

Soil organic matter mapping using INLA-SPDE with remote sensing based soil moisture indices and Fourier transforms decomposed variables

Chenconghai Yang^a, Lin Yang^{a,c,*}, Lei Zhang^a, Chenghu Zhou^{a,b}

^a School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China

^b State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

^c Frontiers Science Center for Critical Earth Material Cycling, Nanjing University, Nanjing 210023, China

ARTICLE INFO

Handling Editor: Budiman Minasny

Keywords:

Soil organic matter
Digital soil mapping
INLA-SPDE
Soil moisture indices
NSDSI
Uncertainty assessment

ABSTRACT

Generating accurate spatial information on soil organic matter (SOM) is increasingly important in the context of global environmental change. Both prediction models and environmental covariates influence the mapping results and accuracy, making them important factors in SOM mapping. The Bayesian spatial model INLA-SPDE is an emerging model, that has shown potential in digital soil mapping (DSM), but its application is still limited. Soil moisture, which affects soil water status and the decomposition of SOM, can be a potential predictor for mapping SOM. However, the difficulty of obtaining soil moisture measurements over a large area using ground-based methods hinders its application. Recently, high spatial resolution remote sensing (RS) has provided a possible way to generate soil moisture indices over a large area. However, the effectiveness of RS-based soil moisture indices on SOM mapping is unknown. Fourier transforms decomposed (FTD) variables based on vegetation indices have been proven effective in detecting time-series patterns of crop growth, thereby improving the mapping accuracy of farmland. Yet, the effectiveness of FTD variables has not been verified in other vegetation-covered areas. This paper examines the use of INLA-SPDE with three RS-based soil moisture indices (NSDSIs) and six FTD variables for SOM mapping compared to Random Forest (RF), in a study area with diverse vegetation cover in Anhui Province, China. The finding indicates that with the optimal combination of environmental covariates, INLA-SPDE yields a higher prediction accuracy than RF, with an increase of 18% in R^2 . Either the RS-based soil moisture indices covariates or the FTD variables are effective in mapping SOM. When compared to using only natural environmental covariates, the best combination including RS-based soil moisture indices and FTD variables improved the mapping accuracy by 25% in terms of R^2 , 21% of LCCC, and 11% of RMSE. Furthermore, quantitative prediction uncertainty maps are derived based on the INLA-SPDE. This study demonstrates the effectiveness of INLA-SPDE model with the RS-based soil moisture indices and Fourier transforms decomposed variables for SOM mapping.

1. Introduction

Soil organic matter (SOM) plays a crucial role in the global carbon cycle, soil health, and food security (Johnston et al., 2009; McBratney et al., 2014; Wang et al., 2022). The significant amount of carbon present in soils has the potential to impact global climate change, and can serve as a potential sink for carbon sequestration (Lal, 2020; Zhou et al., 2019). Obtaining accurate information on the spatial distribution of SOM is important for understanding soil quality, guiding sustainable

management practices, and providing guidance for climate change mitigation efforts (Lal, 2020; Moeskops et al., 2012).

In recent decades, with the development of geographic information systems, environmental data acquisition technology, and machine learning models, digital soil mapping (DSM) has become an effective method for soil mapping (Arrouays et al., 2021; Chen et al., 2022). DSM involves establishing relationships between soil and its environmental covariates based on soil sample points, and using these relationships to predict the spatial distribution of soil properties (McBratney et al.,

Abbreviations: INLA-SPDE, Integrated Nested Laplace Approximation with the Stochastic Partial Differential Equation; NSDSI, Normalized Shortwave-infrared Difference SM Indices; FTD, Fourier Transforms Decomposed.

* Corresponding author.

E-mail addresses: yanglin@nju.edu.cn (L. Yang), lei.zhang_geo@outlook.com (L. Zhang).

<https://doi.org/10.1016/j.geoderma.2023.116571>

Received 8 January 2023; Received in revised form 11 May 2023; Accepted 11 June 2023

0016-7061/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

2003). The accuracy of DSM largely depends on the approaches/models to establish soil-environment relationships and the environmental covariates used (McBratney et al., 2003).

Numerous models have been developed to map soil properties, including linear regression (e.g. Wu et al., 2009), geographically weighted regression (e.g. Odhiambo et al., 2020), kriging approaches (e.g. Kerry et al., 2012; Zare et al., 2021), machine learning (e.g. Jahan-dideh Mahjenabadi et al., 2022; Purushothaman et al., 2022) and deep learning (e.g. Wadoux, 2019; Zhang et al., 2022), etc. Meanwhile, there has been increasing attention paid to quantifying the uncertainty of DSM as it provides an important reference for the reliability of soil maps (Huang et al., 2017; Poggio et al., 2016). Besides, for calculating uncertainty from a large number of realizations or through sample bootstrapping, Bayesian approaches have been developed to explicitly describe uncertainty explicitly by considering the model parameters as probabilistic variables with posterior probability density functions (pdfs) (Huang et al., 2017; Minasny and McBratney, 2016; Poggio et al., 2016). Markov Chain Monte Carlo (MCMC) simulations are a commonly used method for obtaining the posterior distribution of model parameters in Bayesian inference, as demonstrated in studies by Minasny et al. (2011) and Sun et al. (2013). Furthermore, in order to address the challenge of the large computational volume of MCMC, the Integrated Nested Laplace Approximation with the Stochastic Partial Differential Equation approach (INLA-SPDE) has been developed. This involved computing numerical approximations to the marginal density for both hyper-parameters and latent variables (Huang et al., 2017; Rue et al., 2009; Ryan et al., 2015). This approach offers a more efficient alternative to MCMC for Bayesian inference, particularly in cases where the model involves complex and high-dimensional data, and has been successfully applied in many research fields, such as mapping the spatial distribution of geochemical patterns (Wang and Zuo, 2021), ambient air pollution (Wright et al., 2021), spatio-temporal characteristics of disease spreading (Moraga et al., 2021), etc.

The INLA-SPDE model was first applied in DSM by Poggio et al. (2016), and their study verified the feasibility of using INLA-SPDE for mapping SOM in the Grampian region of Scotland, with the advantages of good assessment of uncertainty. Huang et al. (2017) demonstrated that INLA-SPDE is as robust as the residual maximum likelihood for mapping SOC, Ti/Zr ratio and pH in the Nowley farm of Australia with sparse datasets. Beguin et al. (2017) suggested that INLA-SPDE has generated higher prediction accuracy to predict key forest soil properties in the case of limited samples when compared with Ordinary Kriging (OK), Random Forest (RF) and Boosted Regression Tree (BRT). Sun et al. (2021) further adopted INLA-SPDE for modeling spatial-temporal change of SOM on a regional scale. These previous studies demonstrate the potential of INLA-SPDE in DSM, but its application is still limited.

The spatial distribution of SOM is influenced by various environmental factors such as climate, topography, soil properties, and vegetation, etc (Minasny et al., 2013). Remote sensing (RS) has become a valuable tool for predicting SOM distribution due to its high spatial resolution, long-term data availability, and accessibility (He et al., 2021; Omrani et al., 2021). Vegetation-related indices, such as the Normalized Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI), and Enhanced Vegetation Index (EVI), are commonly used covariates extracted from RS data. Long-term time series of vegetation indices provide valuable information on the vegetation growth status, which can be used to map SOM distribution. These techniques have been increasingly applied in recent years. (Maynard and Levi, 2017). The behind theory is that different growth characteristics within a long period relating to behaviors of vegetation influence soil carbon concentrations' absorption and decomposition (Richardson et al., 2010). Fourier transforms decomposed (FTD) covariate, generated by decomposing the time series NDVI into periodic signals in the frequency domain, is one of the promising predictors for mapping SOM developed based on time-series RS images (Yang et al., 2019). Yang et al. (2019) demonstrated the effectiveness of FTD covariates in reflecting

agricultural activities in a typical agricultural area. This may be due the fact that the periodic patterns indicated by FTD covariates are related to different crop species/crop rotations, or different growth stages within one crop species. However, FTD covariates have only been applied in a farmland areas. Therefore it is necessary to explore the effectiveness of these covariates on DSM in other vegetation-covered areas.

Besides vegetation, soil moisture content affects soil water status and the decomposition of SOM (Yoshida et al., 2018), and also influences vegetation evapotranspiration and vegetation growth (Zhang et al., 2019a; Zhang and Zhou, 2016), thus significantly relating to SOM content (Yoshida et al., 2018). As a result, soil moisture can be a potential covariate for predicting the spatial distribution of SOM. However, obtaining large-scale soil moisture estimations based on ground-based observations is costly (Rao et al., 2022; Tian et al., 2021), which may be a reason why the use of soil moisture in DSM is rare. Recently, microwave and thermal remote sensing have been adopted to reflect soil moisture over large areas. However, the spatial resolution of microwave remote sensing is usually coarse (Wigneron et al., 2017; Yue et al., 2019), and thermal remote sensing is strongly influenced by atmospheric factors (near-surface air temperature, etc.) (Sandholt et al., 2002). Optical remote sensing has been utilized to assist in the production of soil moisture information through water absorption bands (Tian et al., 2021; Yue et al., 2019). Several soil moisture indices are developed, such as the Water Index SOIL (WISOIL), the Normalized Index of Nswir domain for Smc estimatiOn from Linear correlation (NINSOIL), the Normalized Index of Nswir domain for Smc estimatiOn from Non-linear correlation (NINSON) (Fabre et al., 2015). However, the indices currently available are primarily generated from hyperspectral (narrow-band) remote sensing, which has limited spatial resolutions (Yue et al., 2019; Zhang et al., 2019b). To address this issue, Yue et al. (2019) proposed the use of normalized shortwave-infrared difference bare soil moisture indices (NSDSI) based on Landsat data with a spatial resolution of 30 m. This approach provides a better spatial resolution and has been shown to achieve comparable accuracy to narrow-band soil moisture indices. The development of these indices has expanded the potential for using RS-based soil moisture indices to map SOM.

In this study, we employed INLA-SPDE to predict the spatial distribution of SOM with soil moisture indices (NSDSI) and Fourier transforms decomposed (FTD) covariates in a study area located in Xuancheng, Anhui province, China. The objectives of this paper are, (1) to explore the prediction accuracy of INLA-SPDE approach to SOM by compared with Random Forest (RF); (2) to evaluate the effectiveness of NSDSI and FTD as environmental covariates to predict SOM spatial distribution (3) to analyze the uncertainty spatial distribution of SOM in this study area by INLA-SPDE and RF models.

2. Material and methods

2.1. Study area

The study area is Xuancheng city and Langxi County, located in Anhui province of China (30°33'59"-31°18'14" N, 118°27'38"-119°22'55" W, as shown in Fig. 1). This area is situated in the transition zone between the Southeast Hills and Yangtze Valley Plain, covering approximately 2552 km² (Yang et al., 2019). The subtropical monsoon climate of this region is characterized by warm and humid summers, and mild and dry winters, with an average annual temperature ranging from 12 to 18 °C and an annual precipitation of 1200 to 1800 mm, which is concentrated between April and October (Yang et al., 2021). The terrain in this area is generally flat plains in the northwest, and hilly in the south and north-east, with elevations ranging from 0 to 1058 m. The local parent materials of the soil are diverse, including clay-silt-gravel, stone, shale, conglomerate, pyroclastic rocks, granite, granodiorite, limestone, and other weathered residuals (Yang et al., 2019). According to the Chinese Soil Taxonomy, the soil types in the region are mainly Ferric-Udic

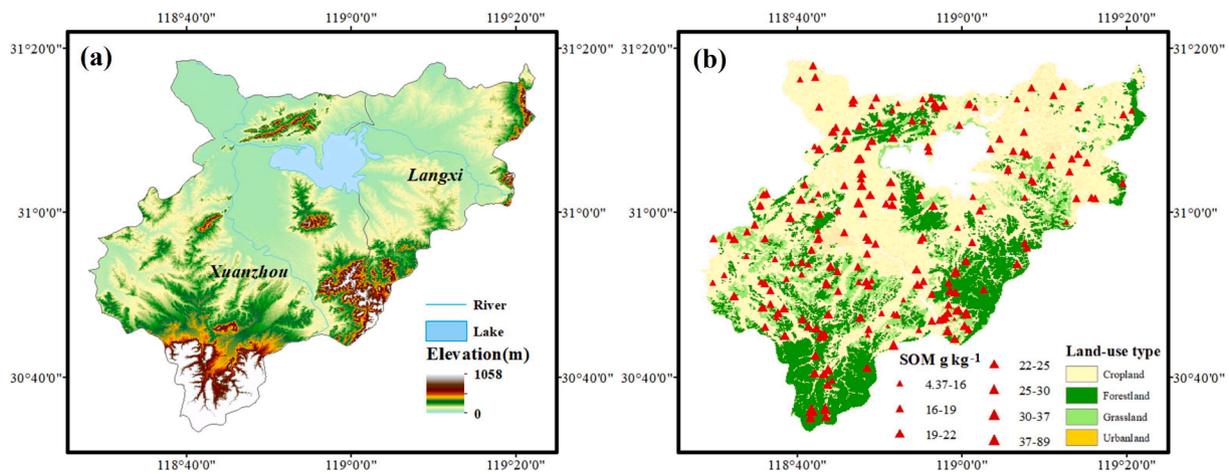


Fig. 1. Digital elevation model (DEM) (a), land-use and sample distribution (b) of the study area.

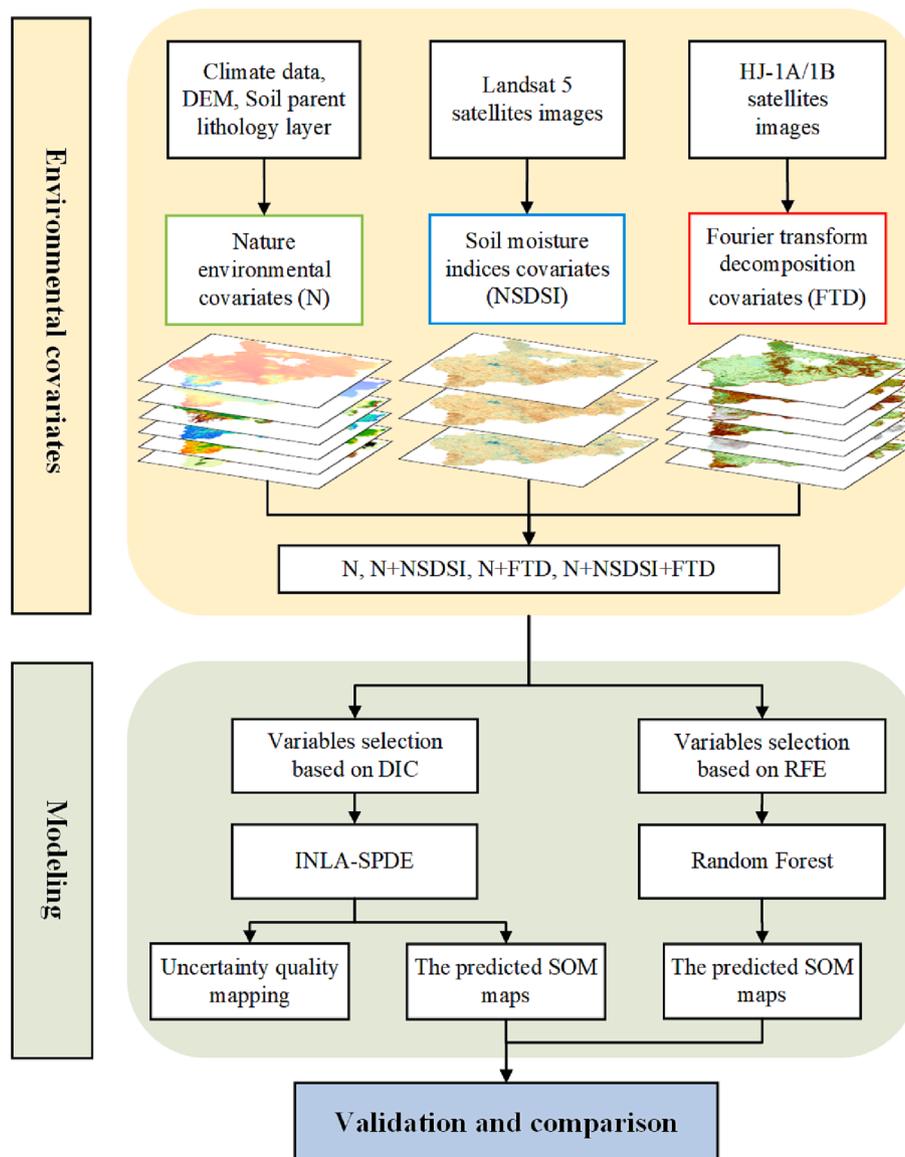


Fig. 2. The flow chart of using different covariates and models for SOM mapping.

Argosols in their natural state, and Stagnic Anthrosols formed by long-term artificial cultivation (Qin et al., 2021), corresponding to the Ferric Lixisols and Hydragric Anthrosols in the World Reference Base for Soil Resources (WRB), respectively.

According to data from the Resource and Environment Science and Data Center (RESDC) (<https://www.resdc.cn/>), in 2010 (the sampling period), farmland covered approximately 62.7% of the vegetation-covered areas in Xuancheng, forestland and grassland accounted for 27.2% and 10.1%, respectively. Various crop rotation types were observed in the area, including rice–wheat, double rice, rice–rape, single rice, and minor rice–tobacco (Yang et al., 2020). And the main wild forest tree types were *Ginkgo biloba*, *Parrotia subaequalis*, *Emmenopterys henryi* Oliv and *Metasequoia glyptostroboides*, etc.

2.2. Soil sample data

Between 2010 and 2011, a total of 212 sampling points were collected at a depth of 0–20 cm using various sampling strategies from previous studies (Fig. 1). A systematic sampling strategy with a 10 km by 10 km grid arrangement was used to select thirty-nine sampling points (Zhang et al., 2022). A stratified random sampling strategy was employed to collect forty-one sampling points (Yang et al., 2013). Thirty-six sampling points were collected using the heuristic uncertainty-directed sampling method (Zhang et al., 2016). The remaining ninety-six sampling points were obtained through representative grade sampling (Yang et al., 2013). A mixed sample was collected at each sample point, and multiple samples within 10 m of the surrounding area were mixed into one sample (Zhou et al., 2019).

Among all samples, 142, 43, and 14 sampling points were located in farmland, forestland, and grassland, respectively. The remaining 13 sampling points were collected from other land use areas. The density of soil samples in this area was one observation per 12.03 km². SOM was measured using the dichromate oxidation method (external heat applied) (Sparks et al., 1996). As the SOM data were skewed distributed, we transformed the original data into normal distribution using logarithm transformation ($\log_{10}(y + 1)$) in modeling to satisfy the requirement of INLA-SPDE. Hereafter we use NorSOM to represent the transformed SOM. After the modeling was completed, the predicted soil property values were back-transformed to their original scale (Sun et al., 2021).

2.3. The overall framework of method

The method of our study is illustrated in the flow chart (Fig. 2). We used three types of environmental covariates, including the commonly-used natural covariates, RS-based soil moisture indices (NSDSI) covariates, and Fourier transform decomposition (FTD) covariates. Those covariates were combined into different combinations for evaluating their effectiveness. Then, different groups of environmental covariates were taken as input to build INLA-SPDE and RF models. Finally, the predicted soil organic matter (SOM) maps generated by INLA-SPDE and RF, using different combinations of environmental covariates, were compared using accuracy indices based on cross-validation. This comparison aimed to assess the performance of the two methods in predicting SOM.

2.4. Environmental covariates

2.4.1. Natural environmental covariates

We collected six natural environmental covariates representing climate, topography, and soil parent materials, i.e. annual mean temperature (TEM), annual mean precipitation (PRE), elevation (DEM), slope, topographic wetness index (TWI), and soil parent lithology (ParentL).

The annual mean temperature and annual precipitation in 2011 were downloaded from RESDC. This climate dataset with a 1 km original

spatial resolution was generated based on interpolation of daily observation data at about 2,400 meteorological stations through the software ANUSPLIN (Hutchinson, 1998). We resampled the 1 km data to 90 m spatial resolution with a cubic resampling approach in ArcGIS.

The topographic covariates were generated from the Digital Elevation Model (DEM), which was produced by the Shuttle Radar Topographic Mission (SRTM) at a 90 m resolution (<https://srtm.csi.cgiar.org/srtmdata>). The elevation (DEM), slope, and TWI (Kopecký et al., 2021) were generated from DEM data using the terrain analysis toolbox in ArcGIS.

The 1:500,000 Chinese geological maps were employed to generate the soil parent lithology layer. The soil parent lithology in this region contains eight types: pyroclastic rocks, shale, sandstone, conglomerate, granite and granodiorite, limestone, quaternary clay-silt-gravel, quaternary vermicule boulder, and grave clay (He et al., 2021).

2.4.2. Soil moisture indices

The basis for obtaining soil moisture information from RS images is that soil spectral reflectance is associated with soil moisture content (Taneja et al., 2021). When the crevices around the soil particle change from air pores to water, its refractive index decreases relatively (Tweymey et al., 1986), and the SWIR band reflectance is sensitive to the changes in soil moisture from dry to saturated (Tian and Philpot, 2015). Yue et al. (2019) developed three NSDSI indices based on different water absorption in shortwave-infrared bands, i.e. SWIR1 and SWIR2 of Landsat images (Table 1). These indices were proven to have comparable or higher accuracy than other soil moisture indices such as WISOIL, NSMI, NINSOL, and NINSON. Importantly, these indices have a stronger application ability because they can be generated from commonly used multispectral high spatial resolution RS data and are less affected by atmospheric water vapor absorption (Yue et al., 2019).

We utilized the above three soil moisture indices, NSDSI1, NSDSI2, and NSDSI3 in our study. We initially gathered Landsat 5 images with minimal cloud interference on various dates throughout the sampling period (February 21st, May 18th, July 21st, August 19th, October 19th, and December 9th) with a 30 m resolution from the GS Cloud platform (<https://www.gscloud.cn/>). We employed two criteria to determine the final images for computing soil moisture indices. Firstly, we selected images with minimal vegetation cover, such as during the winter season, to enable spectral information from remote sensing to more accurately reflect soil information. Secondly, we chose images with strong correlations with predicted soil properties.

2.4.3. Generation of the Fourier transform decomposition covariates

The Fourier transform has been proven effective in detecting periodic patterns in time-series vegetation-related data (Chen et al., 2018; Mingwei et al., 2008). Yang et al., (2019) conducted Fourier transform on time-series NDVI profiles and generated six new variables, i.e. the Amplitude (Amp1, Amp2, Amp3) and Phase (Pha1, Pha2, and Pha3) of three Fourier harmonics. More detailed information on FTD covariates was referred to Yang et al. (2019). These variables were solely used for soil mapping in farmland in their study. We adopted these six variables for predicting SOM over the whole study area.

Table 1

The descriptions of Normalized Shortwave-infrared (SWIR) Difference Bare Soil moisture Indices (NSDSI, 1 ~ 3).

Indices	Formula	Reference
NSDSI1	$\frac{B_{SWIR1} - B_{SWIR2}}{B_{SWIR1}}$	(Yue et al., 2019)
NSDSI2	$\frac{B_{SWIR1} - B_{SWIR2}}{B_{SWIR2}}$	
NSDSI3	$\frac{B_{SWIR1} - B_{SWIR2}}{B_{SWIR1} + B_{SWIR2}}$	

Notes: B_{SWIR1} and B_{SWIR2} represent the SWIR₁ (1.55 ~ 1.75 μm) and SWIR₂ (2.08 ~ 2.35 μm) wave bands in Landsat 5 satellite, respectively.

2.4.4. Development of different covariates combinations

To explore the effectiveness of NSDSI and FTD covariates in predicting SOM with different models, four covariates combinations, i.e., “N”, “N + NSDSI”, “N + FTD”, and “N + NSDSI + FTD” were established based on the three pools of environmental variables. The first combination was only natural environmental covariates. The second and third combinations were composed of the natural environmental covariates with the addition of the NSDSI covariates and FTD covariates, respectively. And the fourth combination consisted of all the covariates.

2.5. INLA-SPDE model

2.5.1. The basis of INLA-SPDE

Unlike frequentist theory, the Bayesian approach considers the model parameters as probabilistic variables with joint posterior probability density functions (pdfs). The estimated posterior marginal distribution of the model parameters is obtained using Bayes' theorem (Huang et al., 2017; Li et al., 2018). In INLA-SPDE, the posterior probability distribution of each parameter is inferred by incorporating external information, such as measured values of soil properties and environmental covariates, based on the Laplace approximation instead of simulations (i.e. MCMC approach) (Martins et al., 2013). Considering soil sample data as variables that change continuously in location s_1, s_2, \dots, s_n , the target soil property can be regarded as the spatial random field model of $X(s_1), X(s_2), \dots, X(s_n)$. If this spatial random field satisfies the normal distribution and has a defined covariance function, it can be represented as a Gaussian field (GF) (Wu, 2021). The Matérn covariance function is a commonly used function to define the spatial random field of GF to represent the spatial autocorrelation (Huang et al., 2017; Li et al., 2018; Wu, 2021):

$$\text{Cov}(X(s_i), X(s_j)) = \sigma_e^2 \delta_{ij} + \frac{\sigma_x^2}{2^{\nu-1} \Gamma(\nu)} (k \|s_i - s_j\|)^{\nu} (k \|s_i - s_j\|) \quad (1)$$

where $\|s_i - s_j\|$ represents the Euclidean distance between two point s_i and s_j in the two dimension space; σ_x^2 is the spatial variance of the Matérn variogram, correlating with the space effect; k is the scale parameter; ν is the smoothness parameter; k^{ν} is the modified Bessel function for the second type; $\Gamma(\nu)$ is Gamma function (Huang et al., 2017; Minasny and McBratney, 2005). The Matérn function also has the noise variance σ_e^2 , which is uncorrelated with space effect (i.e. nugget effect), δ_{ij} is the Kronecker function; if $i = j$, $\delta_{ij} = 1$, otherwise $\delta_{ij} = 0$.

If the model is a full spatial covariance function, it causes high computational costs, so the “big n problem” is involved (Banerjee et al., 2008; Poggio et al., 2016). Lindgren et al. (2011) suggested that a large class of GF (such as the spatial distribution of SOM) can be as solutions to stochastic partial differential equations (SPDE), so that a continuous spatial process of spatial random field can be represented by a discretely indexed spatial random process, and GF can be converted to the Gaussian-Markov random field (GMRF), which is the sparse matrix, thus significantly reducing the time-cost. The SPDE was based on the representation of the Matérn covariance function as a solution (Huang et al., 2017; Lindgren et al., 2011; Wu, 2021):

$$(k^2 - \Delta)^{\frac{\alpha}{2}} X(s) = W(s), s \in \mathbb{R} \quad (2)$$

where $X(s)$ is the spatial random field of location s , $W(s)$ is the white noise of s , s belongs to the two-dimensional spatial domain \mathbb{R} ; Δ is the Laplace operator $\frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2}$, k is a scale parameter, associated with the empirical range in geostatistics and jointly defined by the other parameters of the INLA-SPDE model (max.edge, cutoff, and offset) (Wu, 2021), α is a positive integer related to the smoothness parameter ν ($\alpha = \nu + d/2$, $d = 2$, representing the two-dimensional spatial) (Lindgren and Rue, 2015a; Lindgren and Rue, 2015b; Poggio et al., 2016), α was set to 2 (so that $\nu = 1$ and $d = 2$) as recommended by previous studies (Sun

et al., 2021; Whittle, 1954). In practice, the SPDE is solved by the finite element method (FEM) that approximates the continuous GF as the discrete GMRF, that is, using the Delaunay triangulation network (i.e. MESH) to denote continuous two dimension space in INLA-SPDE model (Lindgren and Rue, 2015a; Wu, 2021). The INLA-SPDE model was implemented using the “INLA” package in R language (<https://www.r-inla.org/>).

2.5.2. The procedure to implement INLA-SPDE model

(1) Construct the mesh

The INLA-SPDE model utilizes a Delaunay triangular network mesh to project the continuous space onto a discrete space, which is composed of small triangles that approximate the Matérn Gaussian field (Lindgren et al., 2011; Poggio et al., 2016). To construct the mesh set of the INLA-SPDE model in this study, several parameters, including max.edge, cutoff, and offset, needed to be set. In this study, we set the three parameters as (0.03, 0.1), 0.02, and (-0.5, 0.1), respectively (see Fig. 3). These values were determined through experiments to obtain the most regular small triangles possible (Huang et al., 2017), while balancing computational cost and modeling accuracy. Once the mesh was constructed, the spatial correlation structure of the SPDE was defined using the Matérn function (Poggio et al., 2016).

(2) Build the Bayesian hierarchical model for SOM prediction

The hierarchical model implemented in INLA-SPDE was used to predict SOM (Huang et al., 2017; Wu, 2021). This model consists of three parts: the intercept, the spatial fixed effects consisting of covariates and their coefficients matrices, and the spatial random effect of spatial locations. These three components can be expressed as follows (Arshad et al., 2020; Huang et al., 2017; Wang and Zuo, 2021; Wu, 2021):

$$\eta_k = \beta_0 + \sum_{b=1}^B \beta_b x_{bk} + \xi(s_i, s_j) \quad (3)$$

Where η_k is an additive linear estimation representing the spatially predicted soil properties; the first term β_0 is the intercept; x_{bk} is the environmental covariate of k , β_b is the coefficient of the i th environmental covariate, $\sum_{b=1}^B \beta_b x_{bk}$ represents the linear fixed effects consisting of environmental covariates; $\xi(s_i, s_j)$ is used to represent spatial random effect and can be expressed as the Matérn covariance function solved by SPDE (in section 2.5.1) (Li et al., 2018; Wang and Zuo, 2021). To fit the hierarchical model, NorSOM was used, and the likelihood distribution family for the distribution of the NorSOM was defined as the “Gaussian” distribution.

(3) Define the prior distribution of the model parameters and establish the data stack

The prior distribution of model parameters can be specified as either informative or left as the default (non-informative) (Poggio et al., 2016). Due to the lacking prior knowledge in this study area, the default parameters and hyperparameter in the “INLA” package were adopted (Li et al., 2018; Poggio et al., 2016; Wang and Zuo, 2021). To ensure the proper functioning of the INLA-SPDE model, a data stack consisting samples matrix, covariates matrix, and prediction matrix were established and joined to the model processing (Poggio et al., 2016).

(4) Select environmental covariates based on DIC

The inclusion of redundant covariate information not only adds unnecessary complexity to the model (Beguin et al., 2017) but also introduces unnecessary errors (Hastie, 2009). Therefore, selecting an

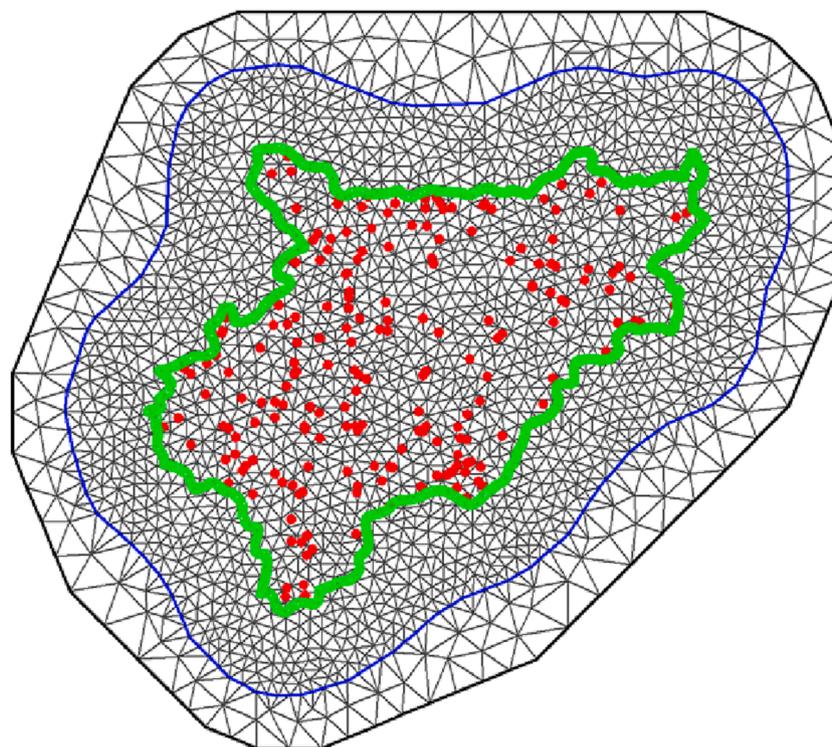


Fig. 3. The constructed mesh network for INLA-SPDE model. The red represents soil sample points, and the green represents the boundary of the study area. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

optimal set of environmental covariates is increasingly becoming an important step in developing a DSM model. The deviance information criterion (DIC), which is a commonly-used index for measuring model performance in Bayesian models (Spiegelhalter et al., 2001), is usually employed to select covariates for INLA-SPDE (Arshad et al., 2020; Huang et al., 2017; Li et al., 2018; Rue et al., 2009). Similar to Akaike's Information Criterion (AIC), a smaller DIC indicates a better model fit (Poggio et al., 2016). The specific procedure for using DIC with INLA-SPDE is as follows. First, fit an original full INLA-SPDE model with all covariates and obtain the original DIC. Second, delete each of the covariates in turn and fit INLA-SPDE again to obtain a new DIC of INLA-SPDE. Third, delete a covariate if the reduced model excluding this covariate has a lower DIC value than the original model or other reduced models; if the result is the opposite, reserve this covariate. Fourth, iterate this process through all covariates and compute DIC in turn. Finally, obtain the final set of covariates that deletes any of the remaining covariates generated models with a larger DIC value than other models in the same processing turn (Arshad et al., 2020; Huang et al., 2017; Li et al., 2018).

2.5.3. Quantifying uncertainty in INLA-SPDE model

The Bayesian approach provides the posterior marginal distribution to quantify the spatial uncertainty of predicted soil properties. This distribution directly represents the probability of the predicted value falling within a credibility interval (Huang et al., 2017; Poggio et al., 2016). In our study, the mean of the predicted SOM in INLA-SPDE modeling output was used as the predicted SOM result (Fig. 8), and the uncertainty of the predicted SOM was reported using two posterior density predictive distribution intervals, namely the 2.5% and 97.5% percentiles, which were derived using the INLA-SPDE model. The 2.5% percentiles and 97.5% percentiles, correspond to the lower limit (LL) and upper limit (UL) of the prediction, respectively. Based on previous studies (Arshad et al., 2020; Li et al., 2018; Wu, 2021), to describe the spatial distribution of uncertainty, we used the 95% highest posterior density credibility interval (95% HPD CI) and the relative width of the

posterior interquartile range (RWPIR) (Wu, 2021; Yuan et al., 2016), which were determined using the following formula ((4) and (5)).

$$95\%HPDCI = 97.5\%percentiles - 2.5\%percentiles \quad (4)$$

$$RWPIR = \frac{97.5\%percentiles - 2.5\%percentiles}{mean} \quad (5)$$

2.6. Random forest as a comparison model

The Random Forest (RF) model was utilized as a reference model. RF is a robust machine-learning tool based on the ensemble of decision trees and is widely used in soil properties prediction (Taghizadeh-Mehrjardi et al., 2020; Tian et al., 2022; Yang et al., 2020; Zhang et al., 2021). RF model has several advantages, including higher tolerance of outliers and noise, and being less prone to overfitting (Breiman, 2001). Previous studies have demonstrated that the RF model often has a better predictive performance than other models (Camera et al., 2017; Pahlavan-Rad and Akbarimoghaddam, 2018).

In RF model, the number of randomly selected predictors for each tree building (m_{try}) and the number of trees to be learned in a forest (n_{tree}) are important parameters. m_{try} was set according to one-third of the total number of model variables (Xia and Zhang, 2022), and n_{tree} was set to 1000, based on several iterations experiments (Yang et al., 2022). Recursive Feature Elimination (RFE) was utilized to select and filter the optimal combination of covariates for the RF model (Chen et al., 2021; Yang et al., 2022). To ensure fairness in model comparison, we selected the best covariates with the highest prediction accuracy for each covariates combinations using DIC-based selection for INLA-SPDE and RFE-based selection for RF, respectively.

To assess the uncertainty of the RF model, we utilized the quantile regression forest (QRF) to estimate the prediction uncertainty (Meinshausen, 2006). This approach calculates quantiles of the conditional probability distribution for each prediction location, which collectively define the distribution at that location (Takoutsing and Heuvelink, 2022). We obtained the 2.5% and 97.5% percentiles from the QRF

model, similar to INLA-SPDE, and computed the 95% highest posterior density (HPD) credible interval and random walk predictive interval regression (RWPIR) using the same method (see section 2.5.3).

We employed the “randomForest” package in the R language to perform the RF models, the “caret” package for the RFE process, and the “ranger” package to execute the QRF models for quantifying uncertainty.

2.7. Evaluation of the predicted soil maps

Ten-fold cross-validation was used to evaluate the accuracy of the predicted SOM using two models with different combinations of covariates. All samples are randomly divided into ten subsets, and in each repetition, nine of them are taken as the training data set and the remaining one as the validation data. The average of these 10 results was used as the final evaluation results. Three commonly used accuracy verification indices, coefficient of determination (R^2), Lin’s concordance correlation coefficient (LCCC) and root mean square error (RMSE, $g\ kg^{-1}$) were used to measure the performance of different covariates combinations in different models. The three indices were calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{6}$$

$$LCCC = \frac{2r\sigma_y\sigma_{\hat{y}}}{\sigma_y^2 + \sigma_{\hat{y}}^2 + (\bar{y} - \bar{\hat{y}})^2} \tag{7}$$

$$RMSE = \frac{1}{\sqrt{n}} \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{8}$$

Where n is the sample size of validation points, y_i and \hat{y}_i is observed and predicted SOM value at the corresponding validation points, r is the correlation coefficient between y_i and \hat{y}_i , \bar{y} and $\bar{\hat{y}}$, σ_y and $\sigma_{\hat{y}}$ are the average and variance of y (observed SOM) and \hat{y} (predicted SOM). To explore the effect of covariate selection on model accuracy, we computed the model accuracy before covariates selection (all covariates in corresponding covariates group) and after covariates selection for each group of covariates.

3. Results

3.1. The descriptive statistics of SOM and its normally distributed transformation

Fig. 4 displays the histograms of SOM and NorSOM, which is SOM after a normally distributed transformation. The original SOM exhibits moderate variation ($CV = 47.1\%$), suggesting that the soil distribution in the study area is heterogeneous. Additionally, the original SOM is left-skewed. After a logarithmic conversion to achieve normal distribution, the NorSOM appears to be approximately normally distributed (Fig. 4b).

3.2. The generated NSDSI

We selected the image on February 21st as the final remote sensing image to calculate the soil moisture indices based on the principles outlined in Section 2.4.2. The generated NSDSI1, NSDSI2 and NSDSI3 are shown in Fig. 5. The spatial distribution of the three indices are similar, especially NSDSI1 and NSDSI3. This is mainly because these three indices only have similar numerators. However, the indices show different variations over the study area. Generally, the southern hilly area and areas in the east part have larger soil moisture. The correlation coefficient between NSDSI1, NSDSI2 and NSDSI3 and NorSOM was 0.16, 0.18 and 0.17 (Fig. 6, significant at 0.05 level), indicating they may provide useful information for SOM prediction.

3.3. The correlation between SOM and environmental covariates

The correlation coefficients between SOM and environmental covariates are presented in Fig. 6. Most of the environmental covariates, except TWI, Amp1, Pha1, and Pha2, showed a significant correlation with SOM at a 0.05 level. Although the correlation coefficients between natural environmental covariates and SOM are generally higher than those of NSDSI and FTD covariates, all three NSDSI covariates showed a significant positive correlation with SOM, which indicated that wetter soils have higher SOM in this area, and three of the six FTD covariates correlated with SOM. Furthermore, the correlation between NorSOM and NSDSI and FTD covariates was slightly higher than that between the original SOM and NSDSI and FTD covariates.

3.4. The covariates selection for INLA-SPDE and RF

Table 2 shows the selected covariates for INLA-SPDE and RF models

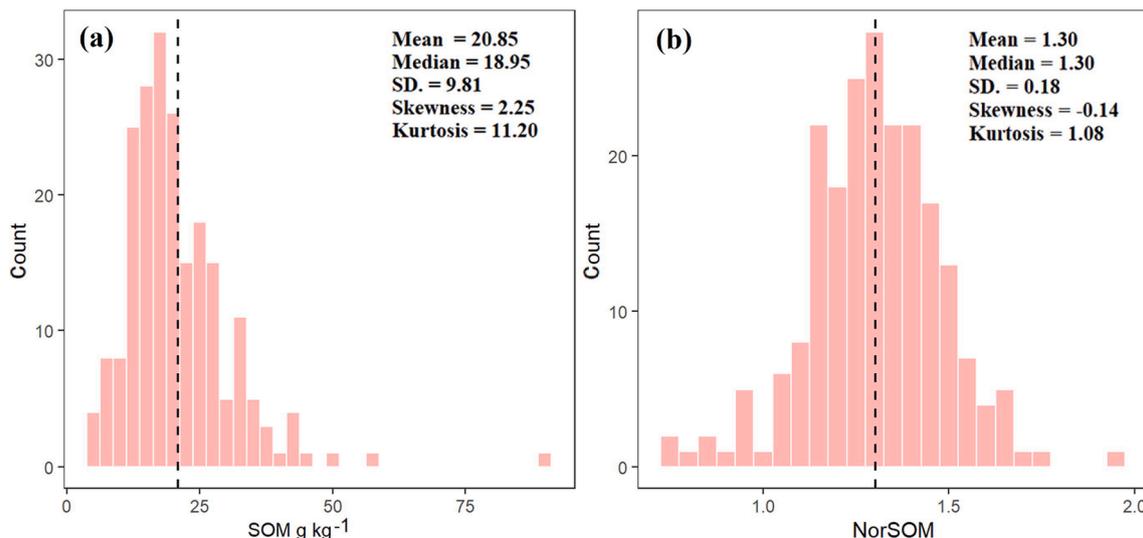


Fig. 4. Distributions of original SOM content (4a) and the logarithmic transformed SOM content (NorSOM) (4b).

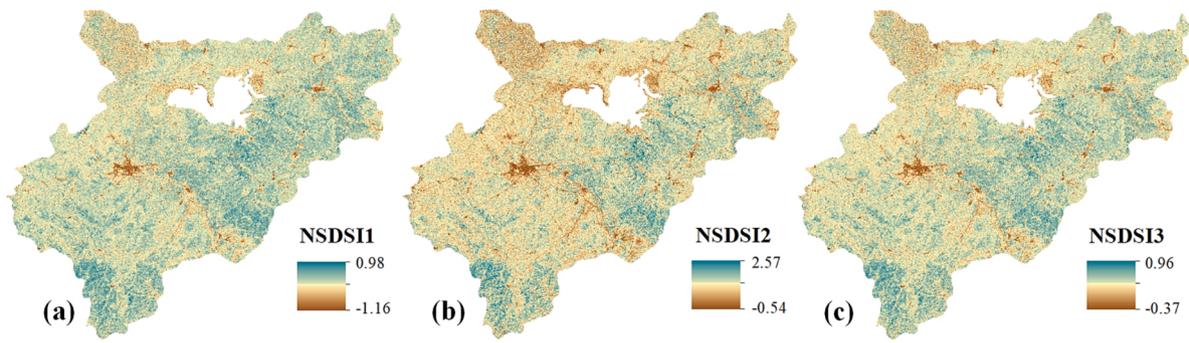


Fig. 5. The spatial distribution of NSDSI1, NSDSI2, and NSDSI3 (a, b, c).

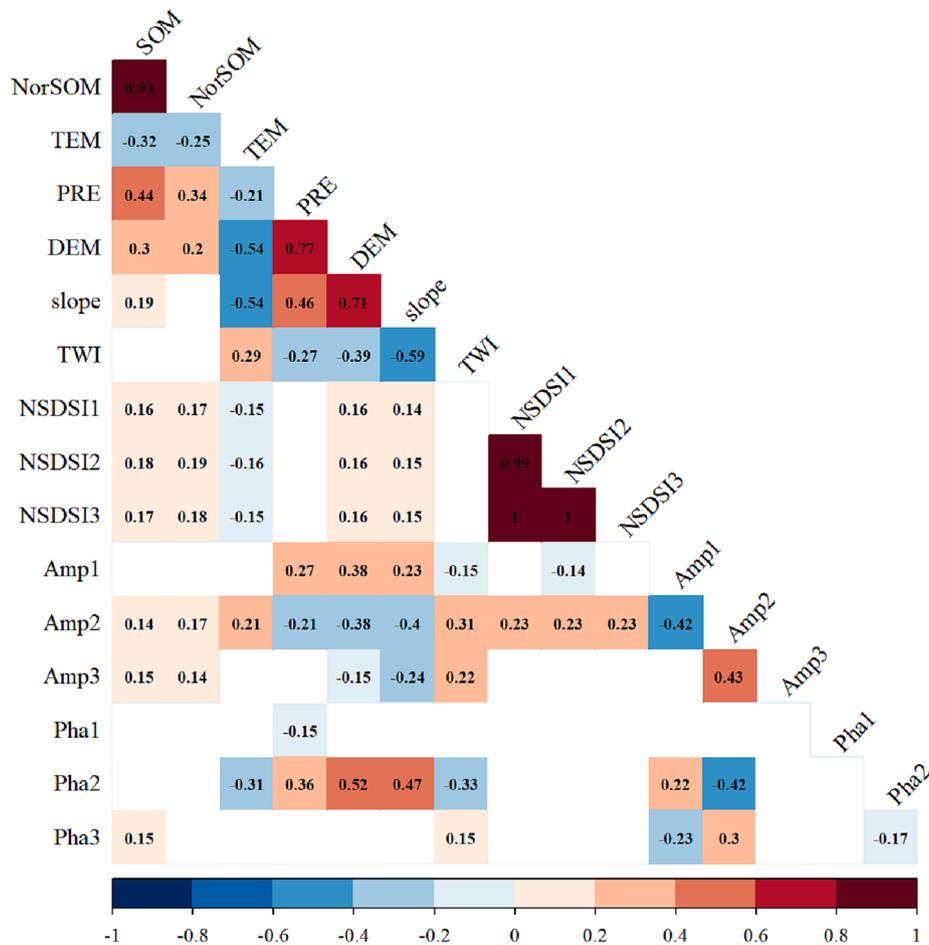


Fig. 6. The correlations between SOM after normally distributed transformation (NorSOM) and natural, FTD and NSDSI environmental covariates. Note: the blank cells indicate that the correlation is not significant at a 0.05 level; Pearson’s correlation was used for original covariates or after normal distribution transformation.

in SOM mapping respectively. It shows that the selected variables for INLA-SPDE and RF are similar. Table 2 displays the covariates selected for the INLA-SPDE and RF models in SOM mapping, respectively. The table indicates that the selected variables for both models are similar. Among the natural environment covariates associated with topography, INLA-SPDE chose the slope while RF chose the DEM, and in phase covariates of FTD, INLA-SPDE chose the first phase Pha1, RF chose the third phase Pha3.

3.5. The prediction accuracies for the INLA-SPDE and RF models with different environmental covariates

Fig. 7 displays the cross-validation outcomes for both models, using

various combinations of environmental covariates. The results indicate that incorporating NSDSI or FTD covariates enhances mapping performance, regardless of whether INLA-SPDE or RF. Compared to the group of “N” which solely included natural environmental covariates, the “N + NSDSI”, “N + FTD”, and “N + NSDSI + FTD” combinations improved the mapping accuracy by 11%, 17%, and 25% in terms of R^2 , 5%, 18%, 21% in terms of LCCC, 4%, 6%, and 11% in term of RMSE when using INLA-SPDE, respectively. The use of FTD covariates resulted in slightly greater improvement than the use of NSDSI. The environmental covariates combination “N + NSDSI + FTD” generally performed the best among all the combinations, indicating that using both NSDSI and FTD covariates showed a more promising predictive power. Additionally, INLA-SPDE outperformed RF in SOM prediction with either covariates

Table 2

The selected covariates after DIC and RFE covariates selection of INLA-SPDE and RF. N: natural environmental covariates; NSDSI: RS-based soil moisture indices; FTD: Fourier transforms decomposed variables.

Model	Combination code	Covariates
INLA-SPDE	N	PRE, slope, ParentL
	N + NSDSI	PRE, slope, ParentL, NSDSI1, NSDSI2, NSDSI3
	N + FTD	PRE, slope, ParentL, Amp2, Amp3, Pha1
	N + NSDSI + FTD	PRE, slope, ParentL, NSDSI1, NSDSI2, NSDSI3, Amp2, Amp3, Pha1
RF	N	PRE, DEM, ParentL
	N + NSDSI	PRE, DEM, ParentL, NSDSI1, NSDSI2, NSDSI3
	N + FTD	PRE, DEM, ParentL, Amp2, Amp3, Pha3
	N + NSDSI + FTD	PRE, DEM, ParentL, NSDSI1, NSDSI2, NSDSI3, Amp2, Amp3, Pha3

combination. Compared to RF, INLA-SPDE had higher accuracies of 11%, 19%, 19%, and 18% in terms of R^2 using covariates combinations “N”, “N + NSDSI”, “N + FTD”, and “N + NSDSI + FTD”, respectively. In addition, the results demonstrate that the selection of covariates enhanced the prediction accuracy for both INLA-SPDE and RF models. Especially, after covariates selection, R^2 of the INLA-SPDE model were improved by 8%, 13%, 10%, and 17% with covariates combinations “N”, “N + NSDSI”, “N + FTD”, and “N + NSDSI + FTD”, respectively.

3.6. The predicted SOM maps

Fig. 8 displays the spatial distribution of topsoil organic matter content predicted using INLA-SPDE and RF. The predicted SOM using INLA-SPDE and RF models with each covariates combination, exhibited a consistent spatial pattern, especially when using the same covariates. Generally, the southern mountainous regions with high elevation, high precipitation, low temperature, and gentle plains in the northwest areas had a relatively higher predicted SOM content, compared to other locations. However, differences obviously existed in the spatial predicted SOM using different environmental combinations or models, and using more predictors produced more details in the predicted SOM maps. Maps of the difference between INLA-SPDE and RF showed that INLA-SPDE predicts higher SOM in some plain areas and some hilly areas in south, but lower SOM in other mountains in south. Furthermore, INLA-SPDE model produced a wider range of the predicted minimum and maximum values in comparison to RF. For instance, in combination “N + NSDSI + FTD”, the predicted range of SOM for INLA-SPDE was from 3.35 $g\ kg^{-1}$ to 77.06 $g\ kg^{-1}$, while the range for RF was from 9.33 $g\ kg^{-1}$ to 37.94 $g\ kg^{-1}$. The predicted range of INLA-SPDE is more similar to the SOM of samples with a range from 4.37 $g\ kg^{-1}$ to 89.12 $g\ kg^{-1}$.

3.7. The generated uncertainty maps

Fig. 9 displays the spatial distributions of the 95% HPD CI and the RWPIR for the INLA-SPDE and RF models using the best combination of covariates (N + FTD + NSDSI). The results for other covariate combinations are similar. The maps illustrate that the two models have different spatial patterns for their 95% HPD CI and RWPIR. INLA-SPDE has a wider CI and RWPIR than RF. Generally, southern hilly areas and plains in the northwest have higher CI values for INLA-SPDE, while southern areas have higher CI values for RF. As for RWPIR with INLA-SPDE, it shows a clear edge effect and mesh effect, which means the RWPIR in the boundary areas of this study area was generally higher, and higher RWPIR occurred in the vertices of small triangles in the generated mesh of SPDE.

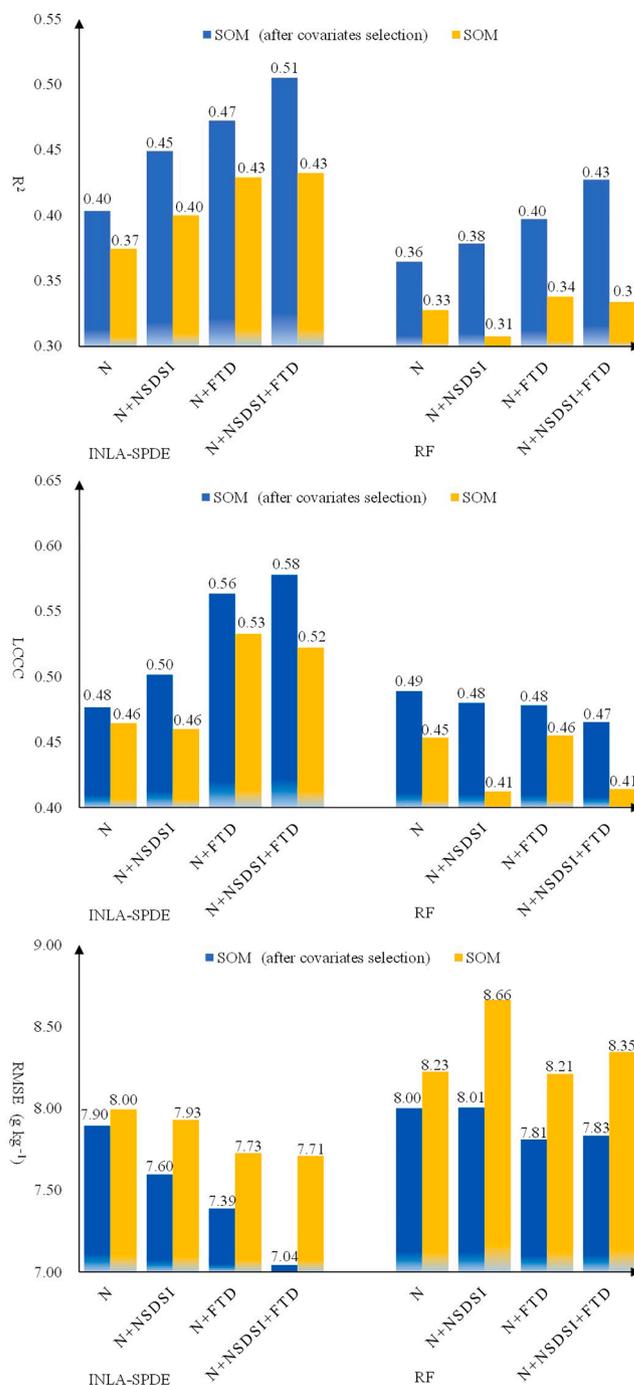


Fig. 7. Modeling performance of INLA-SPDE and RF with different covariates combinations (N, N + NSDSI, N + FTD, and N + NSDSI + FTD). Notes: the blue bar graph represents the validation results of covariates combinations after covariates selection. N: natural environmental covariates; NSDSI: RS-based soil moisture indices; FTD: Fourier transforms decomposed variables. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

4.1. Application of INLA-SPDE model for DSM

In our study, the INLA-SPDE model demonstrated higher prediction accuracy compared to the RF model, which is consistent with the findings of Beguin et al. (2017) and Giannini Kurina et al. (2019). The mapping accuracy of RF model for SOM is similar to the results reported

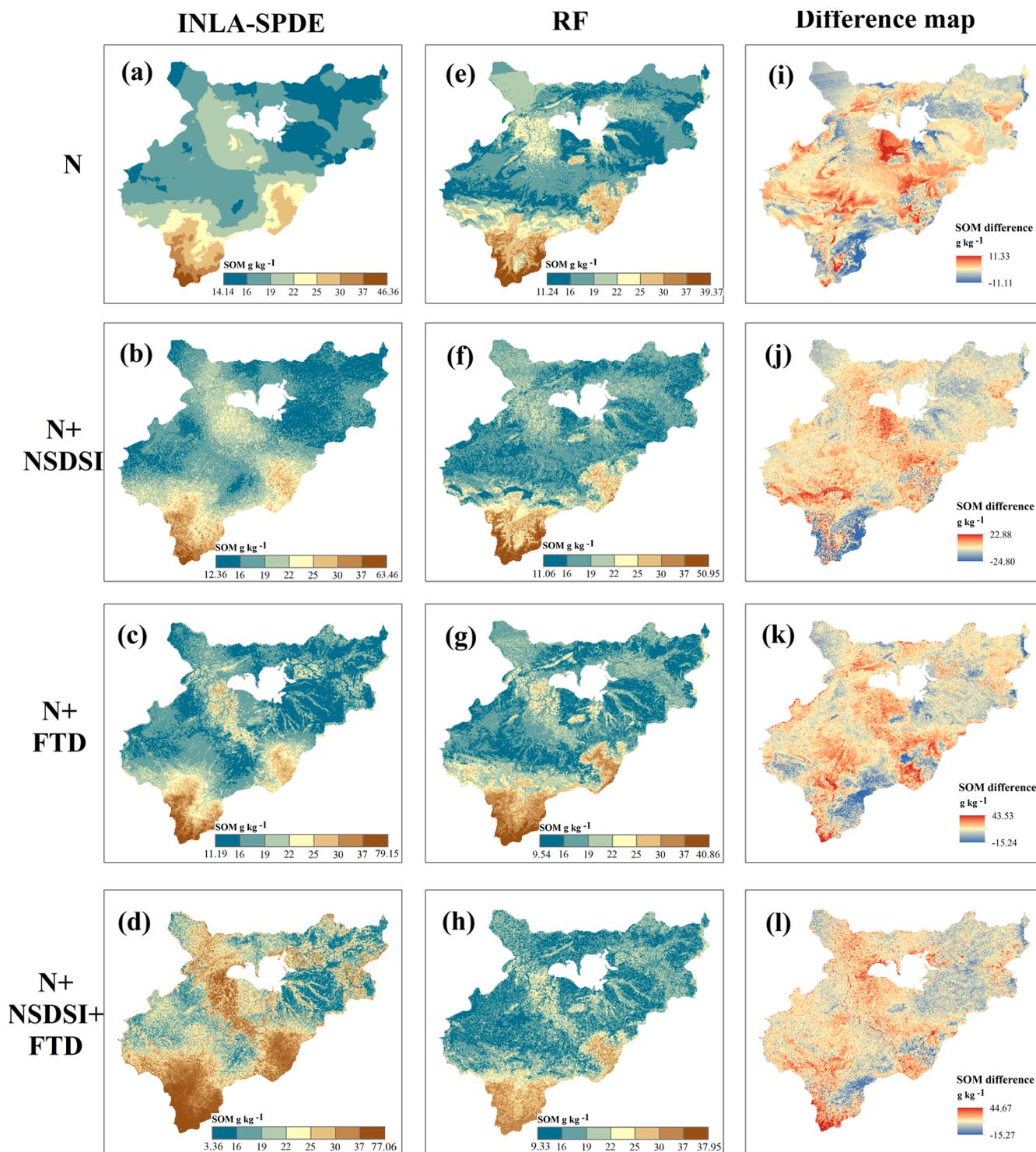


Fig. 8. The predicted SOM maps using INLA-SPDE (a, b, c, d) and RF (e, f, g, h) models (different columns) with different covariates combination (N, N + NSDSI, N + FTD, and N + NSDSI + FTD) and their difference map (INLA-SPDE minus RF, i, j, k, l). N: natural environmental covariates; NSDSI: RS-based soil moisture indices; FTD: Fourier transforms decomposed variables.

by Yang et al. (2019) with RF model using FTD and crop rotation covariates (with an accuracy of $R^2 = 0.43$). The main reason for the higher accuracy of INLA-SPDE is that it considers not only the relationships between soil and environment, but also the geographical locations of the measurements which are ignored by RF (Heuvelink and Webster, 2022). The semivariogram analysis of SOM indicated a significant spatial dependency area, with a nugget-to-sill ratio of 21.5% ($C0 = 0.0404$, $C = 0.1474$, $C0/(C0 + C) = 21.5\%$, exponential model).

The use of Matérn functions in INLA-SPDE allows for the incorporation of spatial dependency information in SOM data. These results may be related to that the INLA-SPDE is a generalized additive model (Banerjee et al., 2014; Beguin et al., 2017), so this model may be sensitive to collinearity and autocorrelation between environmental covariates (Wang et al., 2018).

Different from the Frequentist approach, which calculates uncertainty from a large number of realizations, the INLA-SPDE method,

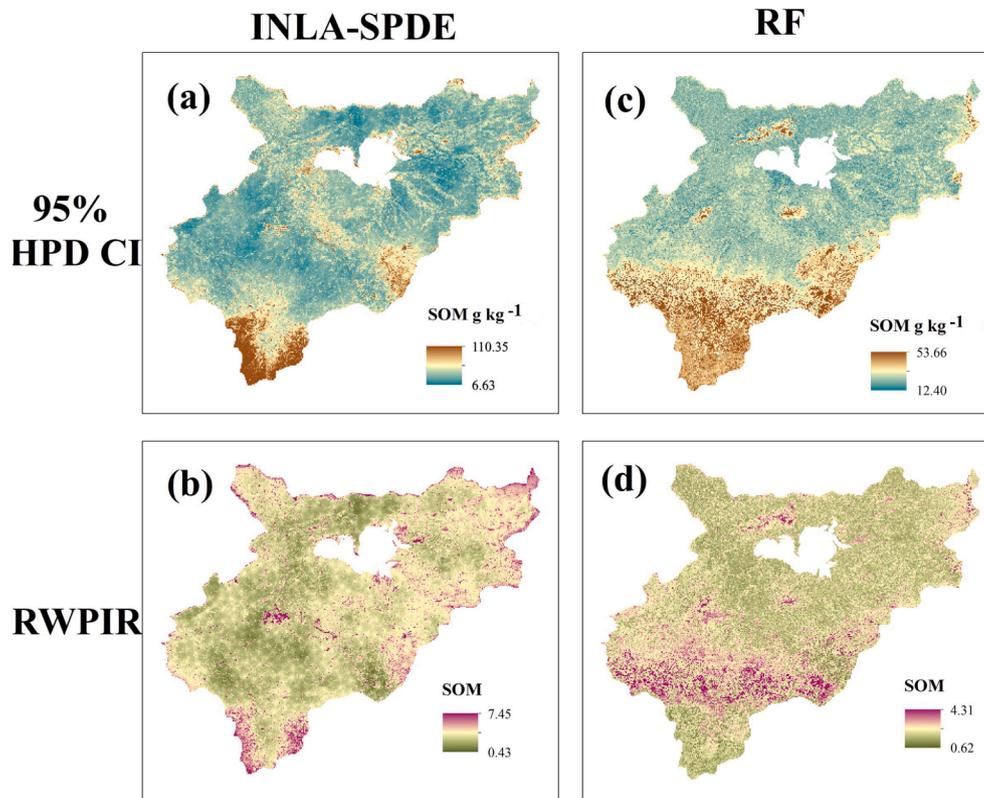


Fig. 9. The maps of 95% HPD CI and RWPIR for INLA-SPDE (a and b) and RF (c and d, based on QRF) model.

based on the Bayesian theorem, can describe uncertainty based on the posterior density of prediction explicitly (Poggio et al., 2016). In our study, the higher uncertainty in the uncertainty map generated by INLA-SPDE is mainly caused by the model calculation method way and limited soil samples. High RWPIR in boundaries is the so-called edge effect (Huang et al., 2017) and high uncertainties occur in nodes of small triangulation in the mesh (Fig. 9(b)). Uncertainty also may be higher in undersampled areas (sparse sample points), which is similar to regression kriging (RK) (Takoutsing and Heuvelink, 2022). In this case, uncertainty information can be used to guide additional soil sampling for improving soil prediction accuracies, for instance, Blackford et al. (2022) and Stumpf et al. (2017) developed a method for uncertainty-guided soil supplemental sampling and ultimately improving cartographic accuracy. In the QRF model, uncertainty quantification based on the Frequentist approach is derived from the ensemble of all regression tree results, resulting in smoother mapping results (Meinshausen, 2006). However, the INLA-SPDE method utilizes joint posterior probability density functions (pdfs) based on Bayesian theorem, which facilitates the incorporation of predictive model uncertainty (Huang et al., 2017; Poggio et al., 2016).

When using INLA-SPDE, it is important to consider the time cost. Although the modeling time of INLA-SPDE is significantly reduced compared to MCMC (Poggio et al., 2016; Wu, 2021), it still requires more time compared to some commonly used models such as RF. Beguin et al. (2017) demonstrated that the time spent on the INLA-SPDE model was approximately 134 times greater than that of RF. In this study, the time difference is about 30 times. If the model complexity increases (such as more small triangles in MESH or more covariates, large areas), the time cost and computing ability requirements will increase exponentially.

Further experiments are necessary to fully develop the potential of the INLA-SPDE model in DSM. For instance, researchers can investigate the MESH settings to balance mapping accuracy and time spent, as well as conduct a quantitative evaluation of uncertainty results. Additionally,

Sun et al. (2021) demonstrated that it is possible to simulate spatial-temporal changes in soil properties using INLA-SPDE if multi-temporal soil sample points are available. Further applications should be conducted to explore the effectiveness of INLA-SPDE in spatial-temporal modeling.

4.2. Application of RS-based soil moisture indices covariates (NSDSIs) in SOM mapping

We utilized the RS-based soil moisture indices (NSDSIs) proposed by Yue et al. (2019) and demonstrated a significant positive correlation between these indices and SOM (Fig. 6), which is consistent with previous research by Qin et al. (2022) and (Nocita et al., 2013). Furthermore, we found that these indices were effective in predicting SOM, as illustrated in Fig. 7. This is because soil moisture is a key environmental factor affecting the process of dissolution, mineralization, synthesis, and material migration of SOM (Li et al., 2022; Li et al., 2019; Yoshida et al., 2018). For instance, Cates et al. (2022) suggested that a greater relative abundance of stable complex compounds in moist soil environments protects the SOM from decomposition. Shabtai et al. (2022) demonstrated that high-moisture soil can alter soil minerals, chemical compositions, and organic-mineral interactions, promoting SOM accumulation. Soil moisture can also indirectly influence SOM by affecting vegetation growth and consequently carbon input (Han et al., 2022). In addition, soil moisture may be related to soil microbial and enzymatic activities, thereby indirectly affecting SOM (Pushkareva et al., 2020; Steinweg et al., 2012). Studies that utilize soil moisture as an environmental covariate to map soil properties are scarce. While previous soil moisture indices (such as WISOIL and NSMI) have been developed, they cannot be applied to conventional multispectral broadband satellites (e.g., Landsat and Sentinel) (Yue et al., 2019). Our study demonstrates that optical remote sensing-based soil moisture indices offer a straightforward and efficient approach to extract soil moisture information as useful predictors for soil mapping.

Yue et al. (2019) utilized the NSDSI indices to examine bare soil regions. The authors suggested that the indices are suitable for areas with low vegetation covers, such as deserts and ploughed farmland. However, they also noted that other surface features, such as green vegetation, exhibit comparable water absorption characteristics between two SWIR bands. To estimate soil moisture in regions with varying levels of vegetation coverage, soil moisture indices combined with NDVI may be employed using the trapezoid feature method (Sadeghi et al., 2017). In our study area, the vegetation coverage is non-uniform due to the presence of few broad-leaved forests in high elevation areas, despite the use of a winter satellite image. Future research can focus on developing NSDSI-NDVI for application.

Yue et al. (2019) pointed out that the variations in NSDSIs between the four soils in their study area are very small when SM is less than 50%, while all SM indices perform poorly when SM is greater than 50%. This may be related to the difference in saturated soil moisture content and rate of different soil types. The main soil type for farmlands in our study area is mainly Ferric-Udic Argosols and Stagnic Anthrosols, and we found that there is no significant difference between NSDSIs of the two soil types. More experiments can be done for areas with multiple soil types to examine how effective the indices are.

The modeling validation results suggest that using all three soil moisture indices (NSDSI1, NSDSI2 and NSDSI3) obtain the best predicting accuracy of SOM (Table.2, Fig. 7, $R^2 = 0.51$, RMSE = 7.04 g kg⁻¹). Yet, when using one of the three soil moisture indices alone, NSDSI2 which has the highest correlation with SOM/NorSOM (Fig. 6) obtains the highest mapping accuracy ($R^2 = 0.50$, RMSE = 7.28 g kg⁻¹), although the improvement compared with NSDSI1 or NSDSI3 is very slight. We recommend that when computational resources are adequate, all soil moisture indices should be utilized as input to provide a variety of information for soil prediction. In cases where computational resources are limited or a concise model is preferred, it may be better to select the soil moisture index that exhibits the highest correlation with the target soil property.

Researchers have devoted significant effort to developing effective soil moisture indices using remote sensing (RS). For instance, Tian et al. (2021) developed a drying process segmentation method to accurately estimate soil moisture from spectral reflectance data and proposed the Shortwave Normalized Index (SNI). However, the SNI index requires experimental measurements of soil samples and is currently not applicable for large-scale soil moisture estimation based on RS images. Nevertheless, emerging indices may have the potential to serve as effective predictors in digital soil mapping (DSM) in future research.

4.3. Application of FTD covariates in SOM mapping

The FTD covariates are generated to detect periodic patterns in time-series vegetation indices with an objective method (Fourier transform). The reason that using the FTD covariates improves the SOM mapping accuracies is mainly because the covariates are related to vegetation growth features. Furthermore, it is suggested that FTD covariates can separate different crop species and crop rotations when combined with expert knowledge and field investigation (Yang et al., 2019). In the study of Yang et al. (2019), the FTD covariates were only adopted in farmland. In our study, we used the FTD covariates in the whole area with farmland, forestland, and grassland. It indicates that the FTD covariates can be applicable in land covered with vegetation. When applying Fourier transform on time series vegetation indices to generate FTD covariates, one important issue is that time series images should cover key phenological dates or periods of different vegetation species in all the land uses. Therefore, it is better to adopt remote sensing images with high temporal resolutions, such as Landsat satellite with a temporal resolution of 16 days, and Sentinel-2 with 5 days. In case of cloud contamination within remote sensing images, multi-source images or data fusion may be an option to complete a time series profile for describing key growth characteristics of all vegetation types in a study

area.

5. Conclusion

In this study, we examine the effectiveness of RS-based soil moisture indices (NSDSI) and Fourier transforms decomposed (FTD) covariates on mapping SOM with INLA-SPDE in a study area in central and eastern China. The results show that using NSDSI or FTD covariates improved SOM mapping accuracies, and INLA-SPDE model exhibited higher mapping accuracies than RF. Compared with environmental combination including only natural environmental covariates, the best combination “N + FTD + NSDSI” which contains natural environmental covariates, FTD covariates, and soil moisture indices improved the SOM accuracy by 23% in terms of R^2 and 9% in terms of RMSE. The study provides new promising predictors for SOM mapping and proved the effectiveness of a Bayesian spatial model, INLA-SPDE. In this paper, we only calculated NSDSI indices in winter to represent more bare soil spectral information. Further explorations on NSDSI indices in other seasons or time-series NSDSI indices can be applied.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This study is supported by the National Key Research and Development Program Plan (Grant No. 2022YFC3800802), the National Natural Science Foundation of China (Project No. 41971054) and the Fundamental Research Funds for the Central Universities (0209-14380115).

References

- Arrouays, D., Mulder, V.L., Richer-de-Forges, A.C., 2021. Soil mapping, digital soil mapping and soil monitoring over large areas and the dimensions of soil security – A review. *Soil Secur.* 5, 100018 <https://doi.org/10.1016/j.soisec.2021.100018>.
- Arshad, M., Li, N., Bella, L.D., Triantafyllis, J., 2020. Field-scale digital soil mapping of clay: Combining different proximal sensed data and comparing various statistical models. *Soil Sci. Soc. Am. J.* 84 (2), 314–330.
- Banerjee, S., Gelfand, A.E., Finley, A.O., Sang, H., 2008. Gaussian predictive process models for large spatial data sets. 70(4), 825–848. <https://doi.org/10.1111/j.1467-9868.2008.00663.x>.
- Banerjee, S., Carlin, B., Gelfand, A., 2014. Hierarchical Modeling and Analysis for Spatial Data. <https://doi.org/10.1201/b17115>.
- Beguín, J., Fuglstad, G.-A., Mansuy, N., Paré, D., 2017. Predicting soil properties in the Canadian boreal forest with limited data: Comparison of spatial and non-spatial statistical approaches. *Geoderma* 306, 195–205. <https://doi.org/10.1016/j.geoderma.2017.06.016>.
- Blackford, C., Heung, B., Webster, K.L., 2022. Incorporating spatial uncertainty maps into soil sampling improves digital soil mapping classification accuracy in Ontario, Canada. *Geoderma Regional* 29, e00495.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Camera, C., Zomeni, Z., Noller, J.S., Zissimos, A.M., Christoforou, I.C., Bruggeman, A., 2017. A high resolution map of soil types and physical properties for Cyprus: A digital soil mapping optimization. *Geoderma* 285, 35–49. <https://doi.org/10.1016/j.geoderma.2016.09.019>.
- Cates, A.M., Jilling, A., Tfaily, M.M., Jackson, R.D., 2022. Temperature and moisture alter organic matter composition across soil fractions. *Geoderma* 409, 115628. <https://doi.org/10.1016/j.geoderma.2021.115628>.
- Chen, Y., Lu, D., Moran, E., Batistella, M., Dutra, L.V., Sanches, I.D.A., da Silva, R.F.B., Huang, J., Luiz, A.J.B., de Oliveira, M.A.F., 2018. Mapping croplands, cropping patterns, and crop types using MODIS time-series data. *Int. J. Appl. Earth Obs. Geoinf.* 69, 133–147. <https://doi.org/10.1016/j.jag.2018.03.005>.
- 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. <https://doi.org/10.1016/j.catena.2020.105062>.

- 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., 2022. Digital mapping of GlobalSoilMap soil properties at a broad scale: A review. *Geoderma* 409, 115567. <https://doi.org/10.1016/j.geoderma.2021.115567>.
- Fabre, S., Briottet, X., Lesaignoux, A., 2015. Estimation of soil moisture content from the spectral reflectance of bare soils in the 0.4–2.5 μm domain. *ISPRS Int. J. Geo-Inf.* 15(2), 3262–3281.
- Giannini Kurina, F., Hang, S., Macchiavelli, R., Balzarini, M., 2019. Spatial predictive modelling essential to assess the environmental impacts of herbicides. *Geoderma* 354, 113874. <https://doi.org/10.1016/j.geoderma.2019.07.032>.
- Han, L., Chang, Y., Chen, R., Liu, Z., Zhao, Y., Zhu, H., Zhao, Z., Gao, Y., Yang, M., Li, Y., Cao, X., 2022. Response of soil moisture to vegetation and trade-off analysis in the hilly area of the Loess Plateau, China. *Ecol. Indic.* 142, 109273 <https://doi.org/10.1016/j.ecolind.2022.109273>.
- Hastie, T., 2009. The elements of statistical learning: data mining, inference, and prediction. <https://doi.org/10.1007/978-0-387-84858-7>.
- He, X., Yang, L., Li, A., Zhang, L., Shen, F., Cai, Y., Zhou, C., 2021. Soil organic carbon prediction using phenological parameters and remote sensing variables generated from Sentinel-2 images. *CATENA* 205, 105442. <https://doi.org/10.1016/j.catena.2021.105442>.
- Heuvelink, G.B.M., Webster, R., 2022. Spatial statistics and soil mapping: A blossoming partnership under pressure. *Spatal Stat.* 50, 100639 <https://doi.org/10.1016/j.spasta.2022.100639>.
- Huang, J., Malone, B.P., Minasny, B., McBratney, A.B., Triantafyllis, J., 2017. Evaluating a Bayesian modelling approach (INLA-SPDE) for environmental mapping. *Sci. Total Environ.* 609, 621–632. <https://doi.org/10.1016/j.scitotenv.2017.07.201>.
- Hutchinson, M.F., 1998. Interpolation of rainfall data with thin plate smoothing splines: II. Analysis of topographic dependence. *J. Geogr. Inf. Decis. Anal.* 2, 168–185.
- Jahandideh Mahjenabadi, V.A., Roohollah Mousavi, S., Rahmani, A., Karami, A., Asadi Rahmani, H., Khavazi, K., Rezaei, M., 2022. Digital mapping of soil biological properties and wheat yield using remotely sensed, soil chemical data and machine learning approaches. *Comput. Electron. Agric.* 197, 106978 <https://doi.org/10.1016/j.compag.2022.106978>.
- Johnston, A.E., Poulton, P.R., Coleman, K., 2009. Chapter 1 Soil Organic Matter: Its Importance in Sustainable Agriculture and Carbon Dioxide Fluxes. In: D.L. Sparks (Ed.), *Adv. Agron. Academic Press*, pp. 1–57. [https://doi.org/10.1016/S0065-2113\(08\)00801-8](https://doi.org/10.1016/S0065-2113(08)00801-8).
- Kerry, R., Goovaerts, P., Rawlins, B.G., Marchant, B.P., 2012. Disaggregation of legacy soil data using area to point kriging for mapping soil organic carbon at the regional scale. *Geoderma* 170, 347–358. <https://doi.org/10.1016/j.geoderma.2011.10.007>.
- Kopecký, M., Macek, M., Wild, J., 2021. Topographic Wetness Index calculation guidelines based on measured soil moisture and plant species composition. *Sci. Total Environ.* 757, 143785 <https://doi.org/10.1016/j.scitotenv.2020.143785>.
- Lal, R., 2020. Food security impacts of the “4 per Thousand” initiative. *Geoderma* 374, 114427. <https://doi.org/10.1016/j.geoderma.2020.114427>.
- Li, Z., Huang, M., Luo, N., Wen, J., Deng, C., Yang, R., 2019. Spectroscopic study of the effects of dissolved organic matter compositional changes on availability of cadmium in paddy soil under different water management practices. *Chemosphere* 225, 414–423. <https://doi.org/10.1016/j.chemosphere.2019.03.059>.
- Li, H., Van den Bulcke, J., Mendoza, O., Deroo, H., Haesaert, G., Dewitte, K., De Neve, S., Sleutel, S., 2022. Soil texture controls added organic matter mineralization by regulating soil moisture—evidence from a field experiment in a maritime climate. *Geoderma* 410, 115690. <https://doi.org/10.1016/j.geoderma.2021.115690>.
- Li, N., Zare, E., Huang, J., Triantafyllis, J., 2018. Mapping soil cation-exchange capacity using Bayesian modeling and proximal sensors at the field scale. *Soil Sci. Soc. Am. J.* 82 (5), 1203–1216.
- Lindgren, F., Rue, H., 2015a. Bayesian spatial modelling with R-INLA. *J. Stat. Softw.* 63 (19), 1–25. <https://doi.org/10.18637/jss.v063.i19>.
- Lindgren, F., Rue, H., 2015b. Bayesian spatial modelling with R-INLA. *J. Stat. Softw.* 63 (19). <http://hdl.handle.net/10.1016/j.geoderma.2016.09.024>.
- Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *J. R. Stat. Soc. Series B* 73 (4), 423–498.
- Martins, T.G., Simpson, D., Lindgren, F., Rue, H., 2013. Bayesian computing with INLA: New features. *Comput. Stat. Data Anal.* 67, 68–83. <https://doi.org/10.1016/j.csda.2013.04.014>.
- 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. <https://doi.org/10.1016/j.geoderma.2016.09.024>.
- McBratney, A., Mendonça Santos, M., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117, 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4).
- McBratney, A., Field, D.J., Koch, A., 2014. The dimensions of soil security. *Geoderma* 213, 203–213. <https://doi.org/10.1016/j.geoderma.2013.08.013>.
- Meinshausen, N., 2006. Quantile regression forests. *J. Mach. Learn. Res.* 7, 983–999. <https://dl.acm.org/doi/10.5555/1248547.1248582>.
- Minasny, B., McBratney, A.B., Malone, B.P., Wheeler, I., 2013. Chapter One - Digital mapping of soil carbon. In: D.L. Sparks (Ed.), *Adv. Agron. Academic Press*, pp. 1–47. <https://doi.org/10.1016/B978-0-12-405942-9.00001-3>.
- Minasny, B., McBratney, A.B., 2005. The Matérn function as a general model for soil variograms. *Geoderma* 128 (3), 192–207. <https://doi.org/10.1016/j.geoderma.2005.04.003>.
- Minasny, B., McBratney, A.B., 2016. Digital soil mapping: A brief history and some lessons. *Geoderma* 264, 301–311. <https://doi.org/10.1016/j.geoderma.2015.07.017>.
- Minasny, B., Vrugt, J.A., McBratney, A.B., 2011. Confronting uncertainty in model-based geostatistics using Markov Chain Monte Carlo simulation. *Geoderma* 163 (3), 150–162. <https://doi.org/10.1016/j.geoderma.2011.03.011>.
- Mingwei, Z., Qingbo, Z., Zhongxin, C., Jia, L., Yong, Z., Chongfa, C., 2008. Crop discrimination in Northern China with double cropping systems using Fourier analysis of time-series MODIS data. *Int. J. Appl. Earth Obs. Geoinf.* 10 (4), 476–485. <https://doi.org/10.1016/j.jag.2007.11.002>.
- Moeskops, B., Buchan, D., Van Beneden, S., Fievez, V., Sleutel, S., Gasper, M.S., D'Hose, T., De Neve, S., 2012. The impact of exogenous organic matter on SOM contents and microbial soil quality. *Pedobiologia* 55 (3), 175–184. <https://doi.org/10.1016/j.pedobi.2012.03.001>.
- Moraga, P., Dean, C., Inoue, J., Morawiecki, P., Noureen, S.R., Wang, F., 2021. Bayesian spatial modelling of geostatistical data using INLA and SPDE methods: A case study predicting malaria risk in Mozambique. *Spatial Spatio-temporal Epidemiol.* 39, 100440 <https://doi.org/10.1016/j.sste.2021.100440>.
- Nocita, M., Stevens, A., Noon, C., van Wesemael, B., 2013. Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy. *Geoderma* 199, 37–42. <https://doi.org/10.1016/j.geoderma.2012.07.020>.
- Odhiambo, B.O., Kenduyiwo, B.K., Were, K., 2020. Spatial prediction and mapping of soil pH across a tropical afro-montane landscape. *Appl. Geogr.* 114, 102129 <https://doi.org/10.1016/j.apgeog.2019.102129>.
- 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 Sens. Appl.: Soc. Environ.* 22, 100498.
- Pahlavan-Rad, M.R., Akbarimoghaddam, A., 2018. Spatial variability of soil texture fractions and pH in a flood plain (case study from eastern Iran). *CATENA* 160, 275–281. <https://doi.org/10.1016/j.catena.2017.10.002>.
- Poggio, L., Gimona, A., Spezia, L., Brewer, M.J., 2016. Bayesian spatial modelling of soil properties and their uncertainty: The example of soil organic matter in Scotland using R-INLA. *Geoderma* 277, 69–82. <https://doi.org/10.1016/j.geoderma.2016.04.026>.
- Purushothaman, N.K., Reddy, N.N., Das, B.S., 2022. National-scale maps for soil aggregate size distribution parameters using pedotransfer functions and digital soil mapping data products. *Geoderma* 424, 116006. <https://doi.org/10.1016/j.geoderma.2022.116006>.
- Pushkareva, E., Eckhardt, K.-U., Hotter, V., Frossard, A., Leinweber, P., Frey, B., Karsten, U., 2020. Chemical composition of soil organic matter and potential enzyme activity in the topsoil along a moisture gradient in the High Arctic (Svalbard). *Geoderma* 368, 114304. <https://doi.org/10.1016/j.geoderma.2020.114304>.
- Qin, C., An, Y., Liang, P., Zhu, A., Yang, L., 2021. Soil property mapping by combining spatial distance information into the Soil Land Inference Model (SoLIM). *Pedosphere* 31 (4), 638–644. [https://doi.org/10.1016/S1002-0160\(20\)60016-9](https://doi.org/10.1016/S1002-0160(20)60016-9).
- Qin, W., Wang, Y., Yuan, X., Zhang, Q., Wang, X., Zhao, H., Zhu, B., 2022. Responses of soil carbon dynamics to precipitation and land use in an Inner Mongolian grassland. *Plant Soil*. <https://doi.org/10.1007/s11104-022-05858-8>.
- Rao, P., Wang, Y., Wang, F., Liu, Y., Wang, X., Wang, Z., 2022. Daily soil moisture mapping at 1 km resolution based on SMAP data for desertification areas in northern China. *Earth Syst. Sci. Data* 14 (7), 3053–3073. <https://doi.org/10.5194/essd-14-3053-2022>.
- Richardson, A.D., Andy Black, T., Ciaia, P., Delbart, N., Friedl, M.A., Gobron, N., Hollinger, D.Y., Kutsch, W.L., Longdoz, B., Luysaert, S., Migliavacca, M., Montagnani, L., William Munger, J., Moors, E., Piao, S., Rebmann, C., Reichstein, M., Saigusa, N., Tomelleri, E., Vargas, R., Varlagin, A., 2010. Influence of spring and autumn phenological transitions on forest ecosystem productivity. *Philos. Trans. R. Soc. B: Biol. Sci.* 365 (1555), 3227–3246. <https://doi.org/10.1098/rstb.2010.0102>.
- Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc.: Series B (Stat. Method.)* 71 (2), 319–392.
- Ryan, E.G., Drovandi, C.C., McGree, J.M., Pettitt, A.N., 2015. A review of modern computational algorithms for Bayesian optimal design. *Int. Stat. Rev.* 84 (1), 128–154.
- Sadeghi, M., Babaeian, E., Tuller, M., Jones, S.B., 2017. The optical trapezoid model: A novel approach to remote sensing of soil moisture applied to Sentinel-2 and Landsat-8 observations. *Remote Sens. Environ.* 198, 52–68. <https://doi.org/10.1016/j.rse.2017.05.041>.
- Sandholt, I., Rasmussen, K., Andersen, J., 2002. A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. *Remote Sens. Environ.* 79 (2), 213–224. [https://doi.org/10.1016/S0034-4257\(01\)00274-7](https://doi.org/10.1016/S0034-4257(01)00274-7).
- Shabtai, I.A., Das, S., Inagaki, T.M., Azimzadeh, B., Richards, B., Martínez, C.E., Kögel-Knabner, I., Lehmann, J., 2022. Soil organic carbon accrual due to more efficient microbial utilization of plant inputs at greater long-term soil moisture. *Geochim. Cosmochim. Acta* 327, 170–185. <https://doi.org/10.1016/j.gca.2022.04.028>.
- Sparks, D.L., Page, A., Helmke, P.A., Loeppert, R.H., Nelson, D., Sommers, L., 1996. Total carbon, organic carbon, and organic matter. <https://doi.org/10.2136/sssabookser5.3.c34>.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2001. Bayesian measures of model complexity and fit. *J. R. Stat. Soc.: Series B (Stat. Method.)* 64 (4), 583–639.
- Steinweg, J.M., Dukes, J.S., Wallenstein, M.D., 2012. Modeling the effects of temperature and moisture on soil enzyme activity: Linking laboratory assays to continuous field data. *Soil Biol. Biochem.* 55, 85–92. <https://doi.org/10.1016/j.soilbio.2012.06.015>.
- Stumpf, F., Schmidt, K., Goebes, P., Behrens, T., Schönbrodt-Stitt, S., Wadoux, A., Xiang, W., Scholten, T., 2017. Uncertainty-guided sampling to improve digital soil maps. *CATENA* 153, 30–38. <https://doi.org/10.1016/j.catena.2017.01.033>.
- Sun, X.-L., Wu, S.-C., Wang, H.-L., Zhao, Y.-G., Zhang, G.-L., Man, Y.B., Wong, M.H., 2013. Dealing with spatial outliers and mapping uncertainty for evaluating the

- effects of urbanization on soil: A case study of soil pH and particle fractions in Hong Kong. *Geoderma* 195–196, 220–233. <https://doi.org/10.1016/j.geoderma.2012.11.017>.
- Sun, X.-L., Minasny, B., Wang, H.-L., Zhao, Y.-G., Zhang, G.-L., Wu, Y.-J., 2021. Spatiotemporal modelling of soil organic matter changes in Jiangsu, China between 1980 and 2006 using INLA-SPDE. *Geoderma* 384, 114808. <https://doi.org/10.1016/j.geoderma.2020.114808>.
- Taghizadeh-Mehrjardi, R., Mahdianpari, M., Mohammadimanesh, F., Behrens, T., Toomanian, N., Scholten, T., Schmidt, K., 2020. Multi-task convolutional neural networks outperformed random forest for mapping soil particle size fractions in central Iran. *Geoderma* 376, 114552. <https://doi.org/10.1016/j.geoderma.2020.114552>.
- Takoutsing, B., Heuvelink, G.B.M., 2022. Comparing the prediction performance, uncertainty quantification and extrapolation potential of regression kriging and random forest while accounting for soil measurement errors. *Geoderma* 428, 116192. <https://doi.org/10.1016/j.geoderma.2022.116192>.
- Taneja, P., Vasava, H.K., Daggupati, P., Biswas, A., 2021. Multi-algorithm comparison to predict soil organic matter and soil moisture content from cell phone images. *Geoderma* 385, 114863. <https://doi.org/10.1016/j.geoderma.2020.114863>.
- Tian, Z., Liu, F., Liang, Y., Zhu, X., 2022. Mapping soil erodibility in southeast China at 250 m resolution: Using environmental variables and random forest regression with limited samples. *Int. Soil Water Conserv. Res.* 10 (1), 62–74. <https://doi.org/10.1016/j.iswcr.2021.06.005>.
- Tian, J., Philpot, W.D., 2015. Relationship between surface soil water content, evaporation rate, and water absorption band depths in SWIR reflectance spectra. *Remote Sens. Environ.* 169, 280–289. <https://doi.org/10.1016/j.rse.2015.08.007>.
- Tian, J., Yue, J., Philpot, W.D., Dong, X., Tian, Q., 2021. Soil moisture content estimate with drying process segmentation using shortwave infrared bands. *Remote Sens. Environ.* 263, 112552. <https://doi.org/10.1016/j.rse.2021.112552>.
- Twomey, S.A., Bohren, C.F., Mergenthaler, J.L., 1986. Reflectance and albedo differences between wet and dry surfaces. *Appl. Opt.* 25 (3), 431–437. <https://doi.org/10.1364/AO.25.000431>.
- Wang, X., Yue, Y., Faraway, J., 2018. Bayesian regression modeling with Inla. <https://doi.org/10.1201/9781351165761>.
- Wadoux, A.M.J.C., 2019. Using deep learning for multivariate mapping of soil with quantified uncertainty. *Geoderma* 351, 59–70. <https://doi.org/10.1016/j.geoderma.2019.05.012>.
- Wang, X., Li, L., Liu, H., Song, K., Wang, L., Meng, X., 2022. Prediction of soil organic matter using VNIR spectral parameters extracted from shape characteristics. *Soil Tillage Res.* 216, 105241. <https://doi.org/10.1016/j.still.2021.105241>.
- Wang, J., Zuo, R., 2021. Spatial modelling of hydrothermal mineralization-related geochemical patterns using INLA+SPDE and local singularity analysis. *Comput. Geosci.* 154, 104822. <https://doi.org/10.1016/j.cageo.2021.104822>.
- Whittle, P., 1954. On stationary processes in the plane. *Biometrika* 41 (3–4), 434–449.
- Wigneron, J.P., Jackson, T.J., O'Neill, P., De Lannoy, G., de Rosnay, P., Walker, J.P., Ferrazzoli, P., Mironov, V., Bircher, S., Grant, J.P., Kurum, M., Schwank, M., Munoz-Sabater, J., Das, N., Royer, A., Al-Yaari, A., Al Bitar, A., Fernandez-Moran, R., Lawrence, H., Mialon, A., Parrens, M., Richaume, P., Delwart, S., Kerr, Y., 2017. Modelling the passive microwave signature from land surfaces: A review of recent results and application to the L-band SMOS & SMAP soil moisture retrieval algorithms. *Remote Sens. Environ.* 192, 238–262. <https://doi.org/10.1016/j.rse.2017.01.024>.
- Wright, N., Newell, K., Lam, K.B.H., Kurmi, O., Chen, Z., Kartsonaki, C., 2021. Estimating ambient air pollutant levels in Suzhou through the SPDE approach with R-INLA. *Int. J. Hyg. Environ. Health* 235, 113766. <https://doi.org/10.1016/j.ijheh.2021.113766>.
- Wu, W., 2021. Multiscale Bayesian Spatial Modelling and Mapping of Soil Organic Carbon with the INLA-SPDE Approach (in Chinese). D Thesis. Zhejiang University.
- Wu, C., Wu, J., Luo, Y., Zhang, L., DeGloria, S.D., 2009. Spatial prediction of soil organic matter content using cokriging with remotely sensed data. *Soil Sci. Soc. Am. J.* 73 (4), 1202–1208. <https://doi.org/10.2136/sssaj2008.0045>.
- Xia, C., Zhang, Y., 2022. Comparison of the use of Landsat 8, Sentinel-2, and Gaofen-2 images for mapping soil pH in Dehui, northeastern China. *Ecol. Inf.* 70, 101705. <https://doi.org/10.1016/j.ecoinf.2022.101705>.
- Yang, L., Zhu, A.X., Qi, F., Qin, C.Z., Li, B., Pei, T., 2013. An integrative hierarchical stepwise sampling strategy for spatial sampling and its application in digital soil mapping. *Int. J. Geog. Inf. Sci.* 27 (1), 1–23. <https://doi.org/10.1080/13658816.2012.658053>.
- Yang, L., Song, M., Zhu, A.X., Qin, C., Zhou, C., Qi, F., Li, X., Chen, Z., Gao, B., 2019. Predicting soil organic carbon content in croplands using crop rotation and Fourier transform decomposed variables. *Geoderma* 340, 289–302. <https://doi.org/10.1016/j.geoderma.2019.01.015>.
- 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 Tillage Res.* 196, 104465. <https://doi.org/10.1016/j.still.2019.104465>.
- Yang, L., Li, X., Yang, Q., Zhang, L., Zhang, S., Wu, S., Zhou, C., 2021. Extracting knowledge from legacy maps to delineate eco-geographical regions. *Int. J. Geog. Inf. Sci.* 35 (2), 250–272. <https://doi.org/10.1080/13658816.2020.1806284>.
- 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. <https://doi.org/10.1016/j.geoderma.2021.115407>.
- Yoshida, H., Sazawa, K., Wada, N., Hata, N., Marumo, K., Fukushima, M., Kuramitz, H., 2018. Changes in the chemical composition of soil organic matter including water-soluble component during incubation: A case study of coniferous and broadleaf forest soils. *CATENA* 171, 22–28. <https://doi.org/10.1016/j.catena.2018.06.032>.
- Yuan, Y., Bachl, F.E., Lindgren, F., Brochers, D.L., Illian, J.B., Buckland, S.T., Rue, H., Gerrodette, T., 2016. Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. *arXiv e-prints*, pp. arXiv:1604.06013.
- Yue, J., Tian, J., Tian, Q., Xu, K., Xu, N., 2019. Development of soil moisture indices from differences in water absorption between shortwave-infrared bands. *ISPRS J. Photogramm. Remote Sens.* 154, 216–230. <https://doi.org/10.1016/j.isprsjprs.2019.06.012>.
- Zare, S., Abtahi, A., Fallah Shamsi, S.R., Lagacherie, P., 2021. Combining laboratory measurements and proximal soil sensing data in digital soil mapping approaches. *CATENA* 207, 105702. <https://doi.org/10.1016/j.catena.2021.105702>.
- Zhang, L., Cai, Y., Huang, H., Li, A., Yang, L., Zhou, C., 2022. A CNN-LSTM Model for Soil Organic Carbon Content Prediction with Long Time Series of MODIS-Based Phenological Variables. *Remote Sensing* 14, 4441. <https://doi.org/10.3390/rs14184441>.
- Zhang, K., Chao, L.-J., Wang, Q.-Q., Huang, Y.-C., Liu, R.-H., Hong, Y., Tu, Y., Qu, W., Ye, J.-Y., 2019b. Using multi-satellite microwave remote sensing observations for retrieval of daily surface soil moisture across China. *Water Sci. Eng.* 12 (2), 85–97. <https://doi.org/10.1016/j.wse.2019.06.001>.
- Zhang, C., Mishra, D.R., Pennings, S.C., 2019a. Mapping salt marsh soil properties using imaging spectroscopy. *ISPRS J. Photogramm. Remote Sens.* 148, 221–234. <https://doi.org/10.1016/j.isprsjprs.2019.01.006>.
- Zhang, L., Yang, L., Ma, T., Shen, F., Cai, Y., Zhou, C., 2021. A self-training semi-supervised machine learning method for predictive mapping of soil classes with limited sample data. *Geoderma* 384, 114809. <https://doi.org/10.1016/j.geoderma.2020.114809>.
- Zhang, L., Yang, L., Cai, Y., Huang, H., Shi, J., Zhou, C., 2022. A multiple soil properties oriented representative sampling strategy for digital soil mapping. *Geoderma* 406, 115531. <https://doi.org/10.1016/j.geoderma.2021.115531>.
- Zhang, D., Zhou, G., 2016. Estimation of soil moisture from optical and thermal remote sensing: A review. *Sensors* 16, 1308. <https://doi.org/10.3390/s16081308>.
- Zhang, S.-J., Zhu, A.X., Liu, J., Yang, L., Qin, C.-Z., An, Y.-M., 2016. An heuristic uncertainty directed field sampling design for digital soil mapping. *Geoderma* 267, 123–136. <https://doi.org/10.1016/j.geoderma.2015.12.009>.
- Zhou, Y., Hartemink, A.E., Shi, Z., Liang, Z., Lu, Y., 2019. Land use and climate change effects on soil organic carbon in North and Northeast China. *Sci Total Environ* 647, 1230–1238. <https://doi.org/10.1016/j.scitotenv.2018.08.016>.