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

Wang, X., Geng, Y., **Zhou, T.**, Li, H., Liu, Y., Li, H., Ren, R., Zhang, Y., Xu, X., Liu, T., Si, B., **Lausch, A.** (2024):
Mapping the soil C:N ratio at the European scale by combining multi-year Sentinel radar and optical data via cloud computing
Soil Tillage Res. **245** , art. 106311

The publisher's version is available at:

https://doi.org/10.1016/j.still.2024.106311

Mapping the soil C:N ratio at the European scale by combining multi-year Sentinel radar and optical data via cloud computing

Xinyue Wang[a, b, 1], Yajun Geng[a, 1], Tao Zhou[a, c, d, *], Ying Zhao[a], Hongchen Li[a], Yanfang Liu[a], Huijie Li[a], Ruiqi Ren[a], Yazhou Zhang[a], Xiangrui Xu[e], Tingting Liu[f], Bingcheng Si[a, *] and Angela Lausch[c, d]

a Ludong University, School of Resources and Environmental Engineering, Middle Hongqi Road 186, 264025 Yantai, China

b Ludong University, School of Hydraulic Engineering, Middle Hongqi Road 186, 264025 Yantai, China

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

d Helmholtz Centre for Environmental Research, Department of Computational Landscape Ecology, Permoserstraße 15, 04318 Leipzig, Germany

e Zhejiang University City College, School of Spatial Planning and Design, Huzhou Street 51, 31000 Hangzhou, China

f University of Göttingen, Institute of Applied Plant Nutrition, Carl-Sprengel-Weg 1, 37075 Göttingen, Germany

* Corresponding author.

E-mail address: xinyangzhoutao@163.com (T. Zhou); bing.si@usask.ca (B. Si).

[1] These authors contributed equally to this paper.

Abstract:

Spatial information on the soil carbon-to-nitrogen (C:N) ratio is essential for sustainable soil use and management. The unprecedented availability of Sentinel optical and radar data on cloud computing platforms, such as the Google Earth Engine (GEE), has created new possibilities for developing soil prediction models from the local scale to the planetary scale. However, there is a paucity of literature on the effects of Sentinel sensor selection and integration and radar data utilization strategies on mapping the C:N ratio. In this study, we explored the use of multiyear Sentinel-1 radar and Sentinel-2 optical data obtained from the GEE platform combined with the digital soil mapping (DSM) technique to map the soil C:N ratio at the European scale. The performance of soil prediction models, which were constructed using two modeling techniques (random forest and support vector machine), derived under multiple scenarios based on optical, radar and commonly used auxiliary data (climatic and topographic variables) combined with the LUCAS 2018 soil dataset, was evaluated by a cross-validation technique. The results showed that the modeling performance varied with the selection and integration of Sentinel observations, as well as the configuration of the radar data. Models based on single polarization performed the worst across all scenarios related to Sentinel-1, with cross-polarization performing better than copolarization. Models that utilized Sentinel-1 data from ascending orbits outperformed those that utilized data from descending orbits. The application of Sentinel-1 backscatter information derived from different orbits and polarization modes resulted in improved prediction accuracy. Our study also demonstrated the potential of integrating multiyear Sentinel satellite data via the GEE to map the continental-scale C:N ratio. The model based on Sentinel-1 data

outperformed the one built on Sentinel-2 data, whereas combining Sentinel-2 optical data with Sentinel-1 radar data led to more accurate predictions. The variable importance results indicated that optical data and backscattering information from Sentinel observations are the most important groups of variables for soil C:N ratio mapping compared to the other variable groups (terrain and climate data). The digital soil maps generated under the different scenarios exhibited detailed patterns with significant spatial variation, with similar overall trends but slightly different details.

Keywords: Soil C:N ratio; Cloud computing; Sentinel; Soil mapping

1. Introduction

Soil organic carbon (SOC), as one of the main indicators reflecting soil fertility and ecological functions, affects soil physical, chemical and biological processes and plays a crucial role in maintaining agricultural productivity and soil quality (El-Naggar et al., 2019; Pu et al., 2024; Wang et al., 2023c). The ratio of SOC to total nitrogen (hereafter referred to as the C:N ratio) could be used to measure the balance between these quantities in soils, which affects the crop yield and microbial activity and is one of the important properties of terrestrial ecosystems (Cong et al., 2012; Li et al., 2020a). Therefore, there has been notable interest in understanding the spatial variations and controlling factors of the C:N ratio for use in soil biogeochemical modeling and carbon and nitrogen nutrition balance assessment (Geng et al., 2024; Gholizadeh et al., 2023). Unfortunately, the existing conventional soil maps produced by conventional soil mapping methods are limited in terms of their spatial details and the accuracy of soil attributes (Yang et al., 2011) and most of these soil maps may not meet the

requirements of practical applications at a finer scale (Li et al., 2020b). The production of soil maps using conventional soil mapping methods is very laborious, time-consuming and expensive (Dharumarajan et al., 2019). Therefore, alternative methods are needed to refine conventional soil maps, ensuring detailed and accurate soil information (Taghizadeh-Mehrjardi et al., 2022).

Digital soil mapping (DSM) has been increasingly adopted as an alternative by soil scientists and environmentalists to meet the growing demand for spatial soil information (Duan et al., 2022; Odebiri et al., 2023). The main difference between DSM technique and conventional soil mapping methods is that the soil-landscape model in the latter is a qualitative model based on the expertise of soil surveyors, while the soil-landscape model in DSM is a quantitative model (Kempen et al., 2012). Many positive advances in the DSM technique and practice have been described in the extensive literature, providing better mapping options (Odebiri et al., 2023; Omrani et al., 2021). Many predictive models (e.g., statistical, geostatistical, hybrid, and machine learning approaches) have been used to link the relationships between soil observations and auxiliary variables under the DSM framework (Zhang et al., 2017). Since machine learning techniques are not conditioned to follow any statistical assumptions, their DSM predictions often appear to be more accurate than those from traditional models (Wadoux et al., 2020). Wadoux et al. (2020) reviewed the current use of machine learning techniques for DSM studies and found that their applications are transforming the way in which soil scientists produce soil maps. This literature review also showed that tree-based models (e.g., boosted regression tree and random forest (RF)) are the most popular machine learners for DSM applications. It is generally necessary to retrieve

environmental data from soil observation locations to represent soil formation factors in DSM application, such as topography, climate, and remote sensing data (Padarian et al., 2015). The DSM performance strongly depends on the input variables used to build the soil landscape model (Nussbaum et al., 2018). Therefore, it is critical to identify the main influencing environmental variables and develop effective prediction methods to generate accurate digital soil maps. Given the advantages of Earth observation (EO) datasets, which provide low-cost, reliable surface information and large coverage areas, they have become a promising data source for improving DSM (Luo et al., 2024).

Remote sensing data captured by optical, thermal, or radar imaging systems provide spatially exhaustive environmental covariates for DSM application (Lausch et al., 2020; Richer-de-Forges et al., 2023b). The most commonly used sensor in EO-based DSM research is the optical imaging system, although it is susceptible to cloudy and rainy weather conditions (Pouladi et al., 2023). Synthetic aperture radar (SAR) sensors, which provide the advantages of penetrating clouds and operating at night, have received limited attention in DSM research (Zhou et al., 2020a). As an important part of the new EO program, the Sentinel constellations launched by the European Space Agency (ESA) perform continuous observation missions (Pouladi et al., 2023), providing a continuously updated image stream for DSM research. The comprehensive spatiotemporal coverage and new technical capabilities of the Sentinel-1 radar and Sentinel-2 optical sensors and the potential synergistic utilization of Sentinel observations provide a unique opportunity for the development of digital soil maps. Within this context, scholars have recently evaluated the observation capabilities of Sentinel-1/2 missions in DSM research activities (Deodoro et al., 2023; Li et

al., 2023; Ma et al., 2021; Zhou et al., 2020a; Zhou et al., 2020b). While these previous efforts demonstrated the potential usability of these two sensors in DSM applications, most were too limited in their spatial range to scale up the work and exhibited limitations in building DSM models using dense Sentinel observations with different acquisition configurations.

The availability of SAR sensors with various acquisition modes has increased over the past decade and has significantly contributed to soil research (Liu et al., 2019). The rapid advancement of the SAR sensor technique provides both great opportunities and certain challenges in DSM. For example, soil scientists employing SAR data may be confused by the various possible data configurations, such as the frequency, polarization, orbital direction, and time window, as these configurations may affect the DSM results. The wavelength of the sensor (e.g., C-, X- or L-band) is an important parameter to consider when using SAR observations, as it determines how the radar signal interacts with the surface and is related to the signal penetration depth (McNairn and Shang, 2016). In addition to the wavelength, the polarization mode used greatly affects the features captured in SAR images and should be accounted for. The interaction of the SAR system with different ground objects is through the different polarizations that the sensor can measure (Santos et al., 2023). The properties of ground targets (such as soil and vegetation) were reported to affect amplitude magnitude over time depending on the polarization of the SAR system (Brunelli and Mancini, 2024). Different SAR systems possess varying polarization capabilities, and under various polarizations, a single scene may exhibit different backscattering mechanisms. Compared to dual-polarized and fully-polarized SAR data, single-polarized SAR data provides relatively

limited information about the characteristics of the target ground (Zhang et al., 2016). Multipolarization SAR systems can offer additional details on surface features due to the different and complementary echoes obtained from multiple polarizations (Gao et al., 2023). Additionally, to extract meaningful information from SAR data, it is essential for users to carefully select the appropriate orbital configuration (Wood et al., 2002). SAR satellites can observe the ground in opposite directions, collecting data during both ascending (south-to-north) and descending (north-to-south) passes (Yun et al., 2006). Images acquired from these opposite orbits contain different amounts of backscatter, a factor that has been considered in many previous studies based on SAR data (Sayedain et al., 2020; Wang et al., 2023a). As one of the crucial configurations for EO data acquisition, the number of acquisition dates of optical data has been reported to significantly impact satellite-based DSM applications (Song et al., 2024; Zhang et al., 2024). Within this context, many studies have focused on investigating the most appropriate SAR configurations for various SAR-based applications (Holobâcă et al., 2019; Irwin et al., 2018; Luo et al., 2022b; Qadir et al., 2023). However, there is a lack of consensus on the most appropriate SAR configuration for SAR-based DSM applications, and more thorough research is needed.

Advanced DSM technique combined with cloud computing platforms like Google Earth Engine (GEE) offers a novel opportunity for large-scale soil monitoring. In this study, we aimed to establish continental (European)-scale soil C:N ratio prediction models using multiyear Sentinel radar and optical observation data compiled by the GEE combined with the LUCAS 2018 soil dataset. In Europe, the LUCAS soil survey, carried out by the European Commission, provides the largest scalable soil dataset in Europe (Fernandez-Ugalde et al.,

2022b), and its latest data collection campaign allows the use of updated input datasets to estimate soil properties. The LUCAS soil dataset has been used in many DSM studies to produce updated maps of soil properties (Ballabio et al., 2016; Chen et al., 2023; Xiao et al., 2023). Specifically, the objectives were to (1) explore the impacts of Sentinel-1 data integration strategies on the prediction accuracy; (2) evaluate the capabilities of multiyear Sentinel optical and radar EO data, as well as the combination of Sentinel optical and SAR data, for mapping the soil C:N ratio; and (3) conduct digital mapping of the soil C:N ratio at the continental scale via open data and cloud computing.

2. Materials and methods

2.1. Soil data

The research region includes all European Union (EU) member states and the UK where the soil survey of the 2018 LUCAS (Land Use/Land Cover Area Frame Survey) was carried out (Fig. 1). The LUCAS soil survey is the only mechanism for uniform and regular collection of soil samples implemented across the EU, covering all major land cover types simultaneously. Since 2009, the survey has included the sampling and analysis of the essential characteristics of EU topsoil, and the latest LUCAS database can be traced back to 2022 (Bispo et al., 2021). In the LUCAS 2015 survey, 90% of the locations were retained, and the remaining 10% were replaced by new sampling points, including points above 1000 m in elevation, which were beyond the scope of the LUCAS 2009/2012 surveys (Jones et al., 2020). The LUCAS 2018 survey is the third update of the database, analyzing soil samples collected from 18,984 locations, of which 16,556 points were repeat visits to sites previously sampled in 2015

(Fernandez-Ugalde et al., 2022a). In the 2018 LUCAS soil survey, samples were collected from all land use types (Table S1), with 7429 samples from cropland (constituting 39.1% of the total samples), 3989 from grassland (21.0%), 6092 from woodland (32.1%), 638 from bareland (3.4%), 720 from shrubland (3.8%), 71 from artificial land (0.4%), and 40 from wetland (0.2%). The LUCAS Topsoil data are considered the most detailed and harmonized soil database on a European scale, recording various basic soil properties, and the data have been adopted in many DSM studies (Kebonye et al., 2023; Xiao et al., 2023). The sampling protocol and laboratory analyses of the LUCAS survey have been described in detail by Bispo et al. (2021). In this study, all soil samples in the LUCAS 2018 topsoil dataset were used to build soil prediction models.

2.2. Environmental data

To explore the use of Sentinel radar and optical data combined with DSM technique to map the soil C:N ratio at the European scale, and to further improve model performance by incorporating commonly used environmental variables into the models built from Sentinel observations, this study collected environmental data from the following two pools: commonly used variables (climatic and topographic variables) and multiyear Sentinel EO data. These environmental data were obtained from the GEE platform, and the approach described by Poggio and Gimona (2017) was used to homogenize the resolutions to the selected pixel size (250 m) for soil property modeling. The selected pixel size is consistent with the digital soil map previously generated at the European scale using LUCAS soil data (Ballabio et al., 2019; Lu et al., 2023), serving as a compromise between coarse and

high-resolution auxiliary variables. All environmental variables and the LUCAS 2018 soil dataset were integrated into a geographic information system utilizing the ETRS89-LAEA projection (EPSG: 3035) for model building.

2.2.1. Climatic and topographic indicators

In this study, we used the Copernicus GLO-30 DEM with a resolution of 30 m stored in the GEE to prepare terrain attributes. Derived from the TanDEM-X mission, the Copernicus DEM is an improved version of WorldDEM™ published by the ESA (Moges et al., 2023). Nine terrain indicators were calculated using SAGA GIS software based on the Copernicus GLO-30 DEM covering the study area: elevation, plan curvature (PLC), profile curvature (PRC), length-slope factor (LSF), channel network base level (CNBL), analytical hillshading (AH), relative slope position (RSP), topographic position index (TPI) and channel network distance (CND). Mean annual temperature (MAT) and mean annual precipitation (MAP) data were collected from the WorldClim database with a spatial resolution of 1 km on the GEE platform as climate indicators (Fick and Hijmans, 2017).

2.2.2. Sentinel-1

Sentinel-1, consisting of two satellites, is a C-band (5.4 GHz) SAR mission launched by the ESA (Liang et al., 2021). Sentinel-1 provides four operating modes and multipolarization imaging capabilities, with a revisit cycle of 12 days using one satellite and a shortened revisit cycle of 6 days using two satellites (Bao et al., 2018). The interferometric wide swath mode (IW) exhibits high spatial resolution and wide coverage and is the main operating mode on

land (Nagler et al., 2016). We accessed and processed IW-mode Sentinel-1 ground range-detected (GRD) images on the GEE platform. The GEE platform preprocesses Sentinel-1 data using tools available in the Sentinel-1 SNAP software package to produce backscatter coefficients in each pixel, as described by Nagler et al. (2016). We filtered the Sentinel-1 data based on the properties, namely, acquisition time, orbital direction and polarization mode. Many GEE-based studies have demonstrated that the median composite is a feasible and efficient option to obtain optimal pixels in time-series images (Griffiths et al., 2019; Luo et al., 2023a). The Sentinel-1 data were collected over both single-year (2018) and multi-year (from the availability of Sentinel-1 on GEE up to 2018) time windows (Fig. 2). Sentinel-1 imagery collected during these periods was utilized to generate median composite images at five different time intervals: from one-month to four-month intervals and a six-month interval. Recently, some scholars have used multi-year optical images to generate median composite images at different time intervals in the GEE platform for DSM studies, and found that the prediction accuracy of composite images at different time intervals for soil organic matter varies greatly (Luo et al., 2022b; Song et al., 2024; Zhang et al., 2024). Luo et al. (2022a) conducted a similar study and found that the more years of synthetic images used, the more robust the prediction results. We tested the prediction accuracy of composite images with five different time intervals in single-year and multi-year time windows using all backscatter observations under VV and VH polarizations to select the optimal image synthesis method to obtain better prediction results. Backscatter observations under VV and VH polarizations and extracted SAR indices using backscatter data, including the ratio (VV/VH), difference (VV − VH), sum (VV + VH) and SAR Normalized Difference

Vegetation Index ((VV − VH) / (VV + VH)) (Hu et al., 2021), were used for model building (Table 1, Table S2).

2.2.3. Sentinel-2

Sentinel-2 consists of twin satellites launched by the ESA, with Sentinel-2A launched on June 23, 2015, and Sentinel-2B on March 7, 2017 (Filippucci et al., 2022; Yilmaz et al., 2023). Sentinel-2 is a multispectral imaging mission equipped with multispectral instrument (MSI) sensors, with 13 spectral bands, a spatial resolution of 10-60 m, and a revisit time of 5 days (two satellites) (He et al., 2022; Kumar et al., 2022). The QA60 image band provides information on the coverage of opaque and cirrus clouds and was used for cloud masking (Thorp and Drajat, 2021; Vergni et al., 2021). Similar to the study by Poggio and Gimona (2017) utilizing optical EO data for DSM research, to ensure the composite images contain limited snow-covered areas, all Sentinel-2 surface reflectance images with less than 10% cloud cover from June to October were collected on the GEE platform. This collection, similar to Sentinel-1, was performed within single-year and multi-year time windows (Fig. 2). All Sentinel-2 images acquired during these two time windows were used to generate median composite images by calculating the median of colocated pixels. Similar to our analysis with Sentinel-1, we also tested and compared the prediction accuracy of Sentinel-2 synthetic images in single-year and multi-year time windows. Synthetic optical images from ten Sentinel-2 channels (i.e., bands 2-8a, 11, and 12) were exported from the GEE for modeling analysis (Table 1).

2.3. Scenarios based on different input variables

We designed eleven modeling scenarios using widely adopted variables and multiyear Sentinel EO data to assess the potential and applicability of different data integration strategies for soil C:N ratio mapping at the European scale. The details of the scenarios formed for the different input variables are provided in Table 2.

2.4. Predictor selection

Multicollinearity resulting from high intercorrelations between predictor variables is a common problem in regression modeling that may negatively impact model performance (Arogoundade et al., 2023). Variable selection techniques are ideal to solve this problem to some extent by eliminating redundant variables (Chen et al., 2022b). We used recursive feature elimination (RFE) technique to remove irrelevant input features without losing model performance, selecting the optimal set of predictors for each scenario. The RFE algorithm is a wrapper-type feature ranking technique that operates by searching for a subset of features in the original dataset and eliminating redundant features (Yang et al., 2022). This algorithm begins by fitting a model using all input variables, evaluating both model performance and variable importance (Xiao et al., 2022). In subsequent iterations, the least important predictor is eliminated from the model, and performance metrics are recalculated (Kasraei et al., 2024). This process is repeated to select the most important predictors and determine the subset of environmental variables that yield the best performance (Chen et al., 2021). The RFE algorithm is one of the variable selection techniques commonly used in DSM studies to remove variables with multicollinearity and redundancy (Hounkpatin et al., 2022; Mashalaba et al., 2020; Zhang et al., 2023c). This study used the “caret” package in R to implement the

RFE algorithm to determine the optimal set of predictors for each modeling scenario.

2.5. Predictive models

2.5.1 Random forest

The RF model is an integrated system based on a decision tree developed by Breiman (2001), which can be used to perform classification and regression tasks. A vast number of trees are constructed in the RF model; each regression tree is derived from the training dataset using the unique bootstrap sample, and multiple regression trees are combined to form a forest; the results of all decision trees are averaged as the output of the model (Castaldi et al., 2019; Nguyen et al., 2022). The RF model has been widely used to solve regression problems in DSM (Guo et al., 2020; Stumpf et al., 2018). Two main parameters must be defined by the user in RF model construction: the number of random variables selected in each tree (mtry) and the number of trees in the forest (ntree) (Deng et al., 2018; Hong et al., 2018a). To optimize these parameters, we used the grid search method of the “caret” package in R software combined with ten-fold cross-validation based on the entire original dataset to determine the optimal values of the parameters, following the method of hyperparameter optimization and subsequent validation strategy of previous DSM studies (Wang et al., 2019; Yang et al., 2015). The RF model was constructed using the “randomForest” package in R software.

2.5.2 Support vector machine

Support vector machine (SVM) is a technique that uses statistical theory for classification and

regression analysis (Hasanipanah et al., 2015). The SVM algorithm is a very flexible machine learning technique with high level of model performance and has been widely used in DSM studies (Esfandiarpour-Boroujeni et al., 2020; Wu et al., 2018). This algorithm uses the kernel function to map the dataset matrix into a high-dimensional feature space and construct a hyperplane within it for classification or regression (Hong et al., 2018b; Tian et al., 2018). The radial basis function (RBF) kernel, which has excellent performance in capturing nonlinear relationships and is a commonly used kernel function in DSM studies (Wang et al., 2018; Zhou et al., 2020b), was selected as the kernel model of SVM. SVM modeling involves adjusting two parameters: kernel width and penalty. Similar to the RF algorithm, the optimization of these parameters was performed by the grid search method of the “caret” package in R software. The SVM model was built using the “kernlab” package in the R software. More information about the SVM algorithm is provided by Zendehboudi et al. (2018).

2.6. Model assessment

The prediction quality of the models built by the two machine learners with the best hyperparameters was evaluated using a ten-fold cross-validation technique, similar to the hyperparameter optimization and validation strategy adopted in previous soil mapping studies (Liu et al., 2022a; Yang et al., 2015). The entire original dataset was randomly divided into equal-size subsets; one group was selected for validation, and the other nine groups were used for model training; the process was repeated 10 times, and the validation results were aggregated and averaged to determine the final model performance (Vabalas et al., 2019). The

cross-validation technique has been reported as one of the most commonly used validation methods to evaluate the accuracy of DSM results (Wadoux et al., 2020). We calculated three common performance metrics to evaluate the prediction results: the coefficient of determination (R^2), mean absolute error (MAE) and root mean square error (RMSE).

3. Results

3.1. Comparison of the accuracy results for the different input datasets

We first tested the accuracy results of Sentinel-1 synthetic images with five different time intervals in single-year and multi-year time windows using all backscatter observations under VV and VH polarizations (Table S3), and compared the accuracy results of Sentinel-2 synthetic images in these two time windows (Table S4). The results indicated that the modeling results of Sentinel-1/2 composite images in multi-year time windows were better than those in single-year time windows. The accuracy ranking of composite images based on five different time intervals in these two time windows was as follows: one-month interval > two-month interval > three-month interval > four-month interval > six-month interval. Therefore, Sentinel-2 composite images under a multi-year time window and Sentinel-1 composite images with a one-month interval under this time window were selected as inputs to build the subsequent soil prediction models. The accuracy results of predicting the soil C:N ratio using eleven scenarios are summarized in Table 3. Our results showed that the selection of satellite sensors and predictive models, as well as the utilization strategy of SAR observations, impacted the model performance. Overall, for the two machine learners used, the prediction accuracy levels of the SVM and RF models were similar in most scenarios,

while the latter had a relatively higher prediction performance in the scenario 10 built from terrain and climate data. The prediction accuracy rankings of the scenarios built with different data integration strategies of Sentinel-1/2 were similar under these two modeling techniques. For instance, models trained on cross-polarized (VH) Sentinel-1 data outperformed those trained on co-polarized (VV) data under both SVM and RF algorithms; the synergistic utilization of polarization modes (SC5) improved the model performance over the utilization of a single polarization mode (SC1 and SC2). Moreover, under the two modeling techniques used, the models (SC5) with Sentinel-1 observations from the ascending and descending orbits achieved a higher accuracy than that based on single-orbital data (SC3 and SC4); the model based on ascending orbit data (SC3) from Sentinel-1 exhibited higher R^2 and lower RMSE and MAE values than that based on descending orbit data (SC4). We observed that the model (SC5) with backscattering coefficients exhibited a comparable predictive performance to that (SC6) with SAR indices; these two models showed a very close performance to the model (SC7) trained jointly by backscattering coefficients and SAR indices.

The models built from all Sentinel-2-derived predictors (SC8) performed worse than those built involving all SAR indices (SC6) or backscatter values (SC5) from Sentinel-1. The predictive performance of models (SC9) constructed using combined Sentinel SAR and optical data surpassed that of models built with data from a single Sentinel sensor alone, with both RF and SVM algorithms showing a 13% improvement in R^2 compared to models that solely utilized Sentinel-2 data as input (SC8). This study used the most commonly used auxiliary variables (terrain and climate) in DSM as additional input variables to further improve the performance of our soil prediction models. When using the RF algorithm, the

predictive accuracy (R^2 = 0.49, RMSE = 4.93, and MAE = 2.83) of the model established with terrain and climate variables (SC10) was higher than that of the model established with Sentinel-2 data (SC8), but lower than the predictive accuracy of the model trained with either Sentinel-1 data alone (SC5, SC6, and SC7) or the combined use of Sentinel optical and SAR data (SC9). Among the soil prediction models built with the SVM algorithm, the model (R^2 = 0.44, RMSE = 5.25, and MAE = 2.96) that utilized topographic and climatic variables exhibited the poorest performance, with a predictive accuracy lower than any model involving Sentinel-1 or Sentinel-2 data. When climate and terrain data were added to the model built using a combination of Sentinel optical and radar data, the prediction accuracy for the C:N ratio slightly improved (SC11), with R^2 increasing from 0.54 to 0.59 for the RF algorithm and from 0.55 to 0.58 for the SVM algorithm. As the most frequently used performance indicator in DSM, DSM studies often use R^2 expressed as the percentage of explained variance to represent predictive performance (Chen et al., 2022a). The R^2 values of the models built from Sentinel-1/2 (SC9) and auxiliary data (SC10) indicated that they could explain approximately 54% and 49% of the total C:N ratio variability, respectively, and could jointly (SC11) explained 59% of the total C:N ratio variation.

3.2. Variable importance assessment

The importance scores of the top twenty most important predictors for C:N ratio modeling under the different scenarios are shown in Fig. 3. This study classified all used input variables into four categories (i.e., climate variables, terrain attributes, Sentinel-1 and Sentinel-2 data) and calculated the total importance of each group. Under the experimental scenario with all

environmental data, the Sentinel-1 data received the highest sum of relative importance (48%), followed by climate variables (34%), terrain attributes (11%) and Sentinel-2 data (7%), indicating the varying impacts of these variable groups in the soil prediction model. MAT and MAP were the most important variables in this model built with all variables, and they were among the top two variables in models built from climate and terrain data. Climatic variables were more important inputs than terrain attributes in predicting C:N ratios in the two models involving topography and climate. Under the scenarios involving Sentinel-1 data, SAR indices were better predictors of the C:N ratio than backscatter data; VH polarization scored higher than VV polarization, further confirming the superiority of cross-polarization in C:N ratio modeling. When Sentinel-1 and Sentinel-2 data were applied synergistically, Sentinel data exerted a greater impact on C:N ratio mapping than climate or terrain data, which confirms that Sentinel observations are useful for continental-scale C:N ratio modeling.

3.3. Spatial prediction of the soil C:N ratio

The mapping results of the soil C:N ratio in Europe generated based on the different scenarios are shown in Figs. 4-5. The predicted maps exhibited detailed patterns with significant spatial variation, revealing a high degree of heterogeneity in soil C:N across the geographic space. Although the predictive accuracy of the scenarios with different input variables varied, the spatial patterns of mapping results generated by the models based on these scenarios were basically similar, with slight differences in details. The mean and standard deviation values of the soil C:N ratio predicted by scenario 11, which performed the

best among all scenarios, were 13.03 and 4.34 for RF (ranging from 4.03 to 57.49), and 12.61 and 4.43 for SVM (ranging from 5.29 to 39.06), respectively. We found that the C:N ratio generally showed some strong correlations with natural conditions, with higher levels in northern areas and in areas with higher rainfall.

4. Discussion

4.1. Analysis of the modeling performance for the different input variables

Analysis of the accuracy results indicated that the selection and integration of Sentinel observations, as well as the utilization strategy of radar satellite observations, greatly impacted the modeling performance. The model built with a single polarization mode performed the worst under all scenarios related to Sentinel-1. Among the various options, VH polarization showed a better modeling performance than VV polarization, and the combined use of both polarization modes significantly improved the prediction accuracy (Table 3). DSM technique facilitates the effective prediction of soil attributes by exploiting the valuable information provided by EO data on the controlling factors of soil formation and changes (Pouladi et al., 2023). The dominance of vegetation cover on the Earth's land surface affects the direct detection of soil by satellite sensors (Richer-de-Forges et al., 2023a). Yang and Guo (2019) reported that time-series SAR data can capture potential relationships between soil properties and vegetation, and such captured relationships aid in utilizing EO technique for predicting soil properties. They also found that the ability of radar satellite imagery to predict soil properties depends on how accurately it captures the temporal dynamics of vegetation. The interaction of SAR systems with vegetation is volumetric in nature, and the sensitivity of

SAR backscatter to vegetation dynamics depends on SAR sensor parameters such as polarization and frequency (Patel et al., 2006). Numerous studies have demonstrated that cross-polarization (e.g., VH polarization) backscatter is more sensitive to vegetation dynamics than co-polarization (e.g., VV polarization) backscatter (Cougo et al., 2015; Vreugdenhil et al., 2020; Wu et al., 2011). Compared to co-polarized SAR signals, cross-polarized SAR signals may be more sensitive to vegetation volume scattering, whereas co-polarized signals primarily reflect surface scattering characteristics (Singh et al., 2023; Urbazaev et al., 2015). The features captured in SAR images are strongly affected by the selected polarization mode (Mishra et al., 2016). Single-polarization SAR observations contain less information than multipolarization SAR observations, while the integration of VV and VH polarization modes provides more information on the target scattering mechanism (Chen et al., 2020). Co-polarized images are highly sensitive to surface properties, while cross-polarized images are directly related to canopy volume scattering intensity (Singh et al., 2023), thereby providing supplementary information to co-polarized images. In this study, we further evaluated the impact of Sentinel-1 images retrieved from different orbits on the prediction accuracy. Our results showed that the Sentinel-1-based models with ascending orbit data performed better than those with descending orbit data; the application of information obtained from different orbits could lead to more accurate predictions than using single-orbital images alone. Wood et al. (2002) reported that the differences between image pairs from ascending and descending orbits were attributed to the presence of dew and variations in the incidence angle. In descending orbit mode, Sentinel-1 images are captured in the morning when the water content of vegetation or bare soil is higher than at other times of

the day (Qadir et al., 2024). Data from ascending orbits are less affected by moisture than data from descending orbits, and the moisture levels of their ground targets are more stable and lower (Sayedain et al., 2020), resulting in more consistent backscatter signals that may help improve the stability of soil prediction models. Images captured from different orbits reflect the ground from opposite directions and contain different amounts of backscatter for specific areas (Sayedain et al., 2020). Cross-orbit satellite data could provide information missing from the line of sight of sensors along opposite orbits (Mahdavi et al., 2017). The idea of integrating SAR observations from ascending and descending orbits has been applied in various fields such as agriculture, deformation monitoring, and land use mapping (Fu et al., 2024; Qadir et al., 2023). While the differences in backscattering between images acquired along opposite orbits have been highlighted, the impact of orbits on modeling large-scale C:N ratios has received less attention. This study is the first attempt to assess the impact of different SAR configurations on mapping the continental-scale C:N ratio using multiyear Sentinel-1 observations on the GEE.

Our results demonstrated the potential of integrating multiyear radar EO data via the GEE to model large-scale C:N ratios and revealed that the models based on SAR observations could achieve a higher prediction accuracy than those based on optical observations (Table 3). Current satellite-based DSM research heavily relies on optical imaging systems, although they are susceptible to clouds and rain. SAR images, where cloud cover is less problematic, have not been widely adopted in DSM partly due to their complexity, diversity, and availability (Zhou et al., 2020b). Previous DSM research mainly focused on building models based on a single sensor such as Landsat or MODIS (Zhang et al.,

2023b). In this study, the combination of Sentinel optical-radar data produced better models and more accurate predictions, with R^2 improved by 13% over the use of optical Sentinel-2 data as input (Table 3). The integration of Sentinel optical-SAR observations has great potential to overcome certain issues, especially in areas prone to cloud cover. The comprehensive spatiotemporal coverage of Sentinel optical and radar sensors undoubtedly provides a unique opportunity to explore the benefits of synergistically utilizing Sentinel observations in DSM. In some studies, the potential of combining optical and radar data for improving DSM models has been reported (Azizi et al., 2023; Zhou et al., 2020b). However, these studies mainly only entailed the application of single-date or multiple-date images to build DSM models, ignoring the advantages of long-term satellite observations in soil prediction, as confirmed by Ma et al. (2023). Our results highlighted the advantages of the GEE platform in integrating dense Sentinel observations to build DSM models more efficiently.

Different DSM studies have utilized composite images obtained at varying time intervals, including those spanning multiple years (Poggio and Gimona, 2017; Poppiel et al., 2021; Wang et al., 2023b; Zhang et al., 2023a), within a single year (Liu et al., 2023), or from multiple specific dates (Wang et al., 2022). Our results indicated that the predictions based on Sentinel-1/2 composite images over a multi-year time window were superior to those obtained from a single-year time window. This result is consistent with a previous GEE-based DSM study, which found that multi-year composite images can provide more robust prediction results (Luo et al., 2022a). The study attributed this to the ability of multi-year median composite images to better mitigate the influence of extreme outliers. Research on

crop mapping based on GEE typically uses images from the same year for synthesis, and the shorter the time interval, the better the mapping effect (Luo et al., 2023b). Soil is more stable than crops (Luo et al., 2023b), and many GEE-based DSM studies have used multi-year composite images to obtain more useful pixels (Poggio and Gimona, 2017; Zhang et al., 2021). The results showed that the GEE platform could effectively support the development of DSM models due to its notable data integration and computing capabilities. This is consistent with the results of existing GEE-based DSM studies (Song et al., 2023; Zhang et al., 2021). However, optical sensors are most commonly used in existing efforts, while there are few studies based on SAR technology or combined with optical data. Pérez-Cutillas et al. (2023) reviewed the latest progress in GEE studies and found that the GEE provides a wide range of applications, with water resource applications as the most frequent, followed by land use and natural disaster applications, while the number of soil-related studies is relatively small. In this review, Landsat EO data were reported to be most widely used in GEE studies, with Landsat-8 as the most commonly used satellite, while the use of the Sentinel satellites has significantly increased. The workflow described in this study provides a promising path for integrating the data and computational potential inherent in the GEE to model large-scale soil properties. With the rapid development of the GEE and the increased participation of application experts, the GEE will provide more practical and powerful functions to help solve the problems faced by the DSM community.

4.2. Variable importance analysis

The successful application of Sentinel satellite observations in predicting soil carbon and

nitrogen has been demonstrated in many studies on different scales (Liu et al., 2022b; Xu et al., 2022). However, most studies on this topic mainly focus on the feasibility of Sentinel optical data, especially Sentinel-2, in predicting soil carbon and nitrogen. In this study, we illustrated that optical data and backscattering information from Sentinel observations are the most important groups of variables for predicting the C:N ratio. When utilized collaboratively, Sentinel optical and SAR data scored higher in importance than climate or terrain data (Fig. 3). Several studies also confirmed the usefulness of Sentinel-2's spectral bands for C:N ratio modeling (Geng et al., 2024; Zhou et al., 2021). The backscattering group achieved a higher sum of the importance scores than the optical predictors. This is supported by the results of Sothe et al. (2022), who explained that SAR images could better capture surface targets and are thus more likely to provide direct soil information than optical images. In all experiments involving Sentinel-1, SAR indices have been consistently identified as more important predictors than backscattered bands (Fig. 3). The DSM study involving SAR data of Domenech et al. (2020) also emphasized the importance of SAR indices in constructing DSM models. In this study, we revealed that the contribution of cross-polarization was greater than that of copolarization according to the variable importance. This finding was also obtained in another study using Sentinel-1 to map SOC (Zhou et al., 2022).

Climatic and topographic variables explained additional sources of variation in the C:N ratio estimates in this study. In the model incorporating all environmental data, the Sentinel-1 data exhibited the highest importance score, followed by climate data, terrain attributes, and Sentinel-2 data. In the models involving terrain and climate data, MAT was identified as the most important predictor, followed by MAP. Climate indirectly affects the soil C:N ratio by

controlling plant growth, biological activity, and litter accumulation and decomposition (Zhou et al., 2021). In many studies, a close relationship between temperature, precipitation and soil C:N ratio has been revealed (He et al., 2023; Jiao et al., 2016). Climate was the main parameter influencing carbon accumulation at the European continental scale (Rial et al., 2017). DEM derivatives were the third variable group in the variable importance ranking, revealing their moderate importance for predicting the C:N ratio in the study area (Fig. 3). DEM derivatives describing the shape of the land surface are the most commonly used input variables in DSM (Dharumarajan et al., 2019). Terrain plays a key role in several environmental processes, such as soil erosion and sedimentation, water infiltration and distribution, thereby affecting the soil C:N ratio. In our study, both groups of terrain and climate data enhanced the model performance by introducing information not available in the other groups.

4.3. Spatial analysis of soil C:N ratio maps predicted by the different data integration strategies

The spatial pattern of the soil C:N ratio predicted under the scenarios with different input variables in this study showed notable similarities with other map products produced using different mapping techniques (Matschullat et al., 2018). The maps produced by the different scenarios exhibited detailed patterns with significant spatial variation. The overall trend in the mapping results of the models under the different data integration strategies was consistent, but the details were slightly different (Figs. 4-5). Overall, the C:N ratio exhibited a strong correlation with natural conditions, although it was not uniform at the continental scale

(Matschullat et al., 2018). Higher C:N ratio levels were observed in northern areas and in areas with higher rainfall, consistent with a previous DSM study (Ballabio et al., 2019). It has been observed that high C:N ratio values are generally associated with low temperatures and heavy rainfall from regional to continental and global scales (He et al., 2023). The spatial distribution of the C:N ratio was significantly affected by land use, with coniferous trees and peatlands exhibiting higher C:N ratio levels. High plant litter inputs and lower average temperatures in boreal forests favor the accumulation of carbon in soil. Cotrufo et al. (2019) found that C:N ratio levels in forest and grassland topsoil in European coniferous and mixed forests were the highest, while those in broadleaf forests and pure grasslands were lower. Tree species were reported as the main factor explaining C:N ratios in European forest soils followed by biogeographical zoning (Cools et al., 2014). Some researchers have proposed that combinations of variables such as ecoregion, climate, topography, soil type, and parent material should be investigated when evaluating factors affecting C:N ratios (He et al., 2023; Ostrowska and Porębska, 2015).

5. Conclusions

This study explored the use of multiyear Sentinel radar and optical observation data obtained from the GEE platform combined with the DSM technique to model and map the soil C:N ratio at the continental scale. Multiple prediction scenarios were implemented based on different data integration strategies for optical, radar and commonly used auxiliary data. Our analysis yielded the following key conclusions:

- Our results indicated that the modeling performance varied depending on the selection

and integration of Sentinel observations, as well as the strategy for utilizing radar data.

- The model built with the single polarization mode performed the worst under all scenarios related to Sentinel-1, with cross-polarization performing better than copolarization, while the synergistic use of both polarization modes improved model performance.
- The application of information derived from different orbits led to more accurate predictions than utilizing single-orbit images, with models based on ascending orbit data performing better than those based on descending orbit data.
- In this study, we revealed the importance of integrating multiyear Sentinel observations via the GEE to model large-scale C:N ratios. The radar Sentinel-1-based model performed better than the optical Sentinel-2-based model, and the combined use of Sentinel optical and radar data resulted in more accurate predictions. Furthermore, the analysis of the importance showed that optical and backscatter information from these two sensors are the most important groups of variables for C:N ratio modeling.
- We observed that the maps produced under the different scenarios exhibited detailed patterns with significant spatial variation, with consistent overall trends but slightly different details. Furthermore, the results indicated that the GEE platform is a promising tool for developing DSM products.

Acknowledgments

The authors would like to thank the ESA and the European Soil Data Center for providing the data free of charge and the GEE platform for providing cloud computing capabilities. This

work was supported by the National Key R&D Program of China (2022YFE0100300), the National Natural Science Foundation of China (42101018), the Natural Science Foundation of Shandong Province (ZR2020QD119, ZR2023QD100 and ZR2020QD012), the Ludong University Talent Research Start-up Fund (20220035), and the Shandong Province Higher Education Institutions Youth Entrepreneurship Team Program (2022KJ289).

References

Arogoundade, A.M., Mutanga, O., Odindi, J., Odebiri, O., 2023. Leveraging Google Earth Engine to estimate foliar C: N ratio in an African savannah rangeland using Sentinel 2 data. Remote Sensing Applications: Society and Environment 30, 100981.

Azizi, K., Garosi, Y., Ayoubi, S., Tajik, S., 2023. Integration of Sentinel-1/2 and topographic attributes to predict the spatial distribution of soil texture fractions in some agricultural soils of western Iran. Soil and Tillage Research 229, 105681.

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., Monatanarella, L., 2016. Mapping topsoil physical properties at European scale using the LUCAS database. Geoderma 261, 110-123.

Bao, Y., Lin, L., Wu, S., Kwal Deng, K.A., Petropoulos, G.P., 2018. Surface soil moisture retrievals over partially vegetated areas from the synergy of Sentinel-1 and Landsat 8 data using a modified water-cloud model. International Journal of Applied Earth Observation and Geoinformation 72, 76-85.

Bispo, A., Arrouays, D., Saby, N., Boulonne, L., Fantappiè, M., 2021. Proposal of methodological development for the LUCAS programme in accordance with national monitoring programmes. Towards climate-smart sustainable management of agricultural soils EJP Soil, 135 p.

Breiman, L., 2001. Random forests. Machine learning 45, 5-32.

Brunelli, B., Mancini, F., 2024. Comparative analysis of SAOCOM and Sentinel-1 data for surface soil moisture retrieval using a change detection method in a semiarid region (Douro River's basin, Spain). International Journal of Applied Earth Observation and Geoinformation 129, 103874.

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, Q., Yang, H., Li, L., Liu, X., 2020. A Novel Statistical Texture Feature for SAR Building Damage Assessment in Different Polarization Modes. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13, 154-165.

Chen, S., Arrouays, D., Leatitia Mulder, V., Poggio, L., Minasny, B., Roudier, P., Libohova, Z., Lagacherie, P., Shi, Z., Hannam, J., Meersmans, J., Richer-de-Forges, A.C., Walter, C., 2022a. Digital mapping of GlobalSoilMap soil properties at a broad scale: A review. Geoderma 409, 115567.

Chen, S., Richer-de-Forges, A.C., Leatitia Mulder, V., Martelet, G., Loiseau, T., Lehmann, S., Arrouays, D., 2021. Digital mapping of the soil thickness of loess deposits over a calcareous bedrock in central France. Catena 198, 105062.

Chen, S., Saby, N.P.A., Martin, M.P., Barthès, B.G., Gomez, C., Shi, Z., Arrouays, D., 2023. Integrating additional spectroscopically inferred soil data improves the accuracy of digital soil mapping. Geoderma 433, 116467.

Chen, Y., Ma, L., Yu, D., Zhang, H., Feng, K., Wang, X., Song, J., 2022b. Comparison of feature selection methods for mapping soil organic matter in subtropical restored forests. Ecological Indicators 135, 108545.

Cong, R.H., Wang, X.J., Xu, M.G., Zhang, W.J., Xie, L.J., Yang, X.Y., Huang, S.M., Wang, B.R., 2012. Dynamics of soil carbon to nitrogen ratio changes under long-term fertilizer addition in wheat-corn double cropping systems of China. European Journal of Soil Science 63(3), 341-350.

Cools, N., Vesterdal, L., De Vos, B., Vanguelova, E., Hansen, K., 2014. Tree species is the major factor explaining C:N ratios in European forest soils. Forest Ecology and Management 311, 3-16.

Cotrufo, M.F., Ranalli, M.G., Haddix, M.L., Six, J., Lugato, E., 2019. Soil carbon storage informed by particulate and mineral-associated organic matter. Nature Geoscience 12(12), 989-994.

Cougo, M.F., Souza-Filho, P.W.M., Silva, A.Q., Fernandes, M.E.B., Santos, J.R.d., Abreu, M.R.S., Nascimento, W.R., Simard, M., 2015. Radarsat-2 Backscattering for the Modeling of Biophysical Parameters of Regenerating Mangrove Forests. Remote Sensing 7(12), 17097-17112.

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. Agriculture, Ecosystems & Environment 254, 213-223.

Deodoro, S.C., Moral, R.A., Fealy, R., McCarthy, T., Fealy, R., 2023. An assessment of Sentinel-1 synthetic aperture radar, geophysical and topographical covariates for estimating topsoil particle-size fractions. European Journal of Soil Science 74(5), e13414.

Dharumarajan, S., Hegde, R., Janani, N., Singh, S.K., 2019. The need for digital soil mapping in India. Geoderma Regional 16, e00204.

Domenech, M.B., Amiotti, N.M., Costa, J.L., Castro-Franco, M., 2020. Prediction of topsoil properties at field-scale by using C-band SAR data. International Journal of Applied Earth Observation and Geoinformation 93, 102197.

Duan, M., Song, X., Liu, X., Cui, D., Zhang, X., 2022. Mapping the soil types combining multi-temporal remote sensing data with texture features. Computers and Electronics in Agriculture 200, 107230.

El-Naggar, A., El-Naggar, A.H., Shaheen, S.M., Sarkar, B., Chang, S.X., Tsang, D.C.W., Rinklebe, J., Ok, Y.S., 2019. Biochar composition-dependent impacts on soil nutrient release, carbon mineralization, and potential environmental risk: A review. Journal of Environmental Management 241, 458-467.

Esfandiarpour-Boroujeni, I., Shahini-Shamsabadi, M., Shirani, H., Mosleh, Z., Bagheri-Bodaghabadi, M., Salehi, M.H., 2020. Assessment of different digital soil mapping methods for prediction of soil classes in the Shahrekord plain, Central Iran. Catena 193, 104648.

Fernandez-Ugalde, O., Scarpa, S., Orgiazzi, A., Panagos, P., Van Liedekerke, M., A., M., Jones, A., 2022a. LUCAS 2018 Soil Module. Presentation of dataset and results.

Fernandez-Ugalde, O., Scarpa, S., Orgiazzi, A., Panagos, P., Van Liedekerke, M., Marechal, A., Jones, A., 2022b. LUCAS 2018 Soil Module. Presentation of dataset and results, EUR 31144.

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.

Filippucci, P., Brocca, L., Bonafoni, S., Saltalippi, C., Wagner, W., Tarpanelli, A., 2022. Sentinel-2 high-resolution data for river discharge monitoring. Remote Sensing of Environment 281, 113255.

Fu, Z., Li, N., Guo, Z., 2024. Locating RFI Source by Integrating Ascending and Descending Passes Based on SAR Raw Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 17, 9376-9388.

Gao, G., Bai, Q., Zhang, C., Zhang, L., Yao, L., 2023. Dualistic cascade convolutional neural network dedicated to fully PolSAR image ship detection. ISPRS Journal of Photogrammetry and Remote Sensing 202, 663-681.

Geng, J., Tan, Q., Lv, J., Fang, H., 2024. Assessing spatial variations in soil organic carbon and C:N ratio in Northeast China's black soil region: Insights from Landsat-9 satellite and crop growth information. Soil and Tillage Research 235, 105897.

Gholizadeh, A., Saberioon, M., Pouladi, N., Ben-Dor, E., 2023. Quantification and depth distribution analysis of carbon to nitrogen ratio in forest soils using reflectance spectroscopy. International Soil and Water Conservation Research 11(1), 112-124.

Griffiths, P., Nendel, C., Hostert, P., 2019. Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. Remote Sensing of Environment 220, 135-151.

Guo, X., Bian, Z., Wang, S., Wang, Q., Zhang, Y., Zhou, J., Lin, L., 2020. Prediction of the spatial distribution of soil arthropods using a random forest model: A case study in Changtu County, Northeast China. Agriculture, Ecosystems & Environment 292, 106818.

Hasanipanah, M., Monjezi, M., Shahnazar, A., Jahed Armaghani, D., Farazmand, A., 2015. Feasibility of indirect determination of blast induced ground vibration based on support vector machine. Measurement 75, 289-297.

He, J., Chen, B., Xu, W., Xiang, C., Kuang, W., Zhao, X., 2023. Driving factors for soil C:N ratio in woody plant communities across northeastern Qinghai-Tibetan Plateau. Catena 233, 107504.

He, S., Shao, H., Xian, W., Yin, Z., You, M., Zhong, J., Qi, J., 2022. Monitoring Cropland Abandonment in Hilly Areas with Sentinel-1 and Sentinel-2 Timeseries. Remote Sensing 14(15).

Holobâcă, I.-H., Ivan, K., Alexe, M., 2019. Extracting built-up areas from Sentinel-1 imagery using land-cover classification and texture analysis. International Journal of Remote Sensing 40(20), 8054-8069.

Hong, H., Tsangaratos, P., Ilia, I., Liu, J., Zhu, A.X., Xu, C., 2018a. Applying genetic algorithms to set the optimal combination of forest fire related variables and model forest fire susceptibility based on data mining models. The case of Dayu County, China. Science of the Total Environment 630, 1044-1056.

Hong, Y., Chen, S., Zhang, Y., Chen, Y., Yu, L., Liu, Y., Liu, Y., Cheng, H., Liu, Y., 2018b. Rapid identification of soil organic matter level via visible and near-infrared spectroscopy: Effects of two-dimensional correlation coefficient and extreme learning machine. Science of the Total Environment 644, 1232-1243.

Hounkpatin, K.O.L., Bossa, A.Y., Yira, Y., Igue, M.A., Sinsin, B.A., 2022. Assessment of the soil fertility status in Benin (West Africa) – Digital soil mapping using machine learning. Geoderma Regional 28, e00444.

Hu, Y., Tian, B., Yuan, L., Li, X., Huang, Y., Shi, R., Jiang, X., Wang, l., Sun, C., 2021. Mapping coastal salt marshes in China using time series of Sentinel-1 SAR. ISPRS Journal of Photogrammetry and Remote Sensing 173, 122-134.

Irwin, K., Braun, A., Fotopoulos, G., Roth, A., Wessel, B., 2018. Assessing Single-Polarization and Dual-Polarization TerraSAR-X Data for Surface Water Monitoring. Remote Sensing 10(6), 949.

Jiao, F., Shi, X.-R., Han, F.-P., Yuan, Z.-Y., 2016. Increasing aridity, temperature and soil pH induce soil C-N-P imbalance in grasslands. Scientific Reports 6(1), 19601.

Jones, A., Fernandez-Ugalde, O., Scarpa, S., 2020. Lucas 2015 topsoil survey. Presentation of dataset and results, EUR 30332.

Kasraei, B., Schmidt, M.G., Zhang, J., Bulmer, C.E., Filatow, D.S., Arbor, A., Pennell, T., Heung, B., 2024. A framework for optimizing environmental covariates to support model interpretability in digital soil mapping. Geoderma 445, 116873.

Kebonye, N.M., Agyeman, P.C., Biney, J.K.M., 2023. Optimized modelling of countrywide soil organic carbon levels via an interpretable decision tree. Smart Agricultural Technology 3, 100106.

Kempen, B., Brus, D.J., Stoorvogel, J.J., Heuvelink, G.B.M., de Vries, F., 2012. Efficiency Comparison of Conventional and Digital Soil Mapping for Updating Soil Maps. Soil Science Society of America Journal 76(6), 2097-2115.

Kumar, H., Karwariya, S.K., Kumar, R., 2022. Google Earth Engine-Based Identification of Flood Extent and Flood-Affected Paddy Rice Fields Using Sentinel-2 MSI and Sentinel-1 SAR Data in Bihar State, India. Journal of the Indian Society of Remote Sensing 50(5), 791-803.

Lausch, A., Schaepman, M.E., Skidmore, A.K., Truckenbrodt, S.C., Hacker, J.M., Baade, J., Bannehr, L., Borg, E., Bumberger, J., Dietrich, P., Gläßer, C., Haase, D., Heurich, M., Jagdhuber, T., Jany, S., Krönert, R., Möller, M., Mollenhauer, H., Montzka, C., Pause, M., Rogass, C., Salepci, N., Schmullius, C., Schrodt, F., Schütze, C., Schweitzer, C., Selsam, P., Spengler, D., Vohland, M., Volk, M., Weber, U., Wellmann, T., Werban, U., Zacharias, S., Thiel, C., 2020. Linking the Remote Sensing of Geodiversity and Traits Relevant to Biodiversity—Part II: Geomorphology, Terrain and Surfaces. Remote Sensing 12(22), 3690.

Li, X., Shang, B., Wang, D., Wang, Z., Wen, X., Kang, Y., 2020a. Mapping soil organic carbon and total nitrogen in croplands of the Corn Belt of Northeast China based on geographically weighted regression kriging model. Computers & Geosciences 135, 104392.

Li, Y., Zhang, N., Li, R.-K., Liu, C.-Y., Shen, J., Jing, Y.-C., 2020b. Soil mapping based on assessment of environmental similarity and selection of calculating samples. Catena 188, 104379.

Li, Z., Liu, F., Peng, X., Hu, B., Song, X., 2023. Synergetic use of DEM derivatives, Sentinel-1 and Sentinel-2 data for mapping soil properties of a sloped cropland based on a two-step ensemble learning method. Science of the Total Environment 866, 161421.

Liang, J., Liang, G., Zhao, Y., Zhang, Y., 2021. A synergic method of Sentinel-1 and Sentinel-2 images for retrieving soil moisture content in agricultural regions. Computers and Electronics in Agriculture 190, 106485.

Liu, C.-a., Chen, Z.-x., Shao, Y., Chen, J.-s., Hasi, T., Pan, H.-z., 2019. Research advances of SAR remote sensing for agriculture applications: A review. Journal of Integrative Agriculture 18(3), 506-525.

Liu, F., Wu, H., Zhao, Y., Li, D., Yang, J.-L., Song, X., Shi, Z., Zhu, A.X., Zhang, G.-L., 2022a. Mapping high resolution National Soil Information Grids of China. Science Bulletin 67(3), 328-340.

Liu, Q., He, L., Guo, L., Wang, M., Deng, D., Lv, P., Wang, R., Jia, Z., Hu, Z., Wu, G., Shi, T., 2022b. Digital mapping of soil organic carbon density using newly developed bare soil spectral indices and deep neural network. Catena 219, 106603.

Liu, Y., Chen, S., Yu, Q., Cai, Z., Zhou, Q., Bellingrath-Kimura, S.D., Wu, W., 2023. Improving digital mapping of soil organic matter in cropland by incorporating crop rotation. Geoderma 438, 116620.

Lu, Q., Tian, S., Wei, L., 2023. Digital mapping of soil pH and carbonates at the European scale using environmental variables and machine learning. Science of the Total Environment 856, 159171.

Luo, C., Zhang, W., Zhang, X., Liu, H., 2023a. Mapping of soil organic matter in a typical black soil area using Landsat-8 synthetic images at different time periods. Catena 231, 107336.

Luo, C., Zhang, W., Zhang, X., Liu, H., 2023b. Mapping soil organic matter content using Sentinel-2 synthetic images at different time intervals in Northeast China. International Journal of Digital Earth 16(1), 1094-1107.

Luo, C., Zhang, W., Zhang, X., Liu, H., 2024. Mapping the soil organic matter content in a typical black-soil area using optical data, radar data and environmental covariates. Soil and Tillage Research 235, 105912.

Luo, C., Zhang, X., Meng, X., Zhu, H., Ni, C., Chen, M., Liu, H., 2022a. Regional mapping of soil organic matter content using multitemporal synthetic Landsat 8 images in Google Earth Engine. Catena 209, 105842.

Luo, C., Zhang, X., Wang, Y., Men, Z., Liu, H., 2022b. Regional soil organic matter mapping models based on the optimal time window, feature selection algorithm and Google Earth Engine. Soil and Tillage Research 219, 105325.

Ma, G., Ding, J., Han, L., Zhang, Z., Ran, S., 2021. Digital mapping of soil salinization based on Sentinel-1 and Sentinel-2 data combined with machine learning algorithms. Regional Sustainability 2(2), 177-188.

Ma, S., He, B., Ge, X., Luo, X., 2023. Spatial prediction of soil salinity based on the Google Earth Engine platform with multitemporal synthetic remote sensing images. Ecological Informatics 75, 102111.

Mahdavi, S., Salehi, B., Amani, M., Granger, J.E., Brisco, B., Huang, W., Hanson, A., 2017. Object-Based Classification of Wetlands in Newfoundland and Labrador Using Multi-Temporal PolSAR Data. Canadian Journal of Remote Sensing 43(5), 432-450.

Mashalaba, L., Galleguillos, M., Seguel, O., Poblete-Olivares, J., 2020. Predicting spatial variability of selected soil properties using digital soil mapping in a rainfed vineyard of central Chile. Geoderma Regional 22, e00289.

Matschullat, J., Reimann, C., Birke, M., dos Santos Carvalho, D., Albanese, S., Anderson, M., Baritz, R., Batista, M.J., Bel-Ian, A., Cicchella, D., Demetriades, A., De Vivo, B., De Vos, W., Dinelli, E., Ďuriš, M., Dusza-Dobek, A., Eggen, O.A., Eklund, M., Ernsten, V., Fabian, K., Filzmoser, P., Flight, D.M.A., Forrester, S., Fügedi, U., Gilucis, A., Gosar, M., Gregorauskiene, V., De Groot, W., Gulan, A., Halamić, J., Haslinger, E., Hayoz, P., Hoogewerff, J., Hrvatovic, H., Husnjak, S., Jähne-Klingberg, F., Janik, L., Jordan, G., Kaminari, M., Kirby, J., Klos, V., Kwećko, P., Kuti, L., Ladenberger, A., Lima, A., Locutura, J., Lucivjansky, P., Mann, A., Mackovych, D., McLaughlin, M., Malyuk, B.I., Maquil, R., Meuli, R.G., Mol, G., Négrel, P., O'Connor, P., Oorts, K., Ottesen, R.T., Pasnieczna, A., Petersell, V., Pfleiderer, S., Poňavič, M., Prazeres, C., Radusinović, S., Rauch, U., Sadeghi, M., Salpeteur, I., Scanlon, R., Schedl, A., Scheib, A., Schoeters, I., Šefčik, P., Sellersjö, E., Slaninka, I., Soriano-Disla, J.M., Šorša, A., Svrkota, R., Stafilov, T., Tarvainen, T., Tendavilov, V., Valera, P., Verougstraete, V., Vidojević, D., Zissimos, A., Zomeni, Z., 2018. GEMAS: CNS concentrations and C/N ratios in European agricultural soil. Science of The Total Environment 627, 975-984.

McNairn, H., Shang, J., 2016. A review of multitemporal synthetic aperture radar (SAR) for crop monitoring. Multitemporal Remote Sensing: Methods and Applications, 317-340.

Mishra, V.N., Prasad, R., Kumar, P., Gupta, D.K., Srivastava, P.K., 2016. Dual-polarimetric C-band SAR data for land use/land cover classification by incorporating textural information. Environmental Earth Sciences 76(1), 26.

Moges, D.M., Virro, H., Kmoch, A., Cibin, R., Rohith, A.N., Martínez-Salvador, A., Conesa-García, C., Uuemaa, E., 2023. How does the choice of DEMs affect catchment hydrological modeling? Science of the Total Environment 892, 164627.

Nagler, T., Rott, H., Ripper, E., Bippus, G., Hetzenecker, M., 2016. Advancements for Snowmelt Monitoring by Means of Sentinel-1 SAR. Remote Sensing 8(4), 348.

Nguyen, T.T., Pham, T.D., Nguyen, C.T., Delfos, J., Archibald, R., Dang, K.B., Hoang, N.B., Guo, W., Ngo, H.H., 2022. A novel intelligence approach based active and ensemble learning for agricultural soil organic carbon prediction using multispectral and SAR data fusion. Sci Total Environ 804, 150187.

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.

Odebiri, O., Mutanga, O., Odindi, J., Naicker, R., Slotow, R., Mngadi, M., 2023. Evaluation of projected soil organic carbon stocks under future climate and land cover changes in South Africa using a deep learning approach. Journal of Environmental Management 330, 117127.

Omrani, M., Shahbazi, F., Feizizadeh, B., Oustan, S., Najafi, N., 2021. Application of remote sensing indices to digital soil salt composition and ionic strength mapping in the east shore of Urmia Lake, Iran. Remote Sensing Applications: Society and Environment 22, 100498.

Ostrowska, A., Porębska, G., 2015. Assessment of the C/N ratio as an indicator of the decomposability of organic matter in forest soils. Ecological Indicators 49, 104-109.

Padarian, J., Minasny, B., McBratney, A.B., 2015. Using Google's cloud-based platform for digital soil mapping. Computers & Geosciences 83, 80-88.

Patel, P., Srivastava, H.S., Panigrahy, S., Parihar, J.S., 2006. Comparative evaluation of the sensitivity of multi-polarized multi-frequency SAR backscatter to plant density. International Journal of Remote Sensing 27(2), 293-305.

Pérez-Cutillas, P., Pérez-Navarro, A., Conesa-García, C., Zema, D.A., Amado-Álvarez, J.P., 2023. What is going on within google earth engine? A systematic review and meta-analysis. Remote Sensing Applications: Society and Environment 29, 100907.

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.

Poppiel, R.R., Demattê, J.A.M., Rosin, N.A., Campos, L.R., Tayebi, M., Bonfatti, B.R., Ayoubi, S., Tajik, S., Afshar, F.A., Jafari, A., Hamzehpour, N., Taghizadeh-Mehrjardi, R., Ostovari, Y., Asgari, N., Naimi, S., Nabiollahi, K., Fathizad, H., Zeraatpisheh, M., Javaheri, F., Doustaky, M., Naderi, M., Dehghani, S., Atash, S., Farshadirad, A., Mirzaee, S., Shahriari, A., Ghorbani, M., Rahmati, M., 2021. High resolution middle eastern soil attributes mapping via open data and cloud computing. Geoderma 385, 114890.

Pouladi, N., Gholizadeh, A., Khosravi, V., Borůvka, L., 2023. Digital mapping of soil organic carbon using remote sensing data: A systematic review. Catena 232, 107409.

Pu, Y., Yang, L., Zhang, L., Huang, H., Zhang, G., Zhou, C., 2024. Major contributions of agricultural management practices to topsoil organic carbon distribution and accumulation in croplands of East China over three decades. Agriculture, Ecosystems & Environment 359, 108749.

Qadir, A., Skakun, S., Eun, J., Prashnani, M., Shumilo, L., 2023. Sentinel-1 time series data for sunflower (Helianthus annuus) phenology monitoring. Remote Sensing of Environment 295, 113689.

Qadir, A., Skakun, S., Kussul, N., Shelestov, A., Becker-Reshef, I., 2024. A generalized model for mapping sunflower areas using Sentinel-1 SAR data. Remote Sensing of Environment 306, 114132.

Rial, M., Mart ́nez Cortizas, A., Rodr ́guez-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.

Richer-de-Forges, A.C., Chen, Q., Baghdadi, N., Chen, S., Gomez, C., Jacquemoud, S., Martelet, G., Mulder, V.L., Urbina-Salazar, D., Vaudour, E., Weiss, M., Wigneron, J.-P., Arrouays, D., 2023a. Remote Sensing Data for Digital Soil Mapping in French Research—A Review. Remote Sensing 15(12), 3070.

Richer-de-Forges, A.C., Chen, Q., Baghdadi, N., Chen, S., Gomez, C., Jacquemoud, S., Martelet, G., Mulder, V.L., Urbina-Salazar, D., Vaudour, E., Weiss, M., Wigneron, J.-P., Arrouays, D., 2023b. Remote Sensing Data for Digital Soil Mapping in French Research—A Review. Remote Sensing 15(12), 3070.

Santos, E.P.d., Moreira, M.C., Fernandes-Filho, E.I., Dematt ê, J.A.M., Dionizio, E.A., Silva, D.D.d., Cruz, R.R.P., Moura-Bueno, J.M., Santos, U.J.d., Costa, M.H., 2023. Sentinel-1 Imagery Used for Estimation of Soil Organic Carbon by Dual-Polarization SAR Vegetation Indices. Remote Sensing 15(23), 5464.

Sayedain, S.A., Maghsoudi, Y., Eini-Zinab, S., 2020. Assessing the use of cross-orbit Sentinel-1 images in land cover classification. International Journal of Remote Sensing 41(20), 7801-7819.

Singh, S.K., Prasad, R., Srivastava, P.K., Yadav, S.A., Yadav, V.P., Sharma, J., 2023. Incorporation of first-order backscattered power in Water Cloud Model for improving the Leaf Area Index and Soil Moisture retrieval using dual-polarized Sentinel-1 SAR data. Remote Sensing of Environment 296, 113756.

Song, J., Yu, D.-s., Wang, S.-w., Zhao, Y.-h., Wang, X., Ma, L.-x., Li, J.-g., 2023. Mapping soil organic matter in cultivated land based on multi-year composite images on monthly time scales. Journal of Integrative Agriculture.

Song, J., Yu, D., Wang, S., Zhao, Y., Wang, X., Ma, L., Li, J., 2024. Mapping soil organic matter in cultivated land based on multi-year composite images on monthly time scales. Journal of Integrative Agriculture 23(4), 1393-1408.

Sothe, C., Gonsamo, A., Arabian, J., Snider, J., 2022. Large scale mapping of soil organic carbon concentration with 3D machine learning and satellite observations. Geoderma 405, 115402.

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. Agriculture, Ecosystems & Environment 258, 129-142.

Taghizadeh-Mehrjardi, R., Sheikhpour, R., Zeraatpisheh, M., Amirian-Chakan, A., Toomanian, N., Kerry, R., Scholten, T., 2022. Semi-supervised learning for the spatial extrapolation of soil information. Geoderma 426, 116094.

Thorp, K.R., Drajat, D., 2021. Deep machine learning with Sentinel satellite data to map paddy rice production stages across West Java, Indonesia. Remote Sensing of Environment 265, 112679.

Tian, Y., Xu, Y.-P., Wang, G., 2018. Agricultural drought prediction using climate indices based on Support Vector Regression in Xiangjiang River basin. Science of the Total Environment 622-623, 710-720.

Urbazaev, M., Thiel, C., Mathieu, R., Naidoo, L., Levick, S.R., Smit, I.P.J., Asner, G.P., Schmullius, C., 2015. Assessment of the mapping of fractional woody cover in southern African savannas using multi-temporal and polarimetric ALOS PALSAR L-band images. Remote Sensing of Environment 166, 138-153.

Vabalas, A., Gowen, E., Poliakoff, E., Casson, A.J., 2019. Machine learning algorithm validation with a limited sample size. PloS One 14(11), e0224365.

Vergni, L., Vinci, A., Todisco, F., Santaga, F.S., Vizzari, M., 2021. Comparing Sentinel-1, Sentinel-2, and Landsat-8 data in the early recognition of irrigated areas in central Italy. Journal of Agricultural Engineering 52(4).

Vreugdenhil, M., Navacchi, C., Bauer-Marschallinger, B., Hahn, S., Steele-Dunne, S., Pfeil, I., Dorigo, W., Wagner, W., 2020. Sentinel-1 Cross Ratio and Vegetation Optical Depth: A Comparison over Europe. Remote Sensing 12(20), 3404.

Wadoux, A.M.J.C., Minasny, B., McBratney, A.B., 2020. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. Earth-Science Reviews 210, 103359.

Wang, B., Waters, C., Orgill, S., Gray, J., Cowie, A., Clark, A., Liu, D.L., 2018. 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, H., Yang, H., Huang, Y., Wu, L., Guo, Z., Li, N., 2023a. Classification of Land Cover in Complex Terrain Using Gaofen-3 SAR Ascending and Descending Orbit Data. Remote Sensing 15(8), 2177.

Wang, L., Wang, X., Kooch, Y., Song, K., Wu, D., 2023b. Improvement of data imbalance for digital soil class mapping in Eastern China. Computers and Electronics in Agriculture 214, 108322.

Wang, X., Wang, L., Li, S., Wang, Z., Zheng, M., Song, K., 2022. Remote estimates of soil organic carbon using multi-temporal synthetic images and the probability hybrid model. Geoderma 425, 116066.

Wang, Y., Chen, S., Hong, Y., Hu, B., Peng, J., Shi, Z., 2023c. A comparison of multiple deep learning methods for predicting soil organic carbon in Southern Xinjiang, China. Computers and Electronics in Agriculture 212, 108067.

Wang, Y., Wang, S., Adhikari, K., Wang, Q., Sui, Y., Xin, G., 2019. Effect of cultivation history on soil organic carbon status of arable land in northeastern China. Geoderma 342, 55-64.

Wood, D., McNairn, H., Brown, R.J., Dixon, R., 2002. The effect of dew on the use of RADARSAT-1 for crop monitoring: Choosing between ascending and descending orbits. Remote Sensing of Environment 80(2), 241-247.

Wu, F., Wang, C., Zhang, H., Zhang, B., Tang, Y., 2011. Rice Crop Monitoring in South China With RADARSAT-2 Quad-Polarization SAR Data. IEEE Geoscience and Remote Sensing Letters 8(2), 196-200.

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.

Xiao, S., Ou, M., Geng, Y., Zhou, T., 2023. Mapping soil pH levels across Europe: An analysis of LUCAS topsoil data using random forest kriging (RFK). Soil Use and Management 39(2), 900-916.

Xiao, Y., Xue, J., Zhang, X., Wang, N., Hong, Y., Jiang, Y., Zhou, Y., Teng, H., Hu, B., Lugato, E., Richer-de-Forges, A.C., Arrouays, D., Shi, Z., Chen, S., 2022. Improving pedotransfer functions for predicting soil mineral associated organic carbon by ensemble machine learning. Geoderma 428, 116208.

Xu, Y., Li, B., Shen, X., Li, K., Cao, X., Cui, G., Yao, Z., 2022. Digital soil mapping of soil total nitrogen based on Landsat 8, Sentinel 2, and WorldView-2 images in smallholder farms in Yellow River Basin, China. Environmental Monitoring and Assessment 194(4), 282.

Yang, L., Jiao, Y., Fahmy, S., Zhu, A.X., Hann, S., Burt, J.E., Qi, F., 2011. Updating Conventional Soil Maps through Digital Soil Mapping. Soil Science Society of America Journal 75(3), 1044-1053.

Yang, R.-M., Guo, W.-W., 2019. 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., Liu, L.-A., Zhang, X., He, R.-X., Zhu, C.-M., Zhang, Z.-Q., Li, J.-G., 2022. The effectiveness of digital soil mapping with temporal variables in modeling soil organic carbon changes. Geoderma 405, 115407.

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.

Yilmaz, O.S., Acar, U., Sanli, F.B., Gulgen, F., Ates, A.M., 2023. Mapping burn severity and monitoring CO content in Turkiye's 2021 Wildfires, using Sentinel-2 and Sentinel-5P satellite data on the GEE platform. Earth Sci Inform 16(1), 221-240.

Yun, S., Segall, P., Zebker, H., 2006. Constraints on magma chamber geometry at Sierra Negra Volcano, Galápagos Islands, based on InSAR observations. Journal of Volcanology and Geothermal Research 150(1), 232-243.

Zendehboudi, A., Baseer, M.A., Saidur, R., 2018. Application of support vector machine models for forecasting solar and wind energy resources: A review. Journal of Cleaner Production 199, 272-285.

Zhang, G.-l., Liu, F., Song, X.-d., 2017. Recent progress and future prospect of digital soil mapping: A review. Journal of Integrative Agriculture 16(12), 2871-2885.

Zhang, H., Lin, H., Li, Y., Zhang, Y., Fang, C., 2016. Mapping urban impervious surface with dual-polarimetric SAR data: An improved method. Landscape and Urban Planning 151, 55-63.

Zhang, M.-W., Wang, X.-Q., Ding, X.-G., Yang, H.-L., Guo, Q., Zeng, L.-T., Cui, Y.-P., Sun, X.-L., 2023a. Monitoring regional soil organic matter content using a spatiotemporal model with time-series synthetic Landsat images. Geoderma Regional 34, e00702.

Zhang, M., Zhang, M., Yang, H., Jin, Y., Zhang, X., Liu, H., 2021. Mapping Regional Soil Organic Matter Based on Sentinel-2A and MODIS Imagery Using Machine Learning Algorithms and Google Earth Engine. Remote Sensing 13(15), 2934.

Zhang, Q., Liu, M., Zhang, Y., Mao, D., Li, F., Wu, F., Song, J., Li, X., Kou, C., Li, C., Man, W., 2023b. Comparison of Machine Learning Methods for Predicting Soil Total Nitrogen Content Using Landsat-8, Sentinel-1, and Sentinel-2 Images. Remote Sensing 15(11), 2907.

Zhang, W., Luo, C., Meng, X., Zang, D., Zhang, X., Liu, H., 2024. Predicting regional soil organic matter content utilizing conventional satellites: Assessing the influence of temporal, spatial, and spectral disparities. Catena 237, 107821.

Zhang, X., Chen, S., Xue, J., Wang, N., Xiao, Y., Chen, Q., Hong, Y., Zhou, Y., Teng, H., Hu, B., Zhuo, Z., Ji, W., Huang, Y., Gou, Y., Richer-de-Forges, A.C., Arrouays, D., Shi, Z., 2023c. Improving model parsimony and accuracy by modified greedy feature selection in digital soil mapping. Geoderma 432, 116383.

Zhou, T., Geng, Y., Chen, J., Liu, M., Haase, D., Lausch, A., 2020a. Mapping soil organic carbon content using multi-source remote sensing variables in the Heihe River Basin in China. Ecological Indicators 114, 106288.

Zhou, T., Geng, Y., Chen, J., Pan, J., Haase, D., Lausch, A., 2020b. 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 729, 138244.

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. Science of the Total Environment 755, 142661.

Zhou, Y., Zhao, X., Guo, X., Li, Y., 2022. Mapping of soil organic carbon using machine learning models: Combination of optical and radar remote sensing data. Soil Science Society of America Journal 86(2), 293-310.

Table 1 Overview of features derived from Sentinel satellite observations.

Sensors	Types	Description
Sentinel-1	Backscattering coefficient	Backscattering coefficient in VV polarization
Sentinel-1	Backscattering coefficient	Backscattering coefficient in VH polarization
Sentinel-1	Backscattering coefficient	Backscatter coefficient from ascending orbit
Sentinel-1	Backscattering coefficient	Backscatter coefficient from descending orbit
Sentinel-1	SAR indices	Ratio images of VV and VH (VV/VH)
Sentinel-1	SAR indices	Difference images of VV and VH (VV − VH)
Sentinel-1	SAR indices	Sum images of VV and VH (VV + VH)
Sentinel-1	SAR indices	Normalized difference of backscattering coefficients for VV and VH polarizations ((VV − VH) / (VV + VH)).
Sentinel-2	Optical images	Sentinel-2 spectral bands

Table 2. Details of the modeling scenarios based on different data integration strategies

Scenarios	Modeling groups	Input data
SC1	Sentinel-1	VV backscattering coefficients
SC2	Sentinel-1	VH backscattering coefficients
SC3	Sentinel-1	Backscatter bands in the ascending orbit
SC4	Sentinel-1	Backscatter bands in the descending orbit
SC5	Sentinel-1	Radar backscatter coefficients (VV and VH)
SC6	Sentinel-1	SAR indices (VV/VH, VV-VH, VV+VH, and (VV − VH) / (VV + VH))
SC7	Sentinel-1	Backscatter bands and SAR indices
SC8	Sentinel-2	Sentinel-2 derived features
SC9	Sentinel-1/2	Sentinel-1/2 derived features
SC10	Terrain and climate	Topographic and climatic indicators
SC11	Sentinel-1/2, terrain and climate	Sentinel-1/2-derived features, climatic and topographic indicators

Table 3. Accuracy results for predicting the soil C:N ratio using eleven scenarios

Predictive models	Scenarios	RMSE	R^2	MAE
RF	SC1	5.06	0.46	2.93
	SC2	5.00	0.48	2.85
	SC3	4.93	0.50	2.79
	SC4	4.99	0.48	2.84
	SC5	4.89	0.51	2.76
	SC6	4.90	0.51	2.78
	SC7	4.88	0.51	2.76
	SC8	5.03	0.47	2.87
	SC9	4.80	0.54	2.68
	SC10	4.93	0.49	2.83
	SC11	4.62	0.59	2.54
SVM	SC1	5.07	0.47	2.87
	SC2	5.00	0.49	2.80
	SC3	4.97	0.50	2.77
	SC4	4.99	0.48	2.78
	SC5	4.85	0.52	2.69
	SC6	4.87	0.52	2.71
	SC7	4.86	0.52	2.69
	SC8	5.07	0.48	2.78
	SC9	4.77	0.55	2.60
	SC10	5.25	0.44	2.96
	SC11	4.70	0.58	2.55

Figure Legends

Fig. 1. Sampling points collected in different land use types (a) and number of sampling points in each country (b) in the LUCAS 2018 survey.

Fig. 2. Spatial pattern of the total number of Sentinel observations in the study area.

Fig. 3. Importance scores of the top twenty most important predictors for C:N ratio modeling using the different scenarios: (a) SC7; (b) SC8; (c) SC10; and (d) SC11 (refer to Table 2 for the modeling scenario details). The names of the Sentinel-1-derived features were abbreviated according to the different periods during which they were composited (refer to Section 2.2 for more details on the abbreviated names of the input variables).

Fig. 4. Mapping results of the soil C:N ratio in Europe using the RF model for different scenarios: (a) SC5; (b) SC6; (c) SC8; (d) SC9; (e) SC10; and (f) SC11 (refer to Table 2 for the modeling scenario details).

Fig. 5. Mapping results of the soil C:N ratio in Europe using the SVM model for different scenarios: (a) SC5; (b) SC6; (c) SC8; (d) SC9; (e) SC10; and (f) SC11 (refer to Table 2 for the modeling scenario details).

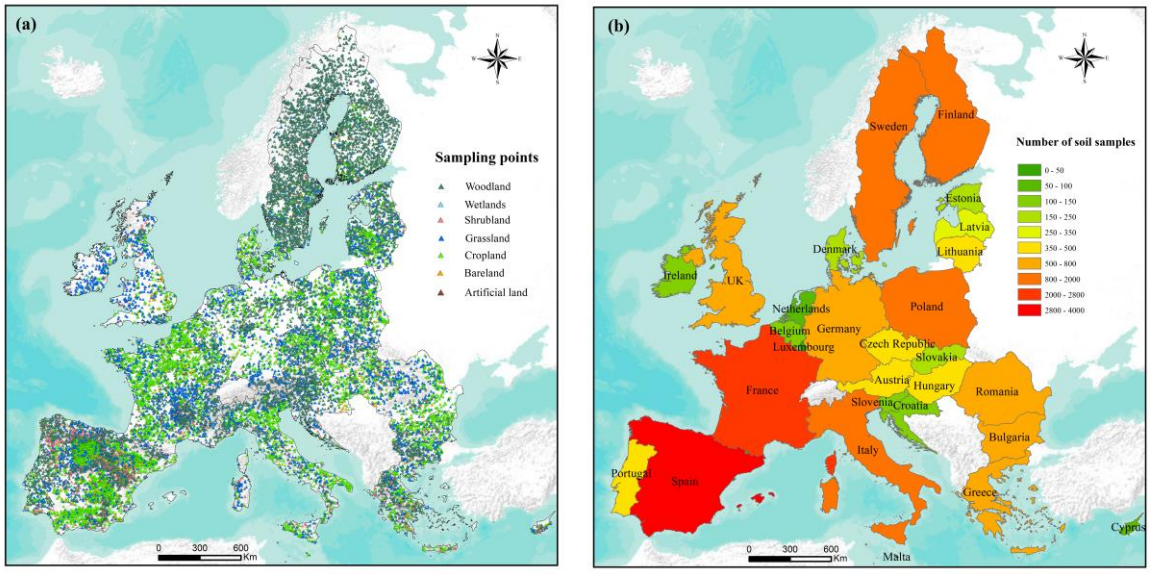

(a)
Sampling points
Woodland
Wetlands
Shrubland
Grassland
Cropland
Bareland
Artificial land
0 300 600 Km
(b)
Number of soil samples
0 - 50
50 - 100
100 - 150
150 - 250
250 - 350
350 - 500
500 - 800
800 - 2000
2000 - 2800
2800 - 4000
Sweden
Finland
Estonia
Latvia
Lithuania
Denmark
Ireland
UK
Netherlands
Belgium
Germany
Luxembourg
Poland
Czech Republic
Slovakia
Austria
Hungary
Romania
France
Slovenia
Croatia
Bulgaria
Italy
Portugal
Spain
Greece
Cyprus
Malta
0 300 600 Km

Fig. 1.

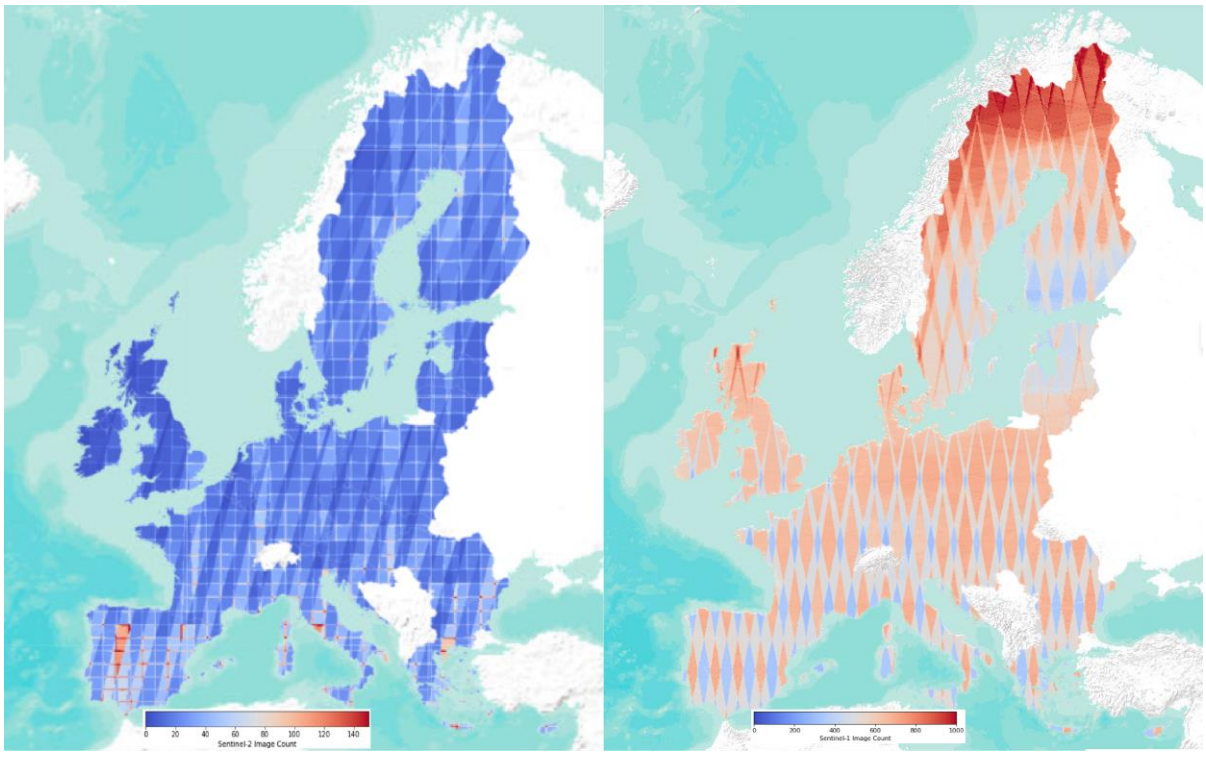

0 20 40 60 80 100 120 140
Sentinel-2 Image Count
0 200 400 600 800 1000
Sentinel-1 Image Count

Fig. 2.

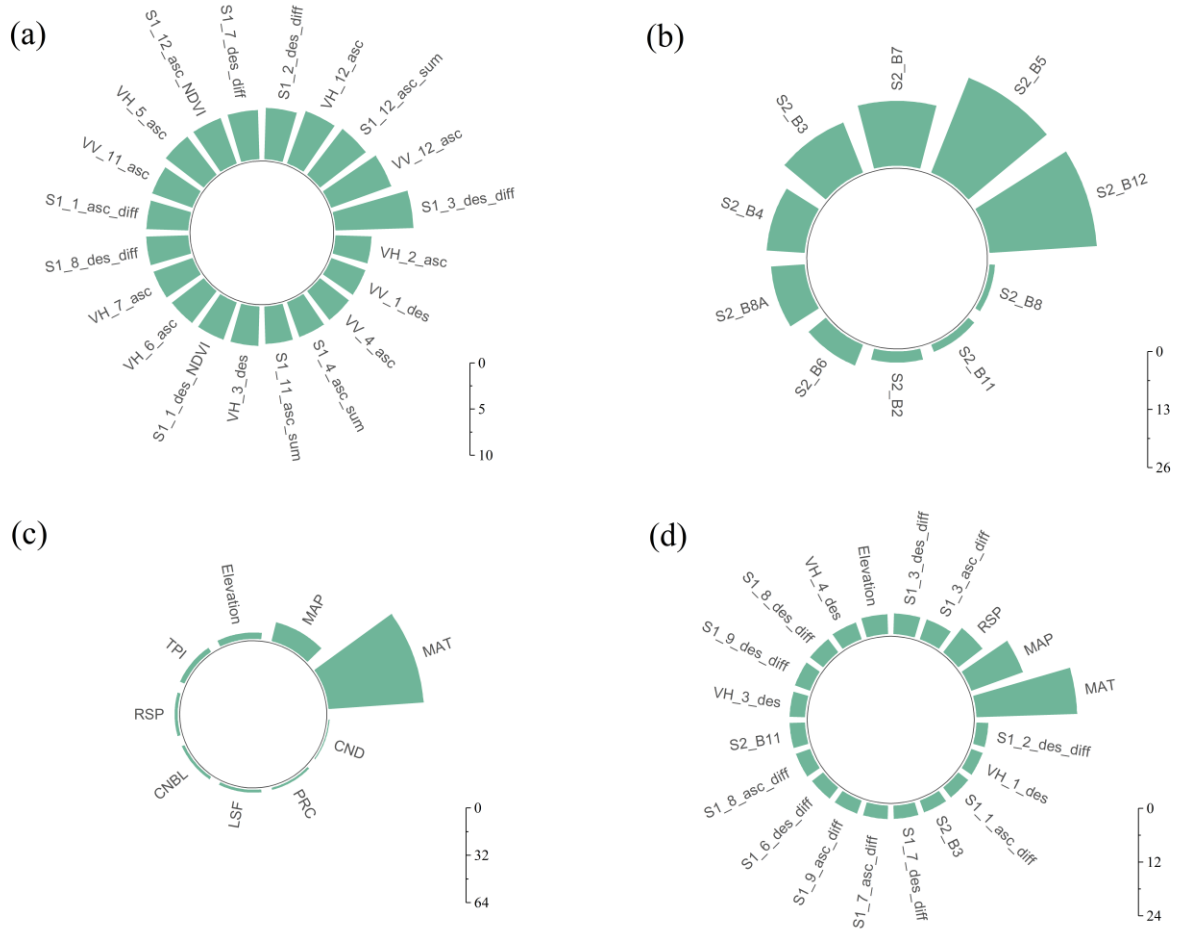

(a)
S1_12_asc_NDVI
S1_7_des_diff
S1_2_des_diff
VH_12_asc
S1_12_asc_sum
VV_12_asc
S1_3_des_diff
VH_2_asc
VV_1_des
VV_4_asc
S1_4_asc_sum
S1_11_asc_sum
VH_3_des
S1_1_des_NDVI
VH_6_asc
VH_7_asc
S1_8_des_diff
S1_1_asc_diff
VV_11_asc
VH_5_asc
0
5
10
(b)
S2_B7
S2_B5
S2_B3
S2_B12
S2_B4
S2_B8
S2_B8A
S2_B11
S2_B6
S2_B2
0
13
26
(c)
Elevation
MAP
MAT
TPI
RSP
CND
CNBL
PRC
LSF
0
32
64
(d)
S1_3_des_diff
S1_3_asc_diff
Elevation
VH_4_des
RSP
S1_8_des_diff
MAP
S1_9_des_diff
MAT
VH_3_des
S1_2_des_diff
S2_B11
VH_1_des
S1_8_asc_diff
S1_1_asc_diff
S1_6_des_diff
S2_B3
S1_9_asc_diff
S1_7_des_diff
S1_7_asc_diff
0
12
24

Fig. 3.

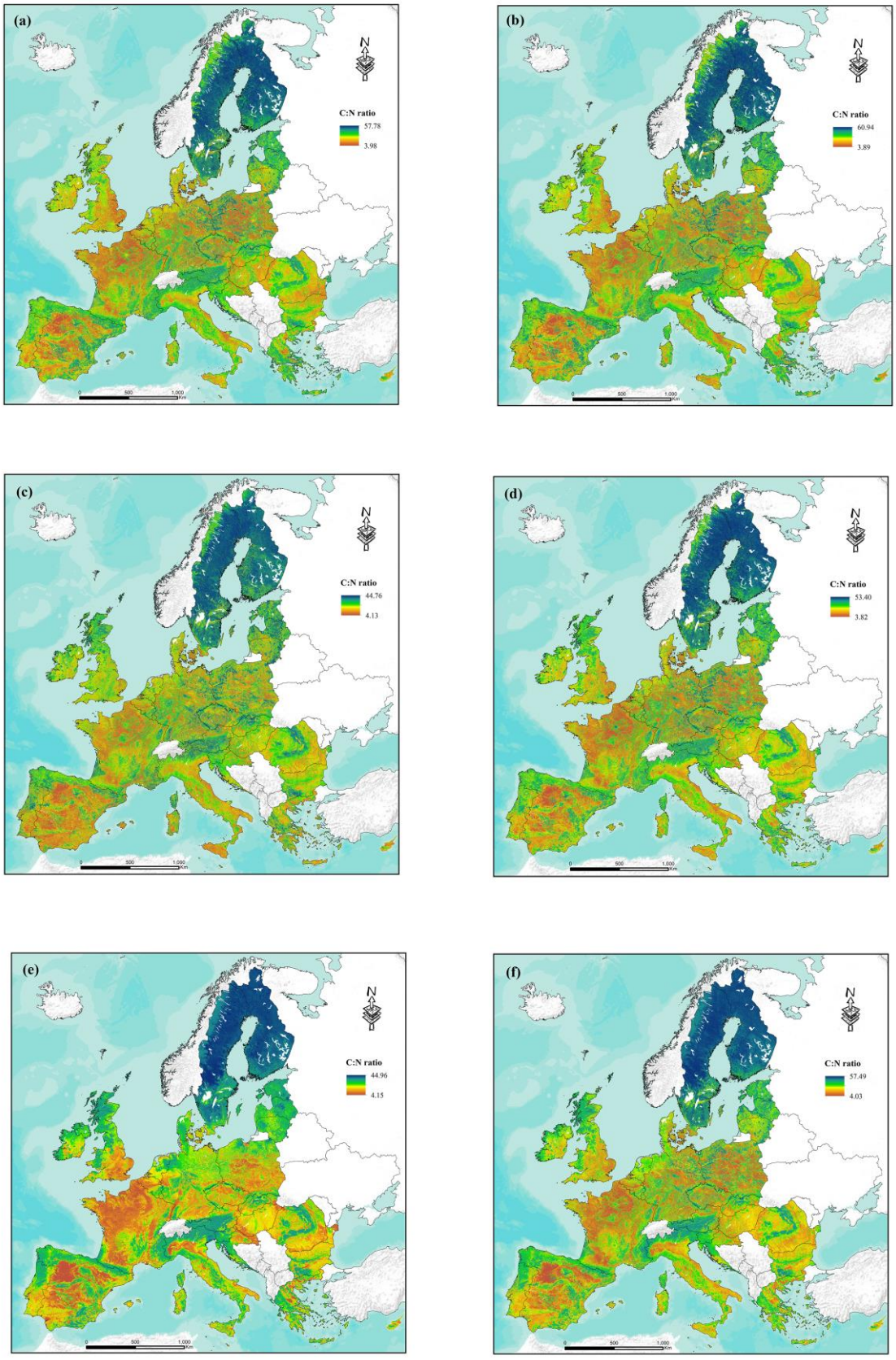

(a)
C:N ratio
57.78
3.98
(b)
C:N ratio
60.94
3.89
(c)
C:N ratio
44.76
4.13
(d)
C:N ratio
53.40
3.82
(e)
C:N ratio
44.96
4.15
(f)
C:N ratio
57.49
4.03

Fig. 4.

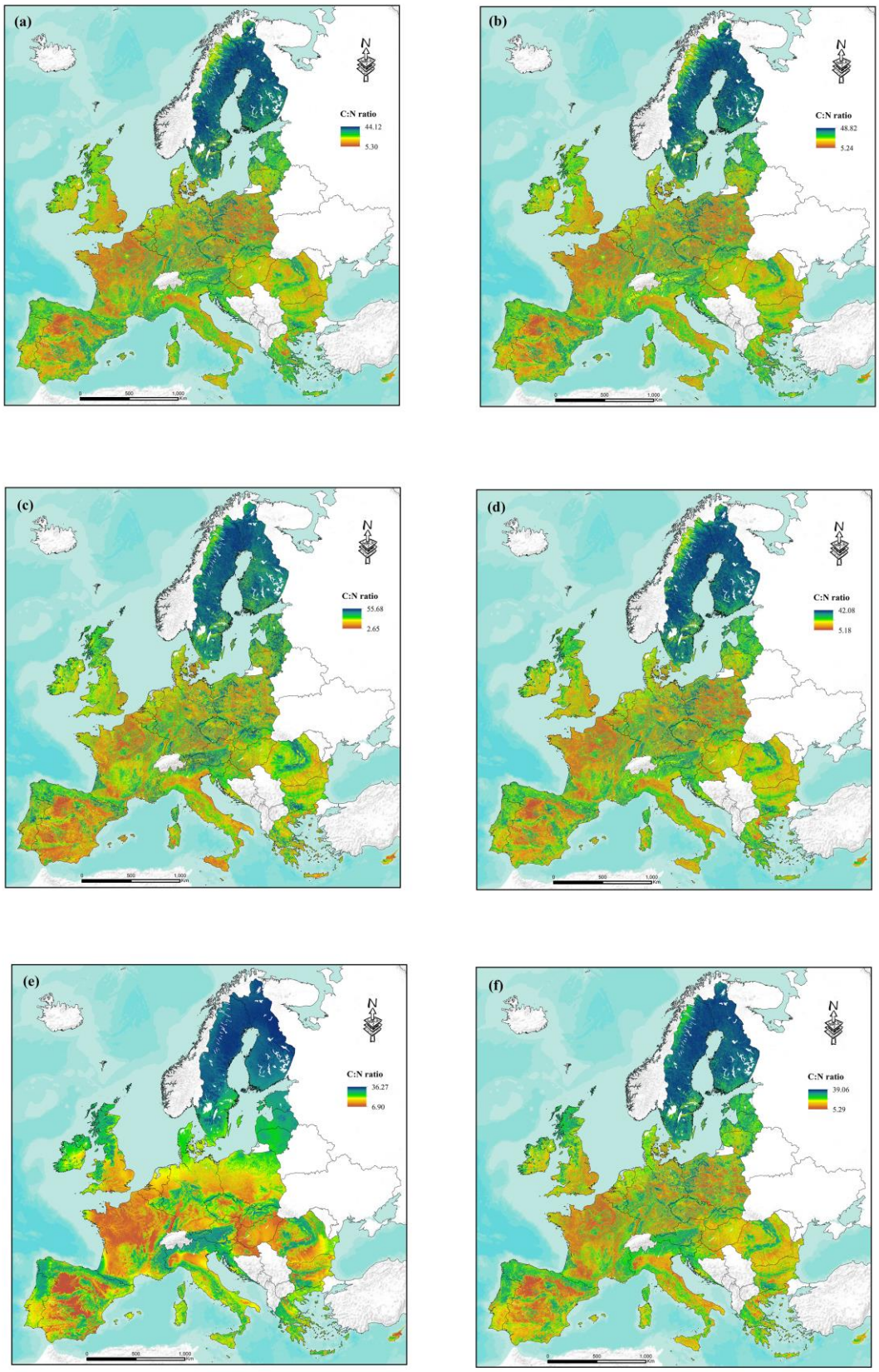

(a)
C:N ratio
44.12
5.30
(b)
C:N ratio
48.82
5.24
(c)
C:N ratio
55.68
2.65
(d)
C:N ratio
42.08
5.18
(e)
C:N ratio
36.27
6.90
(f)
C:N ratio
39.06
5.29

Fig. 5.