., Crawford, C.J., Masek, J.G., Woodcock, C.E., Allen, R.G., Anderson, M.C., Belward, A.S., Cohen, W.B., Dwyer, J., Erb, A., Gao, F., Griffiths, P., Helder, D., Hermosilla, T., Hipple, J.D., Hostert, P., Hughes, M.J., Huntington, J., Johnson, D.M., Kennedy, R., Kilic, A., Li, Z., Lymburner, L., McCorkel, J., Pahlevan, N., Scambos, T.A., Schaaf, C., Schott, J.R., Sheng, Y., Storey, J., Vermote, E., Vogelmann, J., White, J.C., Wynne, R.H., Zhu, Z., 2019. Current status of Landsat program, science, and applications. Remote Sensing of Environment 225, 127-147.

Xie, Y., Sha, Z., Yu, M., 2008. Remote sensing imagery in vegetation mapping: a review. Journal of Plant Ecology 1(1), 9-23.

Xiong, X., Grunwald, S., Myers, D.B., Kim, J., Harris, W.G., Comerford, N.B., 2014. Holistic environmental soil-landscape modeling of soil organic carbon. Environ Model Software 57, 202-215.

Xu, E., Zhang, H., Xu, Y., 2019. Effect of Large-Scale Cultivated Land Expansion on the Balance of Soil Carbon and Nitrogen in the Tarim Basin. Agronomy 9(2), 86.

Xu, H., Liu, G., Wu, X., Smoak, J.M., Mu, C., Ma, X., Zhang, X., Li, H., Hu, G., 2018a. Soil enzyme response to permafrost collapse in the Northern Qinghai-Tibetan Plateau. Ecol Indicators 85, 585-593.

Xu, Y., Li, Y., Li, H., Wang, L., Liao, X., Wang, J., Kong, C., 2018b. Effects of topography and soil properties on soil selenium distribution and bioavailability (phosphate extraction): A case study in Yongjia County, China. Science of The Total Environment 633, 240-248.

Xu, Y., Smith, S.E., Grunwald, S., Abd-Elrahman, A., Wani, S.P., 2017a. Evaluating the effect of remote sensing image spatial resolution on soil exchangeable potassium prediction models in smallholder farm settings. J Environ Manage 200, 423-433.

Xu, Y., Smith, S.E., Grunwald, S., Abd-Elrahman, A., Wani, S.P., 2017b. Incorporation of satellite remote sensing pan-sharpened imagery into digital soil prediction and mapping models to characterize soil property variability in small agricultural fields. ISPRS Journal of Photogrammetry and Remote Sensing 123, 1-19.

Xu, Y., Wang, X., Bai, J., Wang, D., Wang, W., Guan, Y., 2020. Estimating the spatial distribution of soil total nitrogen and available potassium in coastal wetland soils in the Yellow River Delta by incorporating multi-source data. Ecol Indicators 111, 106002.

Yang, L., He, X., Shen, F., Zhou, C., Zhu, A.X., Gao, B., Chen, Z., Li, M., 2020. Improving prediction of soil organic carbon content in croplands using phenological parameters extracted from NDVI time series data. Soil and Tillage Research 196, 104465.

Yang, R.-M., Guo, W.-W., 2019. 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. Ecol Indicators 60, 870-878.

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.

Yimer, F., Ledin, S., Abdelkadir, A., 2007. Changes in soil organic carbon and total nitrogen contents in three adjacent land use types in the Bale Mountains, south-eastern highlands of Ethiopia. For Ecol Manage 242(2), 337-342.

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.

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. Comput Electron Agric 160, 23-30.

Zhi, J., Zhang, G., Yang, R., Yang, F., Jin, C., Liu, F., Song, X., Zhao, Y., Li, D., 2018. An Insight into Machine Learning Algorithms to Map the Occurrence of the Soil Mattic Horizon in the Northeastern Qinghai-Tibetan Plateau. Pedosphere 28(5), 739-750.

Žížala, D., Minařík, R., Zádorová, T., 2019. Soil Organic Carbon Mapping Using Multispectral Remote Sensing Data: Prediction Ability of Data with Different Spatial and Spectral Resolutions. Remote Sensing 11(24), 2947.

Table 1

Different combinations of environmental variables used as inputs for SOC and the C:N ratio prediction.

NO.	Model	Environmental variables
1	Model I	Landsat-8 predictors
2	Model II	Sentinel-2 predictors
3	Model III	Sentinel-3 predictors
4	Model IV	Climate + topography
5	Model V	All available predictors

Table 2

Descriptive statistics of SOC (g/kg) and the C:N ratio.

	Minimum	Maximum	Mean	Median	Standard deviation (SD)	Skewness
SOC	8.90	151.50	43.93	37.45	27.65	1.57
LnSOC	2.19	5.02	3.61	3.62	0.58	0.13
C:N ratio	3.33	22.26	10.60	9.16	3.24	1.06
LnC:N ratio	1.20	3.10	2.32	2.21	0.29	0.16

Notes: LnSOC, log-transformed SOC; LnC:N ratio, log-transformed C:N ratio.

Table 3

Performance results of RF, BRT and SVM in predicting SOC and the C:N ratio based on different combinations of environmental variables at different spatial resolutions. The most accurate results are shown in bold.

Modeling technique	Model	SOC			C:N ratio		
		MAE	RMSE	R^2	MAE	RMSE	R^2
BRT	Model I						
	20 m	0.398	0.497	0.266	0.179	0.238	0.353
	100 m	0.379	0.468	0.363	0.182	0.241	0.330
	400 m	0.411	0.498	0.279	0.213	0.264	0.198
	800 m	0.425	0.526	0.226	0.225	0.270	0.142
	Model II						
	20 m	0.416	0.525	0.201	0.187	0.245	0.297
	100 m	0.417	0.515	0.209	0.180	0.240	0.328
	400 m	0.429	0.531	0.184	0.217	0.268	0.154
	800 m	0.448	0.558	0.154	0.227	0.276	0.119
	Model III						
	20 m	0.407	0.510	0.243	0.212	0.263	0.203
	100 m	0.402	0.510	0.243	0.211	0.261	0.197
	400 m	0.405	0.503	0.263	0.219	0.268	0.160
	800 m	0.413	0.512	0.253	0.210	0.264	0.208
	Model IV						
	20 m	0.389	0.484	0.355	0.205	0.259	0.241
	100 m	0.384	0.478	0.364	0.207	0.259	0.255
	400 m	0.380	0.476	0.359	0.205	0.257	0.242
	800 m	0.393	0.492	0.357	0.219	0.271	0.171
	Model V						
	20 m	0.369	0.468	0.388	0.178	0.232	0.379
	100 m	**0.336**	**0.437**	**0.470**	**0.167**	**0.223**	**0.404**
	400 m	0.351	0.446	0.433	0.196	0.253	0.268
	800 m	0.360	0.464	0.395	0.210	0.260	0.212
RF	Model I						
	20 m	0.401	0.502	0.252	0.185	0.252	0.279
	100 m	0.394	0.482	0.343	0.187	0.246	0.300
	400 m	0.434	0.518	0.219	0.220	0.276	0.154
	800 m	0.425	0.531	0.197	0.233	0.283	0.100
	Model II						
	20 m	0.435	0.547	0.138	0.190	0.250	0.264
	100 m	0.415	0.513	0.234	0.180	0.244	0.309
	400 m	0.446	0.549	0.154	0.227	0.284	0.117
	800 m	0.453	0.564	0.121	0.229	0.278	0.111
	Model III						
	20 m	0.411	0.517	0.232	0.217	0.274	0.159
	100 m	0.403	0.510	0.244	0.209	0.264	0.188
	400 m	0.420	0.516	0.231	0.229	0.281	0.105
	800 m	0.424	0.525	0.203	0.211	0.265	0.195
	Model IV						
	20 m	0.379	0.479	0.357	0.217	0.271	0.182

	100 m	0.379	0.476	0.352	0.211	0.274	0.191
	400 m	0.377	0.476	0.352	0.210	0.264	0.196
	800 m	0.392	0.488	0.353	0.224	0.284	0.131
	Model V						
	20 m	0.361	0.461	0.379	0.173	0.230	0.375
	100 m	0.342	0.443	0.431	0.166	0.223	0.397
	400 m	0.352	0.448	0.428	0.197	0.253	0.261
	800 m	0.365	0.465	0.398	0.207	0.259	0.234
SVM	Model I						
	20 m	0.388	0.488	0.285	0.174	0.235	0.363
	100 m	0.383	0.478	0.337	0.175	0.236	0.359
	400 m	0.399	0.494	0.280	0.199	0.263	0.215
	800 m	0.418	0.517	0.231	0.220	0.275	0.164
	Model II						
	20 m	0.410	0.519	0.213	0.179	0.241	0.315
	100 m	0.404	0.505	0.253	0.174	0.238	0.334
	400 m	0.433	0.537	0.171	0.201	0.265	0.202
	800 m	0.435	0.550	0.167	0.227	0.283	0.116
	Model III						
	20 m	0.406	0.516	0.236	0.198	0.257	0.247
	100 m	0.406	0.516	0.232	0.200	0.259	0.241
	400 m	0.405	0.509	0.256	0.209	0.271	0.185
	800 m	0.408	0.507	0.290	0.200	0.268	0.239
	Model IV						
	20 m	0.376	0.488	0.335	0.197	0.257	0.264
	100 m	0.355	0.466	0.398	0.193	0.256	0.269
	400 m	0.375	0.479	0.350	0.196	0.259	0.269
	800 m	0.388	0.479	0.372	0.215	0.276	0.195
	Model V						
	20 m	0.371	0.469	0.369	0.173	0.231	0.374
	100 m	0.354	0.451	0.419	0.162	0.223	0.398
	400 m	0.366	0.459	0.401	0.189	0.251	0.295
	800 m	0.369	0.457	0.402	0.195	0.261	0.266

Notes: Model I, Landsat-8 predictors; Model II, Sentinel-2 predictors; Model III, Sentinel-3 predictors; Model IV, climate + topography; Model V, all available predictors.

Figure Legends

Fig. 1. Soil sampling points superimposed on digital elevation model (a), Landsat-8 (b), Sentinel-3 (c) and Sentinel-2 (d) data in Switzerland.

Fig. 2. Overview of the flowchart for SOC and the C:N ratio prediction in Switzerland.

Fig. 3. Relative importance of the twenty most important environmental variables used for the C:N ratio and SOC prediction in Model V at a resolution of 100 m based on BRT and RF. Model V, all available predictors; TWI, SAGA wetness index; VD, valley depth; CS, catchment slope; VDCN, vertical distance to channel network; L8_1 to L8_7 correspond to band 1 to band 7 of Landsat-8 OLI data, respectively; S3_B1 to S3_B21 correspond to band 1 to band 21 of Sentinel-3 OLCI data, respectively; S2_B2 to S2_B12 correspond to band 2 to band 12 of Sentinel-2 MSI data, respectively.

Fig. 4. Mean SOC content and C:N ratio maps predicted by 100 runs of BRT in Model V at a resolution of 100 m and their corresponding standard deviation maps (Model V: all available predictors).

Fig. 5. Maps of SOC predicted by BRT in Model V at different resolutions (Model V: all available predictors).

Fig. 6. Maps of C:N ratio predicted by BRT in Model V at different resolutions (Model V: all available predictors).

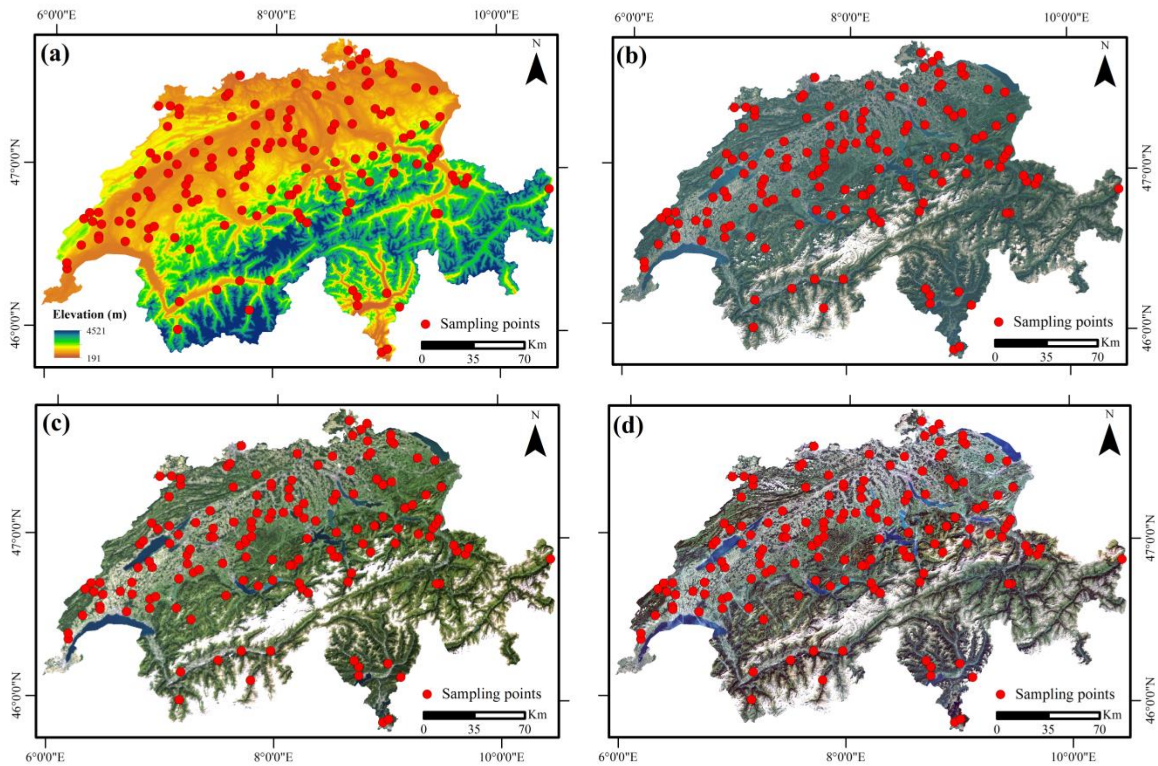

(a)
(b)
(c)
(d)
6°0'0"E
8°0'0"E
10°0'0"E
47°0'0"N
46°0'0"N
N
Elevation (m)
4521
191
Sampling points
Km
0
35
70

Fig. 1

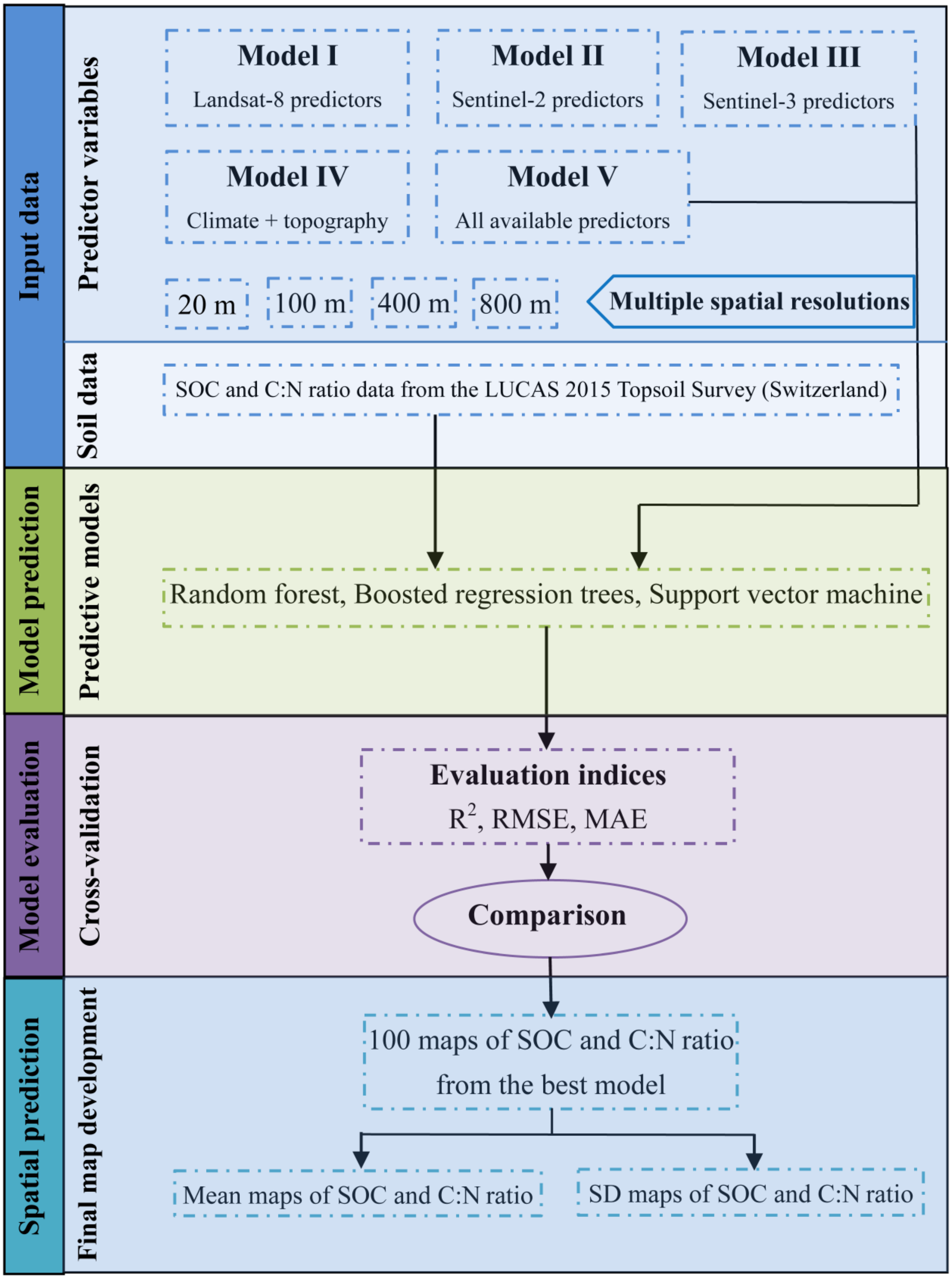
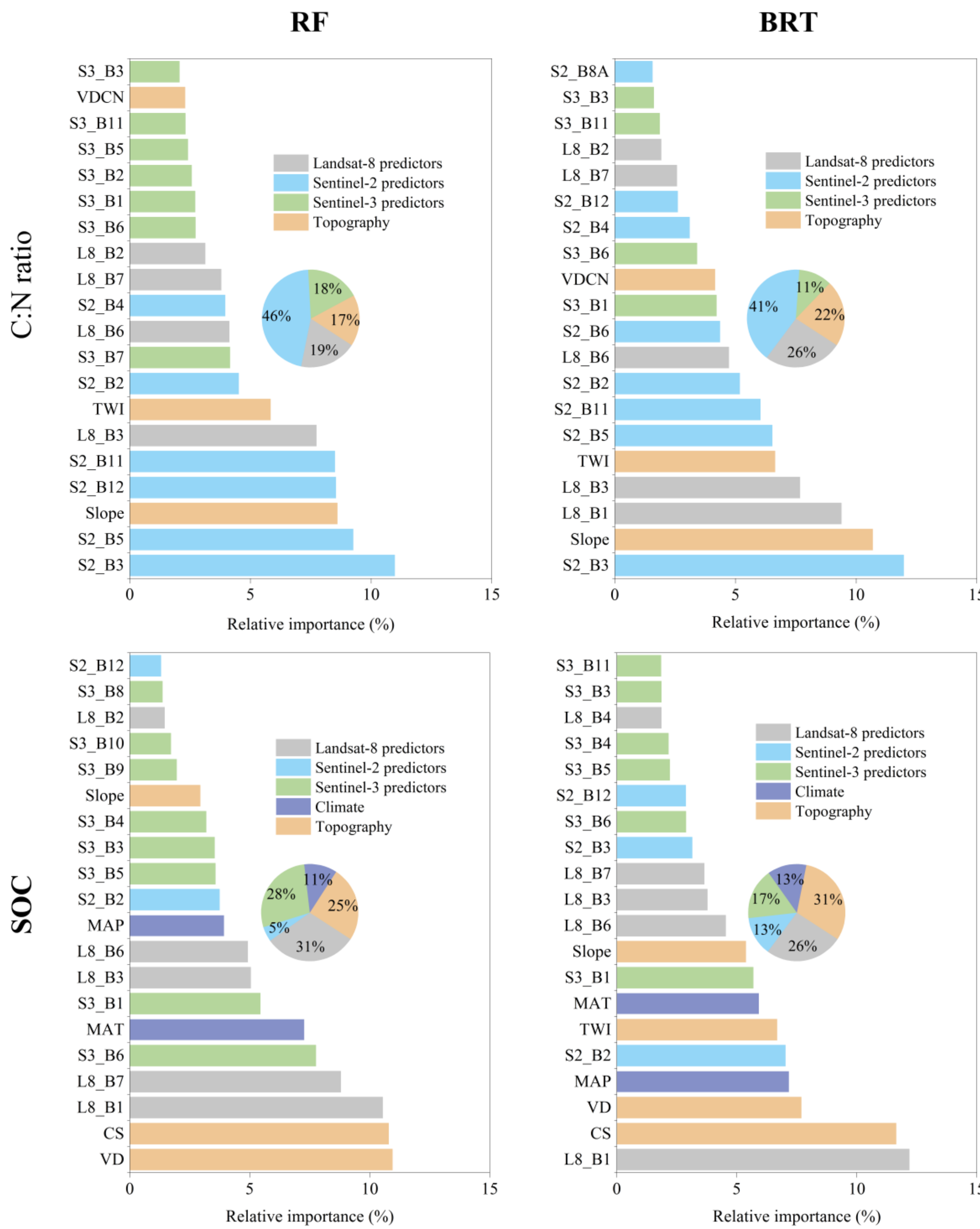

Input data
Predictor variables
Model I
Landsat-8 predictors
Model II
Sentinel-2 predictors
Model III
Sentinel-3 predictors
Model IV
Climate + topography
Model V
All available predictors
20 m
100 m
400 m
800 m
Multiple spatial resolutions
Soil data
SOC and C:N ratio data from the LUCAS 2015 Topsoil Survey (Switzerland)
Model prediction
Predictive models
Random forest, Boosted regression trees, Support vector machine
Model evaluation
Cross-validation
Evaluation indices
R², RMSE, MAE
Comparison
Spatial prediction
Final map development
100 maps of SOC and C:N ratio from the best model
Mean maps of SOC and C:N ratio
SD maps of SOC and C:N ratio

Fig. 2



999

1000 **Fig. 3**

1001

1002

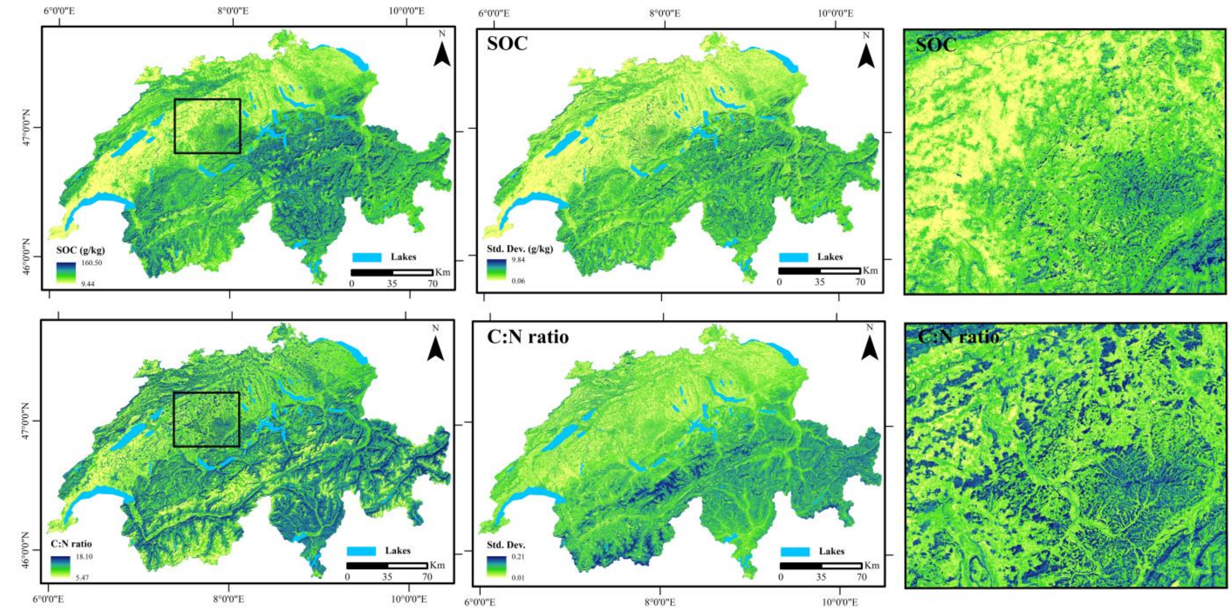
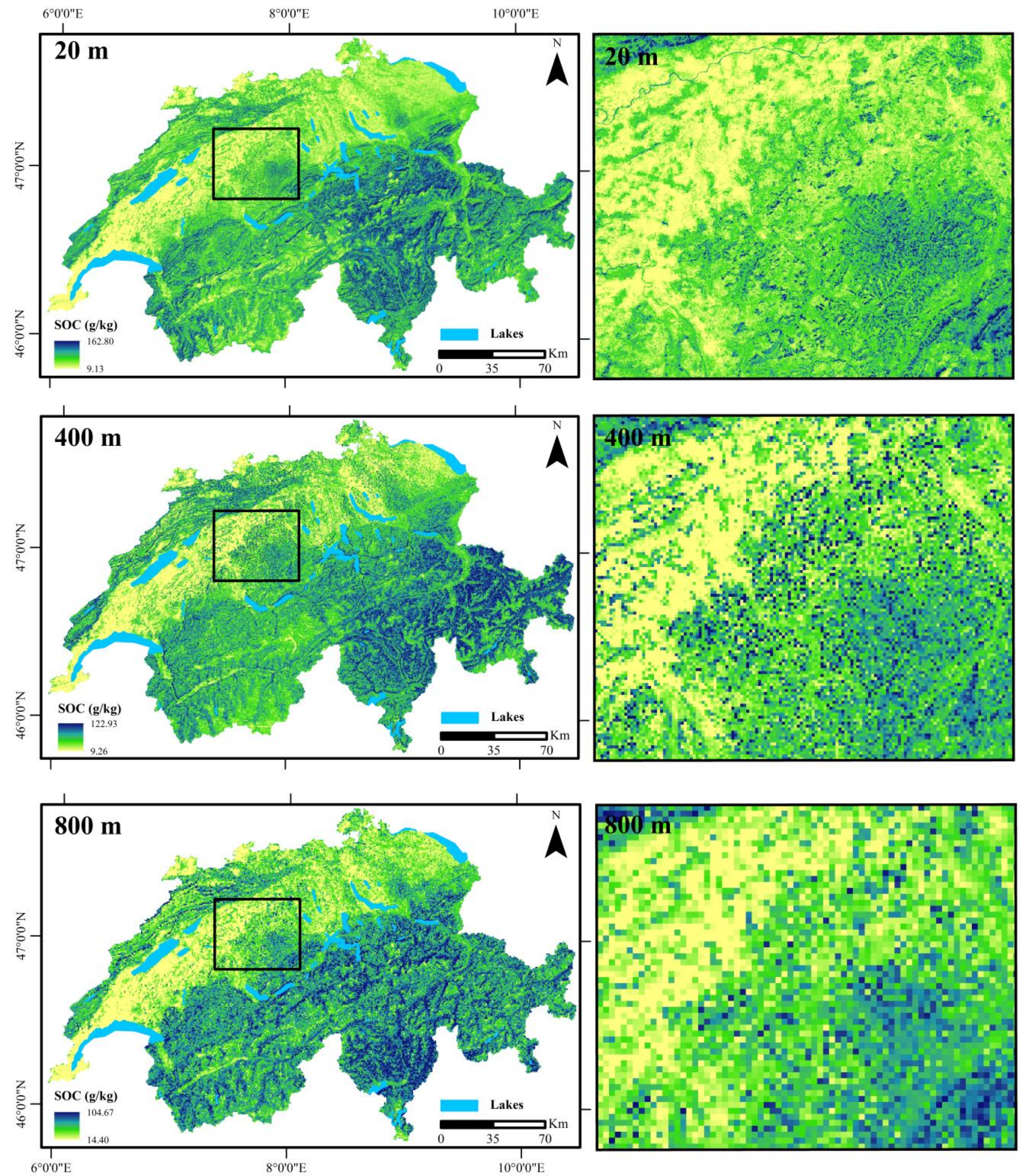

SOC
C:N ratio
SOC (g/kg)
160.50
9.44
Std. Dev. (g/kg)
9.84
0.06
C:N ratio
18.10
5.47
Std. Dev.
0.21
0.01
Lakes
Km
0 35 70
6°0'0"E
8°0'0"E
10°0'0"E
47°0'0"N
46°0'0"N
N

Fig. 4



6°0'0"E
8°0'0"E
10°0'0"E
47°0'0"N
46°0'0"N
20 m
N
SOC (g/kg)
162.80
9.13
Lakes
Km
0
35
70
20 m
400 m
N
SOC (g/kg)
122.93
9.26
Lakes
Km
0
35
70
400 m
800 m
N
SOC (g/kg)
104.67
14.40
Lakes
Km
0
35
70
800 m

Fig. 5

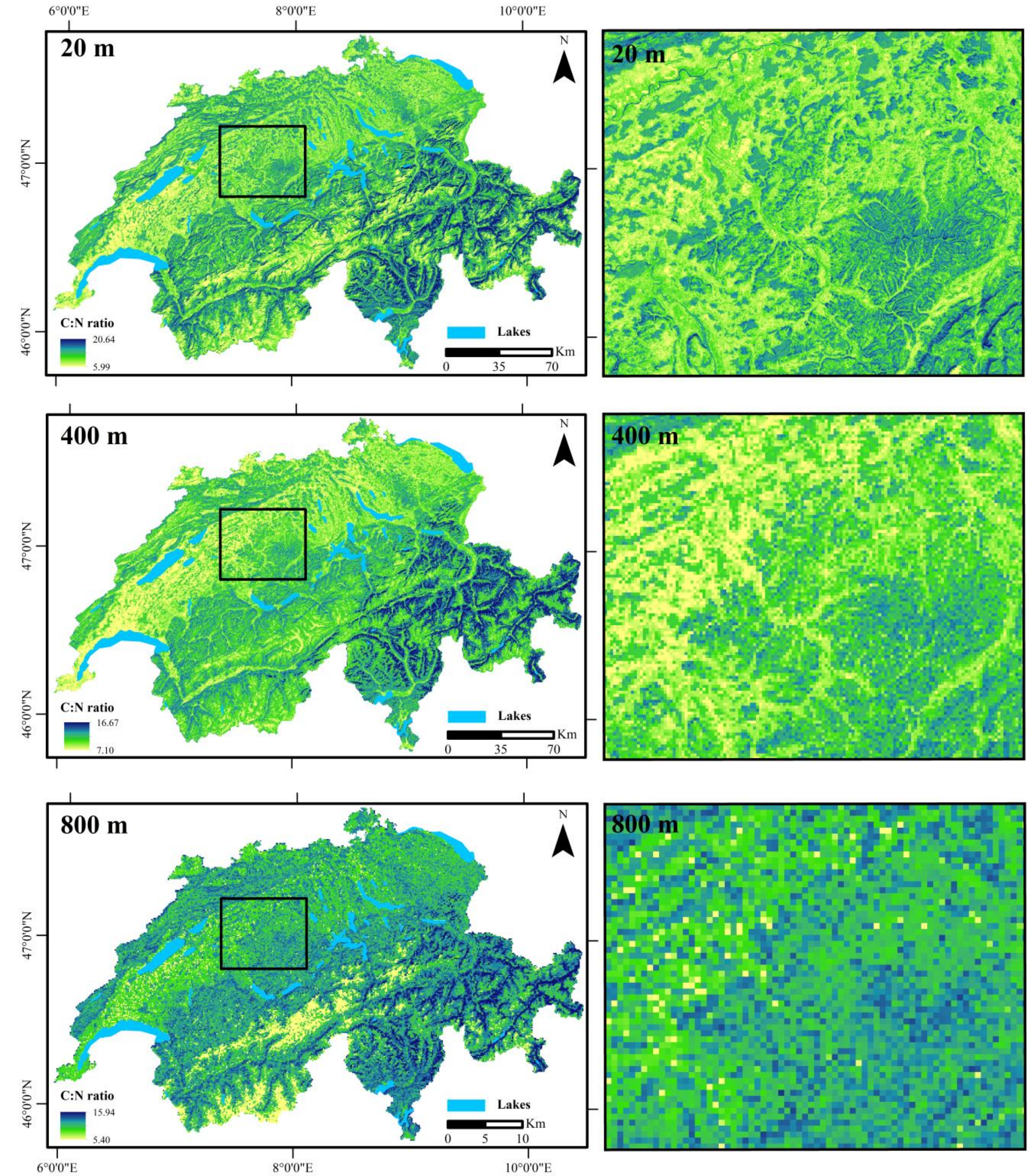

6°0'0"E
8°0'0"E
10°0'0"E
47°0'0"N
46°0'0"N
N
20 m
C:N ratio
20.64
5.99
Lakes
Km
0
35
70
400 m
C:N ratio
16.67
7.10
800 m
C:N ratio
15.94
5.40
5
10

Fig. 6