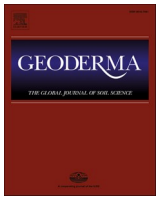
Geoderma 451 (2024) 117067

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma



Space-time modelling of soil organic carbon stock change at multiple scales: Case study from Hungary

Gábor Szatmári [a], László Pásztor [a], Katalin Takács [a,*], János Mészáros [a], András Benő [a,b], Annamária Laborczi [a]

[a] Institute for Soil Sciences, HUN-REN Centre for Agricultural Research, Budapest, Hungary
[b] Doctoral School of Earth Sciences, University of Debrecen, Debrecen, Hungary

ARTICLE INFO

Handling Editor: B. Minasny

Keywords:
Soil organic carbon
Spatiotemporal modelling
Spatiotemporal assessment
Machine learning
Space-time geostatistics
Spatial aggregation
Uncertainty assessment

ABSTRACT

The role of soil organic carbon (SOC) is crucial not only for numerous soil functions and processes but also for addressing various environmental crises and challenges we face. Consequently, the demand for information on the spatiotemporal variability of SOC is increasing, posing new methodological challenges, such as the need for information on SOC and SOC changes with quantified uncertainty across a wide variety of spatial scales and temporal periods. Our objective was to present a methodology based on a combination of machine learning and space–time geostatistics to predict the spatiotemporal variability of SOC stock with quantified uncertainty at various spatial supports (i.e., point support, 1 × 1 km, 5 × 5 km, 10 × 10 km, 25 × 25 km, counties, and the entire country) for Hungary, between 1992 and 2016. The role of geostatistics is pivotal, as it accounts for the spatiotemporal correlation of the interpolation errors, which is essential for reliably quantifying the uncertainty associated with spatially aggregated SOC stock and SOC stock change predictions. Five times repeated 10-fold leave-location-out cross-validation was used to evaluate the point support predictions and uncertainty quantifications, yielding acceptable results for both SOC stock (ME = −0.897, RMSE = 19.358, MEC = 0.321, and G = 0.912) and SOC stock change (ME = 0.414, RMSE = 16.626, MEC = 0.160, and G = 0.952). We compiled a series of maps of SOC stock predictions between 1992 and 2016 for each support, along with the quantified uncertainty, which is unprecedented in Hungary. It was demonstrated that the larger the support, the smaller the prediction uncertainty, which facilitates the identification and delineation of larger areas with statistically significant SOC stock changes. Moreover, the methodology can overcome the limitations of recent approaches in the spatiotemporal modelling of SOC, allowing the prediction of SOC and SOC changes, with quantified uncertainty, for any year, time period, and spatial scale. This methodology is capable of meeting the current and anticipated demands for dynamic information on SOC at both national and international levels.

1. Introduction

Soil organic carbon (SOC) plays a crucial role in the global carbon cycle, as it is the largest terrestrial pool of organic carbon (Lal et al., 2018; Stockmann et al., 2013). SOC is also vital for various soil functions and processes, including agricultural productivity, water retention and management, stabilization of soil aggregates, buffering capacity, and the retention and release of plant nutrients (Centeri et al., 2014; Csikós et al., 2023; Hartemink and McSweeney, 2014; Jakab et al., 2016; Lefévre et al., 2017). However, the significance of SOC extends beyond soil science, attracting interest from multiple disciplines and practical applications such as conservation biology, environmental sciences, meteorology, hydrology, land use planning, and forestry. The multifunctionality of soils and the pivotal role of SOC in addressing environmental crises and challenges are increasingly recognized. The importance of SOC in climate change mitigation (Lal, 2004a; Minasny et al., 2017; Szatmári et al., 2023; Tóth et al., 2018), achieving land degradation neutrality (Keesstra et al., 2018; Stavi and Lal, 2015), maintaining ecosystem services (Jakab et al., 2024; Keesstra et al., 2016; Lal et al., 2018), and ensuring food and water security (Lal, 2004b; Minasny et al., 2017; 2013) is now well-known.

The demand for information on the spatiotemporal variability of SOC is increasing. This interest spans from global and continental scales to national, regional and local levels, highlighting the necessity for

* Corresponding author.
E-mail address: takacs.katalin@atk.hun-ren.hu (K. Takács).

https://doi.org/10.1016/j.geoderma.2024.117067
Received 28 June 2024; Received in revised form 10 October 2024; Accepted 10 October 2024

comprehensive and accurate mapping of SOC over time. Over the past years, numerous mapping activities at different scales have been conducted worldwide, focusing on predicting the spatiotemporal variability of SOC. Examples include Argentina (Heuvelink et al., 2020), Brazil (MapBiomas, https://brasil.mapbiomas.org/en/; Tayebi et al., 2021), China (Chen et al., 2023; Sun et al., 2021; Yang et al., 2023; Zhang et al., 2023a; b), France (Urbina-Salazar et al., 2023), Italy (Schillaci et al., 2017), Mexico and the conterminous United States (Guevara et al., 2020), and the Netherlands (Helfenstein et al., 2024). Additionally, studies have been conducted at continental (De Rosa et al., 2024; Tian et al., 2024) and global scale (Padarian et al., 2022).

In recent years, extensive work has also been conducted in Hungary to better understand the spatiotemporal variability of SOC on a nationwide scale. Various digital soil mapping techniques were employed and compared to identify the technique that provides the most accurate prediction and the smallest yet reliable uncertainty quantification for the country (Szatmári and Pásztor, 2019). Building on this research, SOC stock change was assessed between 1992 and 2010 (Szatmári et al., 2019). This work was later extended to predict SOC stock change for the same time period, but at various spatial scales, using a combination of multivariate geostatistics and machine learning (Szatmári et al., 2021). Note that in our previous research, SOC data were available for only two years, and therefore, it was not possible to develop a space–time model and predict SOC stock for the intermediate years. Since then, SOC data for several years have become available, paving the way for modelling and predicting SOC annually. This still represents an important knowledge gap in Hungary, requiring further research and methodological development.

Spatiotemporal information on SOC is essential for informed decisions in environmental policy, sustainable land management, and climate action. However, there are a number of end-user demands that need to be addressed urgently, including the expanding demands for dynamic information on SOC and the wide variety of spatial scales and temporal periods for which information on SOC change is required. For the latter one, many examples can be given, for instance, carbon accounting requires the spatial average of SOC for the country, reported annually. SOC stock also serves as an indicator for sustainable development goal 15, target 3 (i.e., combat desertification, restore degraded land and soil). The proposed EU Soil Monitoring Directive will require spatial average or total of SOC for soil governance units to assess soil health (Wadoux et al., 2024). These needs often call for the spatial aggregation of SOC or SOC change predictions for larger areas (or supports), e.g., soil districts, total woodlands or grasslands, or the entire country. Crop simulations, terrestrial ecosystem process models or climate models require detailed SOC inputs (Fodor et al., 2014; Hidy et al., 2016; Koós et al., 2021), and SOC conservation strategies are evaluated at even more detailed scales (de Gruijter et al., 2016; Malone et al., 2018).

In addition, we face several scientific challenges in the spatiotemporal modelling of SOC. For example, the uncertainty of SOC at the level of observations may not be negligible due to sampling issues, and errors in measuring SOC and bulk density (Knotters et al., 2022; Paul et al., 2023). Moreover, spatiotemporal interpolation introduces additional uncertainty, although this issue has been properly addressed in recent studies (e.g., Heuvelink et al., 2020; Szatmári et al., 2021; Zhang et al., 2023a). Assessing SOC change after interpolation should account for the uncertainty in the predictions; however, quantifying this uncertainty still presents methodological difficulties (Helfenstein et al., 2024). Furthermore, if SOC change should be assessed at a support larger than the support of the observations, it can be difficult to quantify the uncertainty associated with these spatially aggregated predictions. Although addressing all these challenges is beyond the scope of our study, we focus on some issues and propose solutions to certain methodological challenges.

Clearly, the limited number of SOC observations over time can constrain the spatiotemporal modeling of SOC. However, if sufficient SOC data are available across space and time, either from a soil monitoring system or from multiple soil surveys, then a space–time model can be developed and used to predict SOC annually. Nowadays, the most commonly adopted approach involves the use of machine learning techniques applied to spatially and temporally exhaustive environmental covariates, where static and dynamic covariates over time are distinguished to capture spatial and spatiotemporal variation of the soil forming factors, respectively (Helfenstein et al., 2024; Heuvelink et al., 2020; Tayebi et al., 2021; Yang et al., 2023; Zhang et al., 2023a). Although the quantification of uncertainty associated with point support SOC predictions is increasingly common, a major limitation of this approach emerges.

Quantifying the uncertainty associated with SOC change predictions at point support, and quantifying the uncertainty associated with SOC and SOC change predictions at larger spatial supports pose significant challenges. In both cases, the difficulty lies in accounting for the spatiotemporal correlation of interpolation errors. Incorporating this correlation is essential for providing reliable uncertainty estimates for SOC change predictions at point support and for spatially aggregated SOC and SOC change predictions (Szatmári et al., 2021; Wadoux and Heuvelink, 2023). To the best of our knowledge, machine learning techniques cannot adequately address this correlation, which means that recent studies relying exclusively on machine learning algorithms are unable to reliably quantify the uncertainty associated with SOC change predictions at point support and with spatially aggregated SOC and SOC change predictions. In our previous research (Szatmári et al., 2021), we applied a linear model of coregionalization (Goovaerts, 1997; Wackernagel, 2003), a common technique in multivariate geostatistics, to account for the correlation of errors and to predict SOC and SOC change at various spatial supports. However, this model can only be developed for years for which SOC data have been collected. As a result, this approach is sub-optimal when the aim is to predict or aggregate SOC or SOC change for years or pairs of years lacking SOC data.

Space-time geostatistics offers a straightforward way to account for the spatiotemporal correlation of the errors (Cressie, 1993; Gräler et al., 2016; Webster and Oliver, 2007) and allows for the quantification of uncertainty associated with point support and spatially aggregated predictions. It also overcomes the limitation of multivariate geostatistics, enabling predictions for years without observations. Heuvelink et al. (2020) also suggested using space–time block kriging to predict SOC over larger areas. Indeed, block kriging is specifically designed to predict the spatial average over supports larger than the support of the observations (Goovaerts, 1997; Webster and Oliver, 2007). In block kriging, uncertainty is often quantified by the computed block kriging variance (Hatvani et al., 2021; Lark and Lapworth, 2012), assuming normality and homoscedasticity of errors. However, SOC data are commonly found to be positively skewed (Orton et al., 2014; Szatmári and Pásztor, 2019), and therefore, prior transformation of the data is needed and geostatistical modelling and prediction is conducted on the transformed, normal scale. However, back-transformation of the results at larger supports could pose difficulties (Cressie, 2006; Orton et al., 2015; Szatmári et al., 2021), and consequently, geostatistical simulation is often employed.

The objective of this research was twofold. First, we aimed to present a methodology that combines machine learning and space–time geostatistics to address the growing demand for dynamic information on SOC across various spatial scales and time periods, while also tackling the related methodological challenges. A key goal was not only to predict SOC stock and SOC stock change across different spatial and temporal scales, but also to quantify the uncertainty associated with these spatially aggregated predictions. To the best of our knowledge, this addresses a significant knowledge gap in the spatiotemporal modelling of SOC. The methodology was demonstrated using Hungary as a case study, covering the period from 1992 to 2016 at different aggregation levels. This leads to the second aim of our study, which was to provide a comprehensive series of maps presenting the spatiotemporal variability

of SOC in Hungary; maps that have not been previously available for the country.

2. Materials and methods

2.1. Soil data and computation of SOC stock

Soil data were derived from the Hungarian Soil Information and Monitoring System (SIMS). SIMS, established in 1992, is a nationwide soil monitoring system aimed at providing geographically referenced physical, chemical, and biological information on the temporal changes of Hungarian soils (total area: 93,030 km^2). It stores measurements from 4859 soil horizons across 1236 soil profiles (sampling density: 0.013 soil profiles per km^2) (Fig. 1), which are also referred to as monitoring sites. These sites serve as benchmark locations representing the spatial diversity of Hungarian soils. SIMS distinguishes three types of monitoring sites: (1) I-points (number of soil profiles: 865) placed on agricultural lands (e.g., arable lands, pastures, orchards, vineyards), (2) E-points (number of soil profiles: 183) placed in forests and woodlands, and (3) S-points (number of soil profiles: 188) located in so-called "hot spot" regions, mostly representing different types of environmental hazards (e.g., mining sites, spoil-banks, landfills, industrial areas). It should be noted that S-points were excluded from this research, as they represent areas extremely influenced by anthropogenic activity and are not relevant for soils (Szatmári et al., 2019). Thus, a total of 1048 monitoring sites formed the basis of this research.

The SIMS operating protocol involves revisiting each soil profile annually and sampling each soil horizon. Laboratory analyses are then conducted on the collected samples. However, not all soil properties are measured every year. Specifically, SOC is measured every three years by wet oxidation of organic carbon using potassium dichromate ($K_2Cr_2O_7$) catalyzed by sulfuric acid (H_2SO_4), according to the Hungarian Standard (MSZ–08–0452:1980, 1980). Note that data on SOC were available between 1992 and 2016; data collected after 2016 are not yet available. The proportion of coarse fragments and bulk density of the soil horizons, which are also needed to compute SOC stock, are among the soil properties that were measured only in the first, so-called base year (i.e., 1992).

We computed SOC stock for each soil horizon at each soil profile for each year in which SOC content was measured using the following equation:

$$SOC_{\text{stock}} = 100 \cdot SOC \cdot BD \cdot (1 - CF) \cdot (Top - Bottom) \quad (1)$$

where SOC_{stock} is the soil organic carbon stock (tons $\cdot$ ha^{-1}), SOC is the soil organic carbon content (wt. %), BD is the bulk density (g $\cdot$ cm^{-3}), CF is the proportion of coarse fragments (vol. %), $(Top - Bottom)$ is the thickness of the given soil horizon (cm), and 100 is used for unit conversion. After computing SOC stock values for each soil horizon, we harmonized them for the topsoil (0–30 cm) at each soil profile. Harmonization was carried out using mass-preserving spline (Bishop et al., 1999), which is one of the most frequently used techniques for modelling the vertical distribution of soil properties. Note that computed SOC stock values were assumed to be certain (i.e., error-free).

2.2. Environmental covariates

A number of environmental covariates were collected and derived based on the *scorpan* model (McBratney et al., 2003). Since the objective is to model the space–time variability of SOC stock, it is necessary to

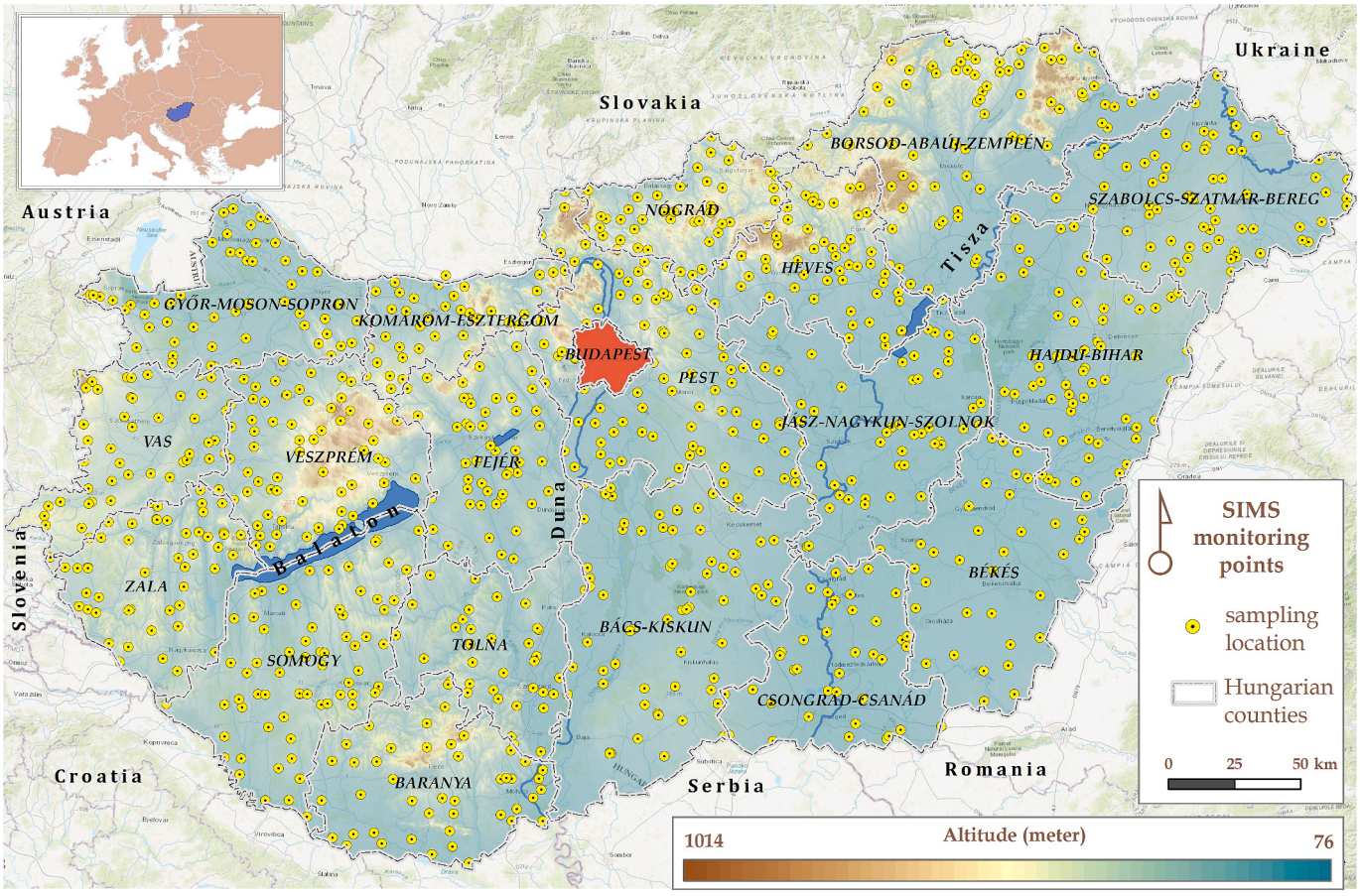


Fig. 1. The location of the soil profiles in the Hungarian Soil Information and Monitoring System (SIMS) presented on a digital elevation model along with the administrative borders of Hungary.

distinguish whether environmental covariates can be considered static or dynamic for the studied time period. For dynamic covariates, data had to be collected and derived for each year between 1992 and 2016. Table 1 summarizes the environmental covariates used in this study.

It was assumed that environmental covariates related to topography and parent material did not change over the studied period, as significant changes in these soil forming factors require more time. Therefore, it is reasonable to consider them static in this research. We used a digital elevation model of Hungary (European Environment Agency, 2016) and many of its derivatives (Table 1). For parent material, we applied the geological map of Hungary (Gyalog and Síkhegyi, 2005), which had been previously correlated with the FAO code system on parent material classes (Bakacsi et al., 2014). Thus, the FAO conform parent material classes were used in this study. In our former research (Szatmári et al., 2021; 2019), it was found that soil type is among the most informative covariates, and thus we also used the soil type map of Hungary (Pásztor et al., 2015). This was also considered static over the studied time period, as significant changes in soil types also need more time.

Table 1
Summary of the environmental covariates. Abbreviations: S: other soil attributes, C: climate, O: organisms, R: relief or topography, P: parent material, DEM: digital elevation model, HungaroMet: Hungarian Meteorological Service, ESA CCI-LC: European Space Agency Climate Change Initiative – Land Cover.

Factor	Name	Type	Source
S	Soil type map of Hungary	static	Pásztor et al. (2015)
C	Seasonal precipitation	dynamic	HungaroMet
	Seasonal maximum temperature	dynamic	HungaroMet
	Seasonal minium temperature	dynamic	HungaroMet
	Seasonal mean temperature	dynamic	HungaroMet
	Seasonal evapotranspiration	dynamic	Cornes et al. (2018)
	Seasonal incoming global radiation	dynamic	Cornes et al. (2018)
O	Land cover	dynamic	ESA CCI-LC
	NIR: Near-infrared band	dynamic	Landsat
	SWIR1: Short wavelength infrared 1 band	dynamic	Landsat
	SWIR2: Short wavelength infrared 2 band	dynamic	Landsat
	NDVI: Normalized difference vegetation index	dynamic	derived from Landsat
	TVI: Transformed vegetation index	dynamic	derived from Landsat
	EVI: Enhanced vegetation index	dynamic	derived from Landsat
	SATVI: Soil-adjusted total vegetation index	dynamic	derived from Landsat
	SAVI: Soil-adjusted vegetation index	dynamic	derived from Landsat
	MSI: Moisture stress index	dynamic	derived from Landsat
	GNDVI: Green normalized difference vegetation index	dynamic	derived from Landsat
	GRVI: Green-red vegetation index	dynamic	derived from Landsat
	LSWI: Land surface water index	dynamic	derived from Landsat
	MSAVI2: Modified soil-adjusted vegetation index (2nd)	dynamic	derived from Landsat
	BI: Brightness index	dynamic	derived from Landsat
	DSWI: Disease water stress index	dynamic	derived from Landsat
R	Elevation (DEM)	static	EU-DEM
	Channel network base level	static	derived from DEM
	Diurnal anisotropic heating	static	derived from DEM
	Horizontal distance to channel network	static	derived from DEM
	LS factor	static	derived from DEM
	Mass balance index	static	derived from DEM
	Multi-resolution ridge top flatness	static	derived from DEM
	Multi-resolution valley bottom flatness	static	derived from DEM
	Multi-scale topographic index	static	derived from DEM
	Profile curvature	static	derived from DEM
	Real surface area	static	derived from DEM
	SAGA wetness index	static	derived from DEM
	Slope	static	derived from DEM
	Stream power index	static	derived from DEM
	Surface roughness	static	derived from DEM
	Total curvature	static	derived from DEM
	Topographic position index	static	derived from DEM
	Topographic ruggedness index	static	derived from DEM
	Topographic wetness index	static	derived from DEM
	Vertical distance to channel network	static	derived from DEM
P	Geological map of Hungary	static	Gyalog and Síkhegyi (2005)

Environmental covariates related to climate were considered as dynamic covariates over the studied period. National and international databases were used, including the meteorological data repository (https://odp.met.hu/) of the Hungarian Meteorological Service (HungaroMet) and E-OBS daily gridded meteorological data for Europe (Cornes et al., 2018). Based on these databases, the climatic parameters listed in Table 1 were derived for each season (spring: March to May; summer: June to August; autumn: September to November; winter: December to February) for each year between 1992 and 2016.

The biosphere (or organisms) is another soil forming factor, which was considered dynamic over the studied time period. We used products from the ESA Climate Change Initiative – Land Cover (Copernicus Climate Change Service - Climate Data Store, 2019), which provides global gridded land cover data with yearly temporal resolution starting from 1992. The dataset is based on the FAO's land cover classification system. Additionally, satellite imagery and several derived spectral indices were used (Table 1). Annual satellite image mosaics, acquired by Landsat-4, −5, −7, −8, and −9 from June to October, were created using cloud masked images and median filter for each year between 1992 and 2016. The formulas used to derive the spectral indices can be found in the Supplementary Material (SM1).

In total, 62 environmental covariates were used in this study. Since these were obtained from different data sources, including national and international databases and repositories, it was necessary to resample them into a common geographic reference system with a resolution of 100 m. This resolution was chosen because it corresponds to the target resolution of the GlobalSoilMap initiative (Arrouays et al., 2014), as well as to the general spatial resolution of the DOSoReMI@hu products (Pásztor et al., 2020). For categorical covariates (e.g., soil type, land cover, parent material), we used a nearest neighbor resampling technique. For continuous covariates (e.g., climatic data layers, digital elevation model and its derivatives), we employed a cubic spline technique.

2.3. Space-time modelling

As mentioned in the Introduction, data on SOC stock frequently found to be positively skewed, and Hungary is no exception (Szatmári et al., 2021; Szatmári and Pásztor, 2019). Therefore, the SOC stock data were transformed using square root transformation and space–time modelling of SOC stock was conducted on the transformed, normal scale. The importance and rationale for this transformation are described in detail in Szatmári and Pásztor (2019).

In this study, the spatiotemporal variation of SOC stock was modelled as a sum of a deterministic component and a stochastic component, that is

$$Z(\mathbf{u},t) = m(\mathbf{u},t) + \varepsilon(\mathbf{u},t) \quad (2)$$

where Z is the transformed SOC stock, m is the deterministic component describing structural variation in space and time, ε is the stochastic part of variation that often show spatiotemporal correlation, $\mathbf{u}$ is the vector of the geographical coordinates, and t refers to time. In the following subsections, it is described how we modelled the deterministic and stochastic components using machine learning and space–time geostatistics, respectively.

2.3.1. Random forest in space and time

We used the random forest (RF) algorithm (Breiman, 2001), which proved to be a well-performing machine learning technique for SOC mapping not just in Hungary (Szatmári et al., 2021; 2019; Szatmári and Pásztor, 2019) but also in other countries (Guevara et al., 2018; Helfenstein et al., 2024; Heuvelink et al., 2020; Veronesi and Schillaci, 2019) and even at continental (Hengl et al., 2015; Wadoux et al., 2023) and global scale (Poggio et al., 2021). In brief, the RF algorithm generates multiple classification or regression trees during training, each tree trained on a random subset of the training data and considering a random subset of predictors (i.e., environmental covariates in our case) at each split. The final prediction for classification is determined by a majority vote of all trees, while for regression, it is the weighted average of the predictions. This ensemble approach helps improve generalization and robustness, making RF a powerful and versatile machine learning algorithm. The number of studies using RF or similar machine learning techniques for space–time modelling is increasing (Erdélyi et al., 2023; Helfenstein et al., 2024; 2022; Hengl et al., 2018; Heuvelink et al., 2020; Zhang et al., 2023a). In such applications, the regression matrix is derived from a space–time overlay of the observations on environmental covariates. This requires spatially and temporally exhaustive covariates and knowledge of the date of sampling. Once the regression matrix obtained, model calibration, cross-validation and prediction are done in the usual way (Hengl et al., 2018; Heuvelink et al., 2020). In this study, we employed the methodology described by Heuvelink et al. (2020).

The hyperparameters of RF can still be fine-tuned when used for space–time modelling. In this study, we fine-tuned the hyperparameter "mtry", which is the number of random subset of environmental covariates at each split. During fine-tuning and RF fitting, 5 times repeated 10-fold cross-validation was used. We should note that "leave-location-out" cross-validation (LLOCV) was used instead of "classical" cross-validation, as the latter often gives over-optimistic picture about the performance of space–time modelling and thus potentially lead to deceiving results (Meyer et al., 2018). In LLOCV, the random subsets (i.e., folds) used in cross-validation are not defined at the level of individual observations, as in "classical" cross-validation, but at the level of the monitoring sites, i.e., folds consist of randomly selected monitoring sites with all their associated observations. For more details and explanation, see Meyer et al. (2018).

After fine-tuning the hyperparameter "mtry", the final RF model was fitted between the transformed SOC stock data and the environmental covariates. Then it was used for giving space–time prediction with a resolution of 100 m for each year between 1992 and 2016, and it was identified as the deterministic part of variation (first term on the right-hand side of Eq. (2)).

2.3.2. Space-time geostatistics

2.3.2.1. Space-time kriging for point support prediction. Geostatistical analysis was performed on the residuals obtained from the fitted RF model. Each residual includes spatial coordinates (i.e., longitude and latitude) as well as temporal information (i.e., year), making it suitable for space–time geostatistical analysis. It was assumed that the residuals represent a stationary spatiotemporal random field (second term on the right-hand side of Eq. (2)), which is multivariate normal with zero mean. In such a case, it can be fully characterized by a covariance function, that is

$$C_{st}(\mathbf{h}, v) = \mathrm{Cov}(\varepsilon(\mathbf{u}, t), \varepsilon(\mathbf{u}+\mathbf{h}, t+v)) \quad (3)$$

where $C_{st}(\mathbf{h}, v)$ is the spatiotemporal covariance function that quantifies the covariance between $\varepsilon(\mathbf{u}, t)$ and $\varepsilon(\mathbf{u}+\mathbf{h}, t+v)$ for any pair of points $(\mathbf{u}, t)$ and $(\mathbf{u}+\mathbf{h}, t+v)$ in the space–time domain, and $C_{st}(\mathbf{h}, v)$ depends only on the separation distance in space (i.e., $\mathbf{h}$) and time (i.e., v). Note that calculating and modelling the spatiotemporal covariance is a real challenge (Gräler et al., 2016; Heuvelink et al., 2017), and therefore we should make additional assumptions. It was supposed that the spatial term of the stochastic component is isotropic, that is the separation vector $\mathbf{h}$ is taken to be scalar, namely, the Euclidean distance (i.e., $h = \|\mathbf{h}\|$). Furthermore, a sum-metric covariance model (Gräler et al., 2016; Snepvangers et al., 2003) was used, which is a combination of a spatial, a temporal and a metric model including an anisotropy parameter:

$$C_{st}(h, v) = C_s(h) + C_t(v) + C_{joint}\left(\sqrt{h^2 + (\kappa \cdot v)^2}\right) \quad (4)$$

where $C_s(h)$ and $C_t(v)$ are the spatial and temporal terms of the spatiotemporal variability, respectively, and the last term on the right-hand side is the joint space–time model with the anisotropy parameter κ. This parameter, which is given as spatial unit per temporal unit, matches space and time by re-scaling temporal distances to equivalent spatial distances (Gräler et al., 2016; Snepvangers et al., 2003). In geostatistics, it is more common to use the variogram instead of the covariance function (Gräler et al., 2016), thus

$$\gamma_{st}(h, v) = \gamma_s(h) + \gamma_t(v) + \gamma_{joint}\left(\sqrt{h^2 + (\kappa \cdot v)^2}\right) \quad (5)$$

where γ_s, γ_t and γ_{joint} are the spatial, temporal and joint variograms, respectively. Once either the covariance or the variogram has been calculated and modelled, kriging can be done in the usual way (Heuvelink and Griffith, 2010; Heuvelink and van Egmond, 2010). In this study, the prediction of the space–time variability of SOC stock at point support, with a resolution of 100 m, was given by adding the space–time kriged RF residuals to the prediction of the RF model (Eq. (2)). The point support SOC stock predictions were then transformed back to the original, positively skewed scale by taking the square of a prediction and adding the computed kriging variance as bias correction. The uncertainty associated with the point support prediction of SOC stock was expressed by the width of the 90 % prediction interval, which is defined by the 0.05 and 0.95 quantile. The quantiles were derived by subtracting and adding 1.64 times the kriging standard deviation to the SOC stock prediction. Finally, they were also transformed back to the original scale by taking their square.

2.3.2.2. 2D + T LU simulation for change of support. As it was indicated in the Introduction, geostatistical simulation was used to predict the spatial average of SOC stock and SOC stock change with quantified uncertainty at various supports. The supports we used in this study are listed in Table 2. We applied LU simulation (Alabert, 1987; Davis, 1987), which has been used in geostatistics for decades to simulate values for larger supports both in 2D and 3D (Deutsch and Journel, 1998; Goovaerts, 1997). In brief, the simulation algorithm is based on the Cholesky factorization (or decomposition) that decomposes the covariance matrix of a given 2D or 3D block into a lower and an upper triangular matrix,

Table 2
List of supports used in this study. In the last column, two numbers separated by the multiplication symbol (×) are presented. The first number means the number of discretization points in the geographical space (2D), whereas the second means the number of time slices (T) from 1992 to 2016 (i.e., 25 years).

Support	Area	Number of 2D + T blocks	Total number of discretization points over a 2D + T block
1 km square blocks	1 × 1 km	94,118	10 × 25
5 km square blocks	5 × 5 km	3938	100 × 25
10 km square blocks	10 × 10 km	1029	200 × 25
25 km square blocks	25 × 25 km	184	1000 × 25
Hungarian counties	525–8443 km^2	19 counties + Budapest*	263–4222 × 25
Hungary	93,030 km^2	1	9000 × 25

* Budapest, as the capital, is officially not part of Pest County

hence the name LU (lower–upper) simulation. First, the given 2D or 3D block is discretized by points often following a grid scheme, then the covariance matrix of the discretization points is subject to Cholesky factorization, and finally the obtained lower triangular matrix is multiplied by a vector of independent random numbers to generate simulated values at the discretizing grid points. By computing the average of the simulated values we obtain a realization of the block average. More simulated values at the discretizing grid points and thus more realizations for the block average can be readily obtained by using other vectors of independent random numbers.

Although the original algorithm is designed to 2D and 3D simulation, it can be readily extended to space–time application if the spatiotemporal covariance matrix is used. In this study, we considered a 2D + T block dimension, spanning spatially over 2D and temporally from 1992 to 2016 with annual breakdowns (i.e., 25 years in total). The discretization points within each 2D + T block followed a grid scheme across space and time with constant geographical locations over time. For instance, if there are 9 discretizing points in the 2D geographical space and 5 time slices, then in each time slice the discretizing points are at the same geographical locations, and the total number of points discretizing the 2D + T block is $9 \times 5 = 45$. Since the block size (or area) varies from support to support, different point densities were used for discretization at each support (Table 2). Point densities in the 2D geographical space ranged from 10 to 0.01 points per km^2. Higher densities were used for smaller supports (e.g., 1 × 1 km square block, 5 × 5 km square block) and lower densities were used for larger supports (e.g., Hungarian counties, entire Hungary). Hungarian counties were discretized with a point density of 0.5 points per km^2.

We used conditional LU simulation (i.e., it was conditioned to the transformed data in space and time), thus the Cholesky decomposition of the spatiotemporal covariance matrix **C** is

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} = \mathbf{L}\cdot\mathbf{U} = \begin{bmatrix} \mathbf{L}_{11} & 0 \\ \mathbf{A}_{21} & \mathbf{L}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{11} & \mathbf{B}_{12} \\ 0 & \mathbf{U}_{22} \end{bmatrix} \tag{6}$$

where $\mathbf{C}_{11}$ is the spatiotemporal covariance matrix between the data points, $\mathbf{C}_{21}$ is the spatiotemporal covariance matrix between the data and the points discretizing the 2D + T block, and $\mathbf{C}_{22}$ is the spatiotemporal covariance matrix between the discretization points. Note that $\mathbf{C}_{12}$ is the transpose of $\mathbf{C}_{21}$. After decomposing the covariance matrix **C** a spatiotemporal stochastic realization at the points discretizing the 2D + T block can be generated by

$$\mathbf{z}^l = \mathbf{A}_{21}\cdot\mathbf{L}_{11}^{-1}\cdot\mathbf{z} + \mathbf{L}_{22}\cdot\boldsymbol{\omega} \tag{7}$$

where $\mathbf{z}^l$ is the vector of a simulated stochastic realization, **z** is the vector of the conditioning data, and ω is the vector of independent random numbers. Note that the first term on the right-hand side is the kriging term and the second is the simulated kriging error. As other geostatistical simulation algorithms, it also reproduces the spatiotemporal correlation structure by the realizations generated. This is indispensable to reliably quantify the uncertainty associated with predictions given at larger supports (e.g., spatial average or total). In this study, 1000 stochastic realizations were generated for each 2D + T block at each support (Table 2).

2.4. Spatial aggregation

To obtain prediction for the spatial average of SOC stock over a given block in a given year we should do the following: (1) transform the simulated SOC stock values for the given block in the given year back to the original, positively skewed scale by taking their square, (2) compute the average of the back-transformed SOC stock values for each realization yielding 1000 average values, which can be interpreted as 1000 realizations of block average of SOC stock, and then (3) compute the average of the 1000 block realizations. The uncertainty associated with this prediction was quantified by the width of the 90 % prediction interval, which was determined by the 0.05 and 0.95 quantiles of the 1000 block realizations.

If we are interested in the spatial average of SOC stock change over a given block between two years, then we should do the following: (1) transform the simulated SOC stock values for the given block in the given two years back to the original, positively skewed scale by taking their square, this yields paired realizations for SOC stock, (2) compute the difference between the paired back-transformed SOC stock realizations, this yields realizations for SOC stock change, (3) compute the average of SOC stock change for each realization yielding 1000 average values, which can be interpreted as 1000 realizations of block average of SOC stock change between the two years, and then (4) compute the average of the 1000 block realizations. Similarly to the case above, the uncertainty associated with the prediction was quantified by the width of the 90 % prediction interval, which was determined by the 0.05 and 0.95 quantiles of the 1000 block realizations of SOC stock change.

To identify significant change in SOC stock between the two years, we examined how many of 1000 block realizations of SOC stock change have a positive or negative change (Szatmári et al., 2021). If the number of positive (or negative) SOC stock changes was larger than 950 (i.e., 95 % of the 1000 block realizations), then we said there was a statistically significant increase (or decrease) in SOC stock over that block between the two years.

2.5. Validation

To assess the performance of the SOC stock and SOC stock change predictions at point support 5 times repeated 10-fold LLOCV was used. The prediction performance was assessed by computing the mean error (ME), root mean square error (RMSE), and model efficiency coefficient (MEC):

$$ME = \frac{1}{n}\sum_{i=1}^{n} P_i - O_i \tag{8}$$

$$RMSE = \sqrt[2]{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{9}$$

$$MEC = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2} \tag{10}$$

where P_i and O_i are the predicted and observed SOC stock (or SOC stock change), respectively, and $\overline{O}$ is the mean of the observed SOC stocks (or SOC stock changes). ME is commonly referred to as bias, and RMSE is the spread of the error distribution. If MEC is between 0 and 1, then it can be similarly interpreted as R-square, and if it takes a negative value, then it means that the prediction is worse than using the mean of the data as prediction.

We also assessed the reliability of the uncertainty quantifications at the level of point support. For this purpose, prediction interval coverage probability (PICP) plot (also known as accuracy plot) was compiled (Shrestha and Solomatine, 2006), and then G-statistic was computed (Deutsch, 1997). A PICP plot graphically presents the expected and observed fraction of validation data falling within a symmetric prediction interval. For instance, for the 90 % prediction interval, we expect 90 % of the validation data to fall within the interval and 10 % outside the interval. This can be generalized to any symmetric prediction interval:

$$PICP(p) = \frac{1}{n}\sum_{i=1}^{n}\delta(l_i \le O_i < u_i) \tag{11}$$

where l_i and u_i are the lower and upper limit of the p-width symmetric prediction interval for P_i, respectively, and $\delta(\bullet)$ is an indicator function taking 1 if the expression in the brackets is true and 0 otherwise. Ideally, the PICP plot follows the 1:1 line. G-statistic measures the overall

closeness of the PICP plot to the 1:1 line, that is

$$G = 1 - \int_0^1 |PICP(p) - p|dp \quad (12)$$

Ideally, the G value is equal to 1.

More recently, Schmidinger and Heuvelink (2023) pointed out a weakness of PICP, namely that it does not take systematic one-sided bias into account. Therefore, they proposed additional validation metric, namely quantile coverage probability (QCP). QCP follows the same logic as PICP, but uses quantiles instead of symmetric prediction intervals, that is:

$$QCP(\tau) = \frac{1}{n}\sum_{i=1}^{n}\delta(O_i \le q_\tau^i) \quad (13)$$

where q_τ^i is the τ quantile for P_i, and $\delta(\bullet)$ is an indicator function taking 1 if the expression in the brackets is true and 0 otherwise. Its advantage is that one-sided bias will not be hidden.

3. Results

3.1. Space-time modelling of SOC stock at point support

In Fig. 2, we present the boxplots for each year and histogram of all SOC stock data before and after square root transformation. The histogram of the original (i.e., untransformed) data confirmed that the distribution of the SOC stock values is positively skewed (Fig. 2, upper left corner). The boxplots of the original data show some outlying observations in the upper tail of the distribution (Fig. 2, lower left corner), which are mostly due to the skewed distribution of the data rather than erroneous observations. Those high SOC stock values were observed in peatlands known to be significant organic carbon pools worldwide (Illés et al., 2019; Minasny et al., 2019). After square root transformation it was found that the yearly and overall distribution of the data are close to normal, furthermore, transformation also handled most of the outliers (Fig. 2, right column). Outliers, which were not handled by transformation, were removed from the dataset. Summary statistics of SOC stock data are shown in Table 3, and the spatial distribution of SOC stock data can be found in the Supplementary Material (SM2).

It was found that the optimal value of the hyperparameter "mtry" is 10 (Fig. 3, left graph), and the associated RMSE and R-square values are 1.367 and 0.421, respectively. The final RF model consists of 500 trees and the minimum node size was set to 5. In Fig. 3 (right graph), we present the 40 most important environmental covariates in the final RF model. Soil type proved to be the most important, which is completely in line with our previous findings (Szatmári et al., 2021; 2019). This can be explained by that the genetic soil types in Hungary (Stefanovits, 1963) strongly depend on the development, amount and quality of soil organic matter, among others. According to the World Reference Base (WRB) international soil classification system (IUSS Working Group WRB, 2022), the soil types with the highest proportion of area are Vertisols (17.1 %), Gleysols (14.5 %), Regosols (13.6 %), Chernozems and Kastanozems (13.2 %), Arenosols (11.5 %), Luvisols and Alisols (10.8 %), and Stagnosols (6.3 %). In addition to soil type, parent material and several geomorphometric parameters (e.g., multi-resolution ridge top flatness, elevation, total curvature) were also found to be informative covariates, which were considered static over the studied time period. Last but not least, covariates related to climate (e.g., seasonal precipitation and temperature) and organisms (e.g., soil-adjusted total

Table 3
Summary statistics of the soil organic carbon (SOC) stock data derived from the Hungarian Soil Information and Monitoring System. Annotation: The unit of the SOC stock data is [tons · ha^{-1}], and n means the number of observations.

Years	n	Min	Max	Mean	Median	Std. deviation
1992	983	2.45	221.44	50.08	46.83	26.06
1995	944	0.50	186.32	50.52	46.43	26.14
1998	995	1.51	207.15	52.36	48.16	26.12
2000	842	1.95	157.35	52.19	49.71	23.40
2004	807	3.61	178.18	52.01	49.13	25.46
2007	999	5.15	171.91	53.99	51.46	24.92
2010	972	7.42	164.01	52.33	50.02	23.96
2013	985	2.14	166.16	51.11	48.02	23.74
2016	939	4.79	162.43	49.81	48.50	23.57
Total	8466	0.50	221.44	51.66	48.43	25.16

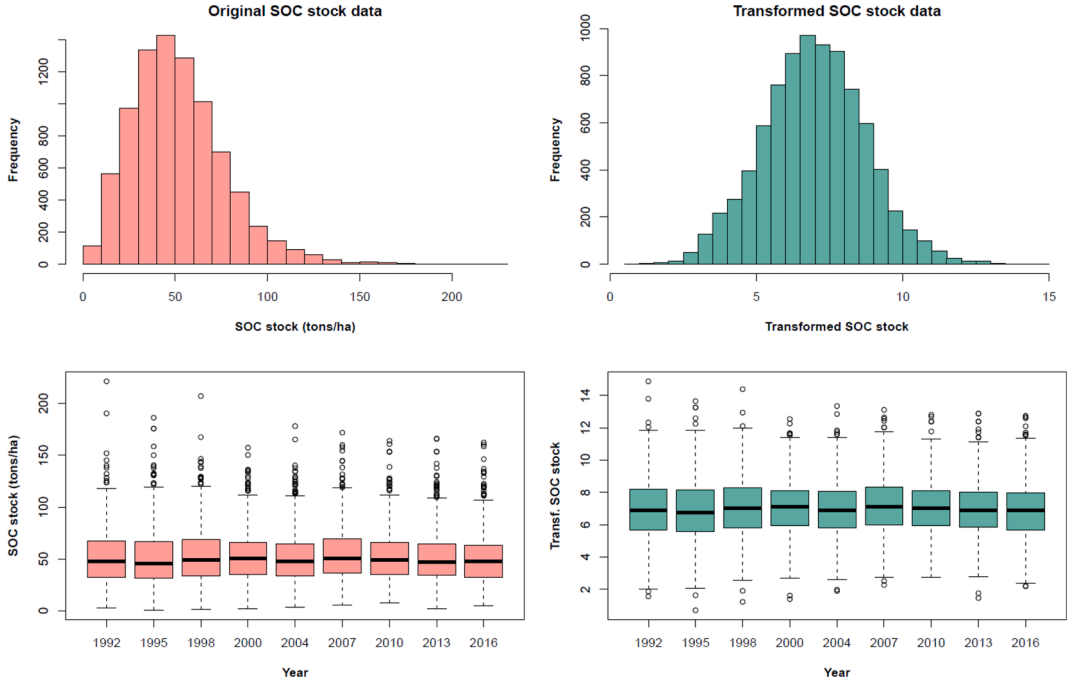


Fig. 2. Graphical summary of the soil organic carbon (SOC) stock data before (left column) and after (right column) square root transformation. First row presents the histogram of the overall SOC stock data, and second row presents the boxplots for each year with measured SOC.

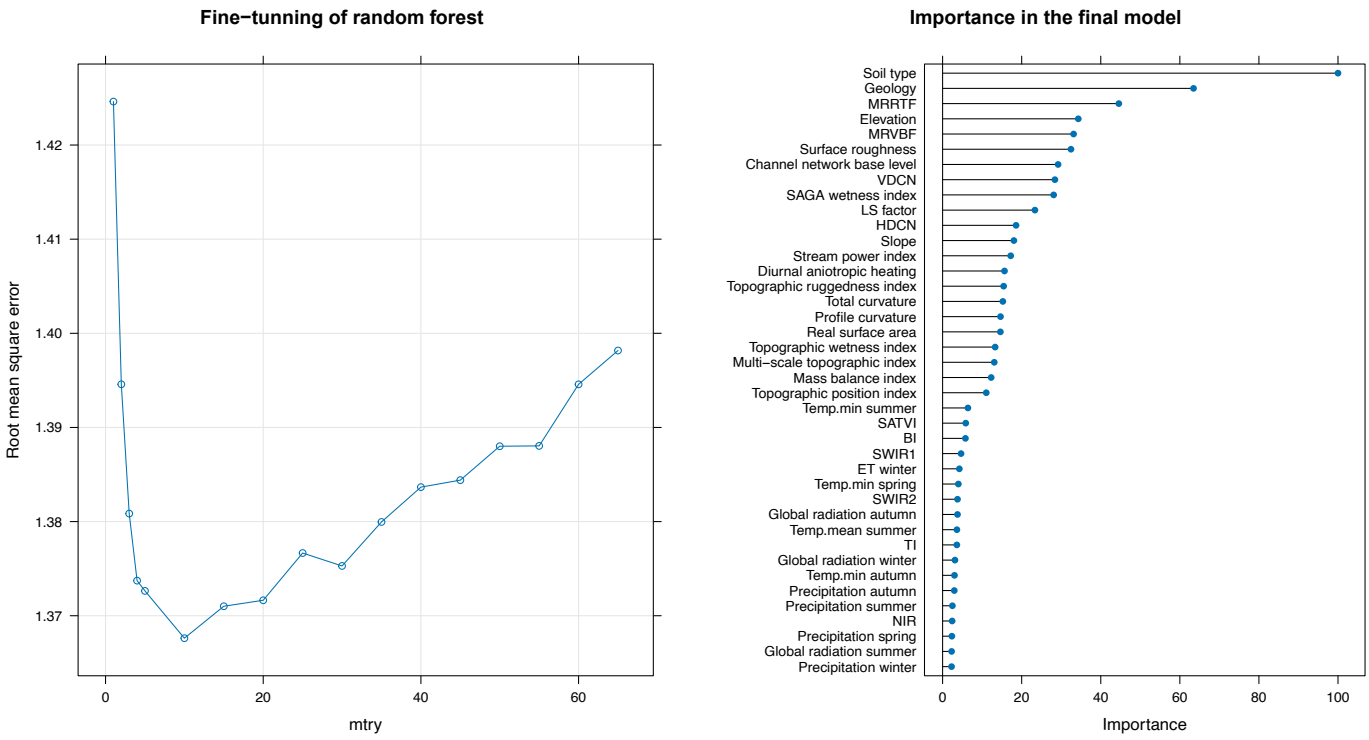


Fig. 3. Fine-tuning of the number of covariates used in each split ("mtry") (left graph), and the 40 most important covariates in the final random forest model (right graph). Abbreviations used in the right graph: MRRTF: Multi-resolution ridge top flatness, MRVBF: Multi-resolution valley bottom flatness, VDCN: Vertical distance to channel network, HDCN: Horizontal distance to channel network, SATVI: Soil-adjusted total vegetation index, SWIR1: Short wavelength infrared (band 1), ET: Evapotranspiration, TI: Transformed vegetation index, BI: Brightness index, SWIR2: Short wavelength infrared (band 2), and Temp: Temperature.

vegetation index, short wavelength infrared, brightness index) are also in the top 40. Note that they were considered dynamic in the respective period.

We derived the RF residuals and then space–time geostatistical analysis was conducted on them. Fig. 4 presents the experimental variogram and the fitted spatiotemporal variogram model in 3D. As it was mentioned, a sum-metric variogram model was fitted, which models the space–time variability as a combination of spatial, temporal and joint variograms (Eq. (5)). Its parameters are summarized in Table 4.

Using the fitted variogram model, space–time kriging was performed on the RF residuals, and then the resulting space–time kriging predictions were added to the predictions obtained by the RF model, according to Eq. (2). Finally, we transformed the point support SOC stock predictions back to the original, positively skewed scale, which were then presented in Fig. 5. The quantified uncertainty associated with the point support prediction of SOC stock can be found in the

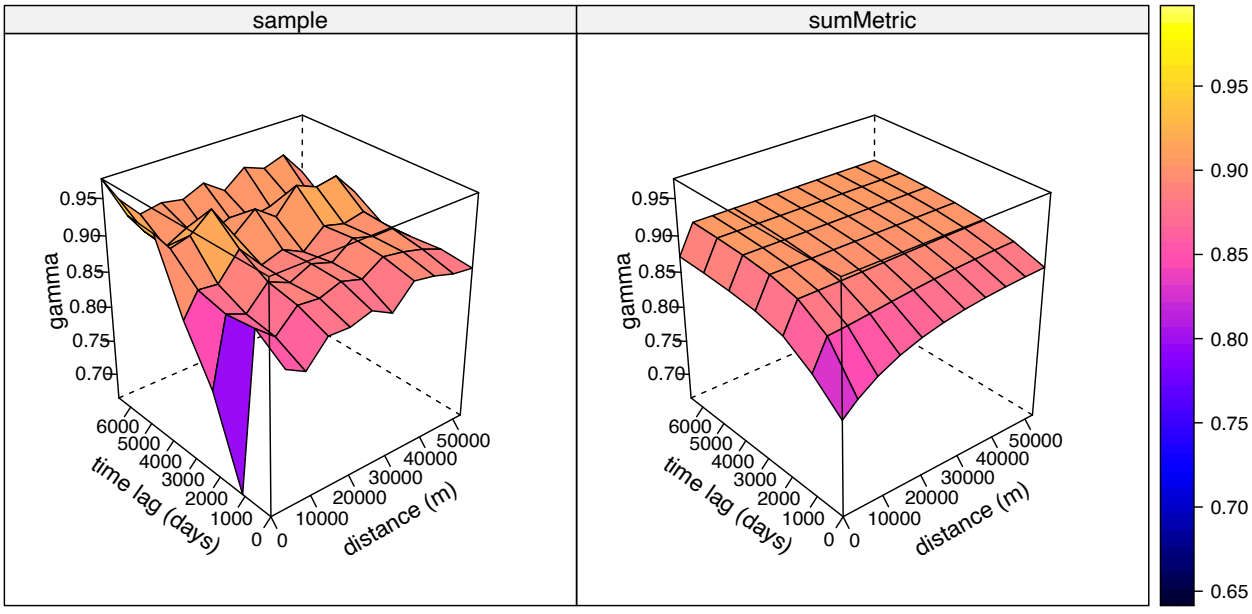


Fig. 4. 3D representation of the spatiotemporal variogram of the random forest residuals (left plot) and the fitted sum-metric variogram model (right plot). Annotation: The parameters of the variogram model are summarized in Table 4.

Table 4
The parameters of the sum-metric variogram model fitted to the random forest residuals. The space–time anisotropy parameter of the model reported in the last row. Annotation: A 3D representation of the fitted variogram model is presented in Fig. 4 (right plot).

Component	Partial sill	Model type	Range	Nugget
Spatial variogram	0.001	Exponential	20,000 m	0.040
Temporal variogram	0.032	Exponential	2000 days*	0.001
Joint variogram	0.083	Exponential	15,000 m	0.758
Anisotropy parameter (κ)	17.764 m · day^{-1}**			

** Approximately 5.5 years, ** Approximately 6.48 km · year^{-1}*

Supplementary Material (SM3).

3.2. Performance of SOC stock predictions and uncertainty quantifications at point support

The performance of the spatiotemporal predictions at point support is presented in Table 5. For SOC stock, it was found that the space–time model slightly underestimates SOC stock by 0.897, whereas for SOC stock change, the space–time model slightly overestimates the actual change by 0.414. However, if we take the standard deviation associated with the computed ME values into account, then we can say that SOC stock and SOC stock change predictions are practically unbiased at point support. In both cases, the magnitude of the computed RMSE and MEC values are similar to those presented in our previous study (Szatmári et al., 2021). However, a deeper comparison cannot be made because in our previous research the computed error measures referred to a non-space–time model, mapping the SOC stock change for a certain time period (i.e., 1992–2010), whereas in this study, these measures refer to a space–time model that is able to predict SOC stock change for any time period between 1992 and 2016. The PICP plots presented in Fig. 6 (left column) has shown that for SOC stock, the uncertainty quantifications at point support slightly underestimate the "real" uncertainty, whereas for SOC stock change, it was found that the uncertainty quantifications slightly overestimate the uncertainty. The computed G-statistics, which measures the overall closeness of PICP to the 1:1 line, were found to be 0.912 and 0.952 for SOC stock and SOC stock change, respectively. They are still quite close to the ideal, expected value (i.e., 1), thus the predicted uncertainty can be considered as a realistic quantification of uncertainty. This was also confirmed by the QCP plots (Fig. 6, right column), furthermore, the QCP plots also revealed that uncertainty

Table 5
The performance of the SOC stock and SOC stock change spatiotemporal predictions at point support using 5 times repeated 10-fold leave-location-out cross-validation. Annotation: In the brackets the corresponding standard deviation is presented. Abbreviations: ME: mean error (or bias), RMSE: root mean square error, and MEC: model efficiency coefficient.

	ME [tons · ha^{-1}]	RMSE [tons · ha^{-1}]	MEC [-]
SOC stock	−0.897 (1.984)	19.358 (2.042)	0.321 (0.087)
SOC stock change	0.414 (1.742)	16.626 (2.112)	0.160 (0.067)

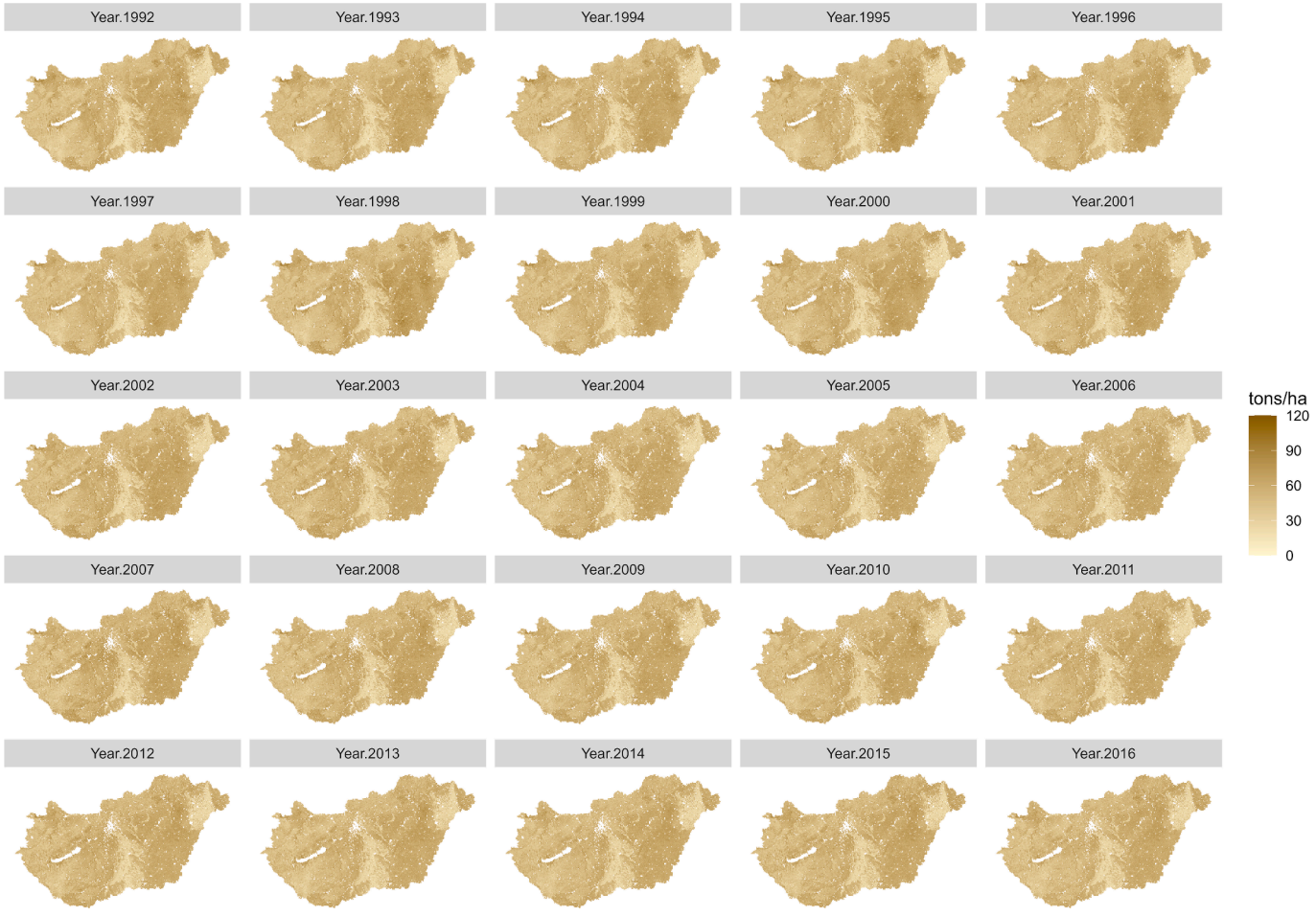


Fig. 5. Spatiotemporal prediction of soil organic carbon (SOC) stock at point support from 1992 to 2016. The spatial resolution is 100 m, the unit is [tons · ha^{-1}], and the maps have been masked with open water bodies and settlements. Annotation: The uncertainty associated with the spatiotemporal predictions can be found in the supplementary material (SM3).

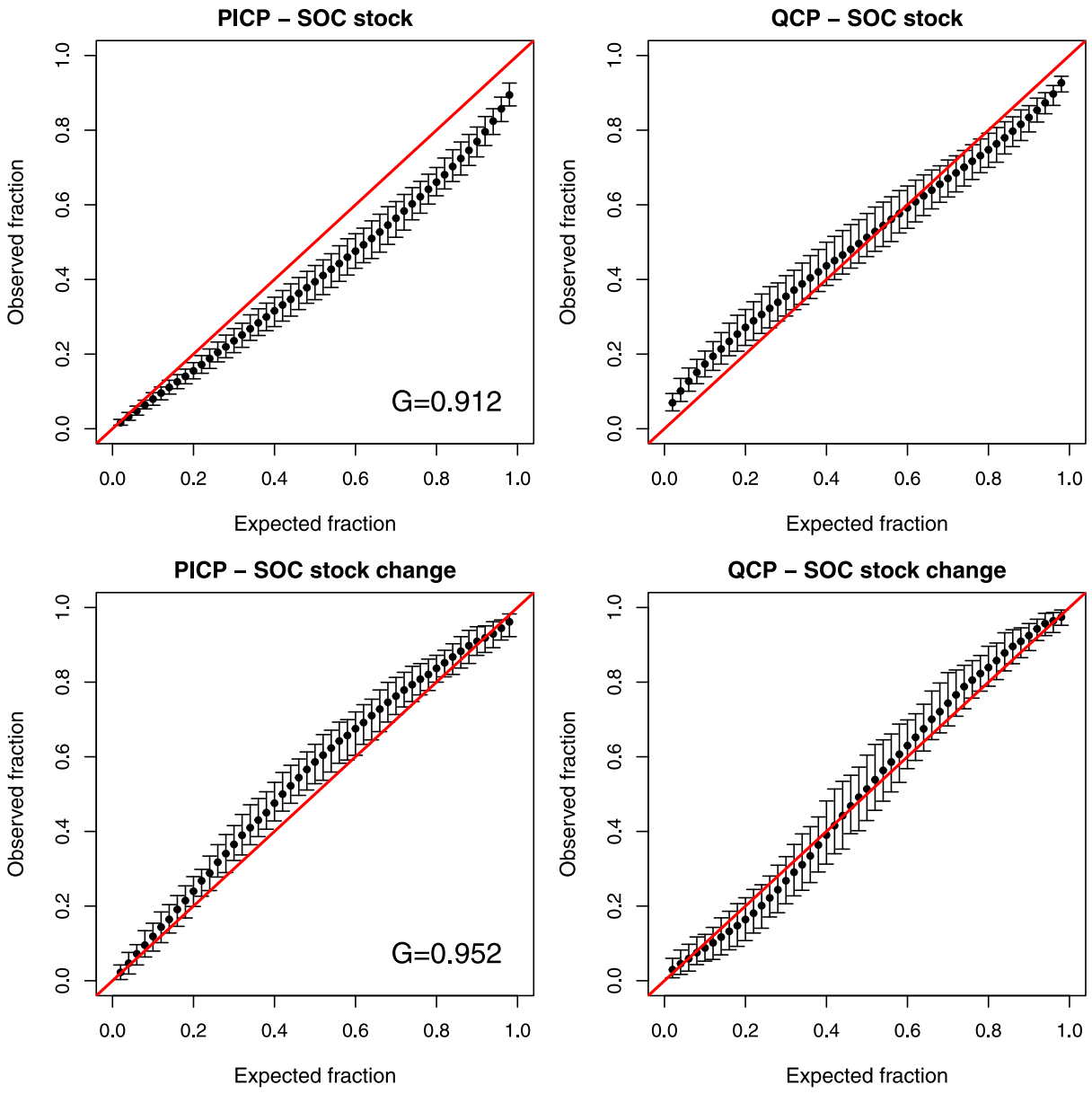


Fig. 6. The prediction interval coverage probability (PICP) plot with the computed G-statistic value (left column) and quantile coverage probability (QCP) plot (right column) for SOC stock (first row) and SOC stock change (last row). Annotation: The error bars were derived from the 5 times repeated 10-fold cross-validation, and the red solid lines show the 1:1 line (i.e., ideal case). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

quantifications do not suffer from one-sided bias over the period under study.

3.3. Spatial aggregation of SOC stock at various supports

After carrying out spatial aggregation, we compiled a series of maps of spatially aggregated SOC stock predictions between 1992 and 2016 for each support, together with a series of maps showing the quantified prediction uncertainty. The map series can be found in the Supplementary Material (SM4–15). In Fig. 7, we present the spatially aggregated SOC stock change between the start and end year of the period under study (i.e., 1992 and 2016). As expected, the larger the support, the smaller the uncertainty associated with the spatially aggregated SOC stock change predictions (Fig. 7, second column) and, consequently, supports to identify and delineate larger areas with statistically significant change (Fig. 7, third column). Indeed, there is a difference of three orders of magnitude in the quantified uncertainty between the first and last rows of Fig. 7 (second column), which is a remarkable decline of uncertainty. It is important to emphasize that Fig. 7 serves as an illustrative example, showing the potential and possibilities of our recent work. Specifically, such investigations and assessments can be conducted for any year, time period, and spatial support, which is unprecedented in Hungary and, to the best of our knowledge, globally. In the Supplementary Material (SM16), we provided a similar assessment for SOC stock change for the period between 1992 and 2010, which was the focus of our previous research (Szatmári et al., 2021). This allows for a direct comparison between our recent results and those from previous study (Szatmári et al., 2021). Furthermore, we could also demonstrate, for example, how SOC stock changed year by year at a given spatial support (e.g., 10 × 10 km) during the respective period (Fig. 8).

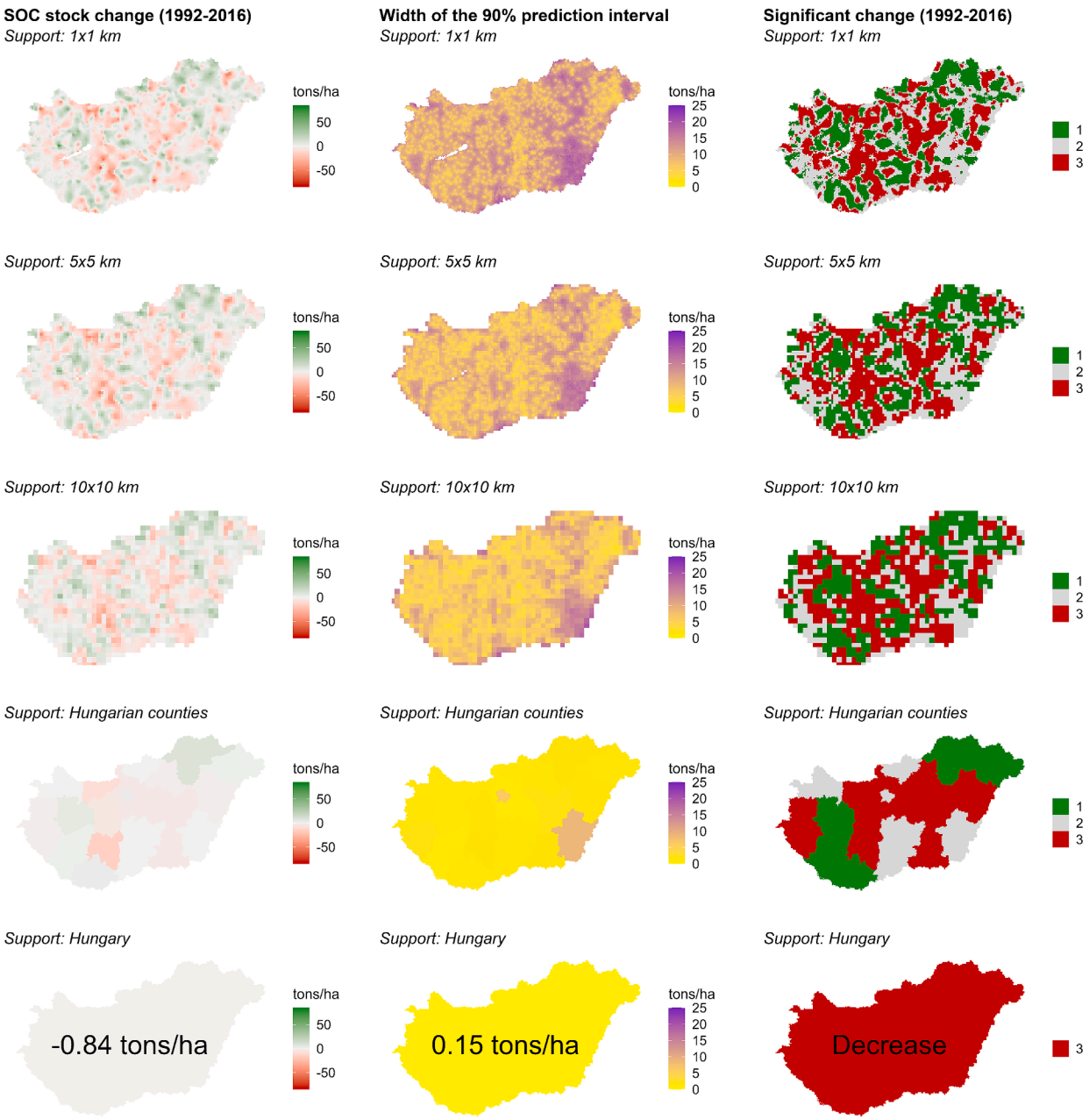


Fig. 7. Spatial aggregation of soil organic carbon (SOC) stock change between 1992 and 2016 at various supports. Spatially aggregated predictions of SOC stock change (first column), the associated uncertainty (second column), and the maps of statistically significant change in SOC stock (third column). Annotation: The codes in the third column are: 1 – Increase, 2 – No significant change, and 3 – Decrease.

4. Discussion

4.1. Interpretation and comparison of the results

It was observed that the static environmental covariates were more influential than the dynamic covariates in the final, fine-tuned RF model (Fig. 3, right graph). Note that special care was taken to avoid overfitting the RF model with static environmental covariates, using leave-location-out cross-validation instead of "classical" cross-validation (Meyer et al., 2018). The predominance of static covariates in the RF model can be attributed to the larger spatial variability of SOC stock compared to its temporal variability, as it was noted by Heuvelink et al. (2020). Regarding the informative covariates, both static and dynamic, identified in our study, they align with findings from studies using RF or other machine learning techniques (e.g., cubist) to model spatiotemporal variability of SOC (Helfenstein et al., 2024; Heuvelink et al., 2020; Tayebi et al., 2021; Zhang et al., 2023a). These covariates are parent material (or geology), soil type (or soil class), digital elevation model and its derivatives (e.g., ridge top and valley bottom flatness, topographic wetness index), climatic data (e.g., seasonal temperature, precipitation), and satellite images and derived indices (e.g., short-wave infrared bands, NDVI). Clearly, the relative importance of these covariates varies from case to case because these SOC mapping campaigns took place in different parts of the world (i.e., Argentina, Brazil, China, and the Netherlands), reflecting diverse conditions of climate, geology, topography, and biotic factors. Moreover, the target resolution of these campaigns ranges widely, from 25 m to 1 km.

In line with our previous findings (Szatmári et al., 2021), it was shown that spatial aggregation significantly reduces the uncertainty associated with spatially aggregated SOC stock and SOC stock change

Fig. 8. Series of maps showing areas with statistically significant change in SOC stock. The maps present how SOC stock changed year after year over 10 × 10 km square blocks between 1992 and 2016. Color key to the maps: Green – Increase, Grey – No significant change, and Red – Decrease. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

predictions (Fig. 7 and SM4–15), even in space–time applications. Generally, it can be said that the larger the support, the smaller the uncertainty associated with the spatially aggregated predictions. In fact, there was a difference of three orders of magnitude between the quantified uncertainties associated with point support predictions and when they were spatially aggregated over the whole territory of Hungary. This is due to the partial cancellation of negative and positive interpolation errors during spatial aggregation, a well-known

phenomenon in geostatistics (Goovaerts, 1997; Webster and Oliver, 2007). It is noteworthy that the decreasing uncertainty enables us to identify and delineate more areas with statistically significant SOC stock changes in Hungary, as it was showed in Fig. 7 (third column) and Fig. 8.

In addition to the validation of SOC stock predictions, SOC stock change predictions were also validated using 5 times repeated 10-fold cross-validation (Table 5). In terms of MEC, the SOC stock predictions (MEC = 0.321) were found to be more accurate than SOC stock change predictions (MEC = 0.160), which is consistent with recent findings (e.g., Helfenstein et al., 2024). This suggests that predicting SOC stock change is more challenging than predicting SOC stock, which is an important finding because an increasing number of end-users and initiatives are interested in SOC stock change in addition to or instead of SOC stock. However, the uncertainty quantification associated with SOC stock change predictions is proved to be accurate (Fig. 6, last row), and it can be said that the uncertainty estimates at point support reliably quantify the "real" uncertainty. For SOC stock (Fig. 6, first row), the uncertainty was slightly underestimated. It is important to note that uncertainty was quantified solely using the kriging variance, which implicitly assumes no uncertainty in the deterministic part of variation (i.e., first term on the right-hand side of Eq. (2)). This can sometimes be an over simplistic assumption (Szatmári and Pásztor, 2019). Therefore, incorporating uncertainty in the deterministic part into space–time modeling is advisable.

It is challenging to directly compare the current results with those published previously (Szatmári et al., 2021; 2019), as such a comprehensive map series of the spatiotemporal variability of SOC stock has not been available for Hungary before. As mentioned in the Introduction, in our previous research, SOC data were available for only two years (1992 and 2010), making it impossible to develop a space–time model capable of predicting SOC stock annually. Szatmári et al. (2021) used a linear model of coregionalization (Goovaerts, 1997) to account for the correlation of the interpolation errors in predicting SOC stock changes at various supports. Comparing those results with the current ones (SM16) reveals significant differences in SOC stock change, associated uncertainty, and areas of significant change. Although the magnitude is similar, the patterns of SOC stock change and uncertainty differ. These differences can be attributed to several factors. First, this study included more SOC stock data and environmental covariates, using annual data for climate and biosphere instead of long-term averages. Second, the methodologies were very different. Regarding uncertainty, Szatmári and Pásztor (2019) demonstrated that the methodologies and assumptions made in quantifying uncertainty can lead to very different uncertainty predictions. The space–time model in this study resulted in relatively smaller prediction uncertainty, allowing for the identification and delineation of more areas with significant change than in our previous study.

It is important to highlight the main limitation of the approach presented in our previous study (Szatmári et al., 2021): SOC stock and SOC stock change, along with quantified uncertainty, whether at point support or larger supports, can only be predicted for years or time periods where SOC stock observations are available. However, there is no guarantee that SOC stock observations will be available for all years that are of interest to end-users. Therefore, this approach is sub-optimal from this point of view. In contrast, the methodology presented in our recent work, which combines space–time geostatistics with machine learning, allows us to predict SOC stock and SOC stock change for any year, time period, and spatial support (Figs. 7 and 8, and SM4–16).

It follows from the above that such comprehensive, coherent, spatially and temporally exhaustive information on SOC stock in Hungary has never been available before. This makes our study and the resulting map series unique and important contribution from a national perspective. It also paves the way for a comprehensive analysis and assessment of the spatial and temporal variability of SOC stored in Hungarian topsoils. This information is crucial for Hungary, and perhaps Europe, to address contemporary environmental crises (e.g., climate change, land and soil degradation, desertification) and challenges (e.g., water-, food-, and soil security).

4.2. General discussion

First and foremost, it is important to emphasize that without accounting for the spatial or spatiotemporal correlation of interpolation errors, it is impossible to reliably quantify the uncertainty associated with spatially aggregated SOC stock or SOC stock change predictions. Note that this principle generally applies to all other soil properties as well. If this correlation is ignored, the quantified uncertainty will not represent the "true" uncertainty. Moreover, Wadoux and Heuvelink (2023) noted that uncertainty will be underestimated. The serious consequences of making decisions based on such unrealistic uncertainty quantification cannot be overstated.

In this study, a methodology combining space–time geostatistics with machine learning was presented and tested on the example of Hungary predicting the spatiotemporal variability of SOC stock and SOC stock change at various supports. In fact, the use of geostatistics in digital soil mapping has been in decline, with most recent studies on spatiotemporal modeling relying solely on machine learning techniques (Helfenstein et al., 2024; 2022; Heuvelink et al., 2020; Tayebi et al., 2021; Yang et al., 2023). However, our study highlighted an important limitation of applying machine learning in spatiotemporal modelling: machine learning alone is inadequate when it comes to quantifying the uncertainty associated with SOC change predictions at point support or when predicting SOC and SOC change at larger spatial scales with quantified uncertainty. Indeed, machine learning techniques are unable to account for the spatial or spatiotemporal correlation of errors, which is essential for reliably quantifying these uncertainties, and to the best of our knowledge, no study has demonstrated otherwise. The methodology presented here effectively addresses these challenges. Space-time geostatistics is specifically designed to account for the spatiotemporal correlation of errors (Cressie, 1993; Gräler et al., 2016; Kyriakidis and Journel, 1999; Webster and Oliver, 2007), and this is the key advantage of using geostatistics in this study. Furthermore, it overcomes the limitations of our previous approach (Szatmári et al., 2021), allowing for the prediction of SOC stock and SOC stock change, with quantified uncertainty, for any year, time period, and spatial scale (Figs. 7 and 8, and SM4–16), thereby meeting a wide range of end-user needs. Note that there are other approaches, tested so far in 2D and 3D aggregation, that do not use geostatistics, such as the Gaussian process regression (Wang et al., 2024) or the Monte Carlo integration of the spatial autocorrelation of the map errors (Wadoux and Heuvelink, 2023). An interesting question is whether these methods can be extended to space–time applications.

Geostatistical simulation was used to generate equally probable stochastic realizations for the given 2D + T blocks, which were regularly or irregularly shaped and varied in size (Table 2). This formed the basis for predicting the spatial average of SOC stock and SOC stock change with quantified prediction uncertainty. Although sequential stochastic simulation (Deutsch and Journel, 1998; Goovaerts, 1997) is often regarded as the gold standard for generating stochastic realizations both univariate and multivariate cases (Angelini et al., 2023; Goovaerts, 2001; Heuvelink, 1998; Heuvelink et al., 2016; Poggio and Gimona, 2014; Szatmári et al., 2021; Szatmári and Pásztor, 2019), we applied 2D + T LU simulation, which relies on the Cholesky decomposition of the spatiotemporal covariance matrix. This method was preferred over sequential simulation because it is a simple technique that is easy to use and implement, can be faster, requires less computing capacity based on our experience, and allows for the generation of a large number of stochastic realizations once the covariance matrix has been decomposed. These factors make it very attractive and time-efficient for working in the space–time domain.

As mentioned in the Introduction, there is a growing demand for information on SOC and, more importantly, on SOC change, both

nationally and internationally. It is well known that the spatial and temporal scales at which such information is required vary widely, highlighting the need for dynamic, scalable information on SOC and SOC change, with quantified uncertainty, over time. Indeed, numerous examples can be given, from GHG inventories (spatial total for the country, reported annually), ecosystem accounts (spatial averages for total grasslands and arable lands, reported every 3–4 years), through crop simulations, terrestrial ecosystem process models, and climate models (desired resolution: from 100 m to 10 km) (Fodor et al., 2014; Hidy et al., 2016; Koós et al., 2021), to carbon farming and sustainable soil and land management (desired resolution: $\leq$ 100 m) (de Gruijter et al., 2016; Malone et al., 2018). To the best of our knowledge, it is not currently mandatory to report uncertainty associated with SOC or SOC change predictions (e.g., in GHG inventories, carbon accounting, and ecosystem condition accounts). However, we believe that involving this uncertainty is crucial for informed decision-making, and therefore, propose that the associated uncertainty should also be reported. As demonstrated, the methodology presented in this study is capable of meeting these requirements, thus addressing an important knowledge gap in the spatiotemporal modelling of SOC.

Several studies have pointed out that the uncertainty associated with point support predictions is too high to reliably identify areas with statistically significant SOC change (Heuvelink et al., 2020; Szatmári et al., 2021). Indeed, our results also revealed a significant difference in uncertainty between point support predictions and spatially aggregated predictions over larger supports. Therefore, it is noteworthy to highlight that spatial aggregation can greatly reduce uncertainty, thus supporting the detection and delineation of larger areas with statistically significant changes in SOC stock (Figs. 7 and 8). However, it should be noted that an important drawback of spatial aggregation is the loss of spatial detail. Consequently, spatial aggregation is not suitable in cases where decisions or actions need to be based on highly detailed spatial information on SOC or SOC change, such as carbon farming or evaluation of SOC conservation strategies. This issue will be addressed in the following subsection.

4.3. Limitations and future research

In this study, a relatively large set of static and dynamic environmental covariates was used. However, capturing the temporal variation of SOC is challenging. In fact, the temporal variation of SOC is influenced not only by present conditions but also by the delayed or long-term effects of factors and processes that occurred or changed in the past. For example, land use change significantly influences SOC, although these effects often manifest later in time. In this work, we did not account for the long-term effects of soil forming factors, which should be addressed in future research to improve the performance of the developed space–time model. Nonetheless, there are promising approaches aimed at incorporating the long-term effects of the soil forming factors. For example, Heuvelink et al. (2020) applied weighted averages to dynamic covariates across multiple years using exponential decay functions, whereas Helfenstein et al. (2024) considered land use changes over 5, 10, 20, and 40-year periods based on historical land use maps. It should be pointed out that, unfortunately, many national and international initiatives link SOC stock changes to the annual changes in land use, as proposed by the LULCF (land use, land-use change, and forestry) framework. We believe that more research is needed to better understand the temporal variation of SOC and how the delayed or long-term effects of certain factors (e.g., land use change, and declining precipitation) influences changes in SOC.

Although spatial aggregation is a potential way to reduce uncertainty, the issue of reducing uncertainty at point support or finer scales remains unresolved. Addressing this challenge is becoming increasingly urgent, as a growing number of applications (e.g., carbon farming, sustainable agriculture, land use planning) and related decision-making processes and policy formulations require reliable information on SOC and SOC change at finer spatial scales. Two factors further complicate this issue. First, there are additional sources of uncertainty that are rarely considered in the spatiotemporal modeling of SOC. As mentioned in the Introduction, uncertainty related to SOC stock data may not be negligible due to sampling issues, measurement errors in SOC and bulk density, etc. (Knotters et al., 2022; Paul et al., 2023). Second, accurately predicting SOC change is difficult, which makes it challenging to reliably track SOC changes over time at finer scales (Helfenstein et al., 2024). In the near future, emphasis should be placed on reducing uncertainty at finer scales and incorporating all possible sources of uncertainty into the modelling process.

We could only validate the predictions and uncertainty quantifications at point support, as we had no SOC stock observations from probability sampling, which are essential for validating block predictions (Brus et al., 2011; Vaysse et al., 2017). Cross-validation statistics at point support can provide an indication of the quality of block support predictions; however, this method is not always reliable for all cases. Therefore, more effort should be made to collect validation data at block support. Collecting such validation data through probability sampling would enable a more comprehensive assessment of prediction accuracy across different spatial scales. This approach aligns with the goals outlined in Pedometrics Challenge 5 (Wadoux et al., 2021), which emphasizes the need for improved validation methods to better support soil property predictions. Implementing probability sampling in future studies would not only strengthen the validation process but also help in refining the predictive models used for SOC stock assessments.

5. Conclusions

Our objective was to present a methodology for predicting the spatiotemporal variability of SOC stock and SOC stock change with quantified uncertainty at various aggregation levels, using Hungary as an example, and to identify and delineate areas with statistically significant SOC stock changes. We used a combination of machine learning and space–time geostatistics to predict SOC stock at point support, then spatial aggregation was performed using conditional 2D + T LU simulation, which relies on Cholesky decomposition of the spatiotemporal covariance matrix. The main conclusions drawn from this study can be summarized as follows:

- *Importance of spatiotemporal correlation:* It is crucial to account for the spatiotemporal correlation of errors to reliably quantify the uncertainty associated with SOC change predictions at point support, as well as the uncertainty related to spatially aggregated SOC or SOC change predictions. Ignoring this correlation results in unrealistic uncertainty quantifications, which can have serious consequences for decision-making processes.
- *Integration of geostatistics and machine learning:* The combination of machine learning and space–time geostatistics offers a robust methodology for predicting the spatiotemporal variability of SOC across various supports. While machine learning alone is insufficient for quantifying the uncertainty associated with SOC change or spatially aggregated SOC and SOC change predictions, geostatistics fills this gap by accounting for the spatiotemporal correlation of errors.
- *Addressing dynamic SOC information needs:* The methodology presented in this study partially meets the increasing demand for dynamic information on SOC and SOC changes across various spatial scales and time periods. By providing comprehensive predictions and quantifying the associated uncertainties, it addresses both current and anticipated demands at national and international levels.
- *Exhaustive information on SOC for Hungary:* The series of maps and the information that can be derived from them are unprecedented and meet the crucial demand for spatially and temporally exhaustive information in Hungary. This also paves the way for a comprehensive

analysis and deeper understanding of the spatiotemporal variability of SOC stored in Hungarian soils.

CRediT authorship contribution statement

Gábor Szatmári: Writing – review & editing, Writing – original draft, Supervision, Software, Methodology, Funding acquisition, Formal analysis, Conceptualization. **László Pásztor:** Writing – review & editing, Supervision, Resources, Methodology, Conceptualization. **Katalin Takács:** Writing – original draft, Visualization, Validation, Investigation, Data curation. **János Mészáros:** Writing – original draft, Software, Data curation. **András Benő:** Investigation, Data curation. **Annamária Laborczi:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Conceptualization.

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.

Acknowledgements

This research was funded by the National Research, Development and Innovation Office (NKFIH, grant number: FK-146391) and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences (Szatmári, G.). The authors thank Judit Matus for her indispensable contribution.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geoderma.2024.117067.

Data availability

The authors do not have permission to share data.

References

Alabert, F., 1987. The practice of fast conditional simulations through the LU decomposition of the covariance matrix. Math. Geol. 19, 369–386. https://doi.org/10.1007/BF00897191/METRICS.

Angelini, M.E., Heuvelink, G.B.M., Lagacherie, P., 2023. A multivariate approach for mapping a soil quality index and its uncertainty in southern France. Eur. J. Soil Sci. 74, e13345.

Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B.M., Hong, S. Y., Lagacherie, P., Lelyk, G., McBratney, A.B., McKenzie, N.J., Mendonca-Santos, M. d.L., Minasny, B., Montanarella, L., Odeh, I.O.A., Sanchez, P.A., Thompson, J.A., Zhang, G.L., 2014. GlobalSoilMap. Toward a fine-resolution global grid of soil properties. Adv. Agron. 125, 93–134. https://doi.org/10.1016/B978-0-12-800137-0.00003-0.

Bakacsi, Z., Laborczi, A., Szabó, J., Takács, K., Pásztor, L., 2014. Az 1:100 000-es földtani térkép jelkulcsának és a FAO rendszer talajképző kőzet kódrendszerének javasolt megfeleltetése. Agrokémia És Talajt. 63, 189–202.

Bishop, T.F.A., McBratney, A.B., Laslett, G.M., 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91, 27–45. https://doi.org/10.1016/S0016-7061(99)00003-8.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. Eur. J. Soil Sci. 62, 394–407. https://doi.org/10.1111/j.1365-2389.2011.01364.x.

Centeri, C., Szabó, B., Jakab, G., Kovács, J., Madarász, B., Szabó, J., Tóth, A., Gelencsér, G., Szalai, Z., Vona, M., 2014. State of soil carbon in Hungarian sites: Loss, pool and management, Soil Carbon: Types, Management Practices and Environmental Benefits.

Chen, J., Biswas, A., Su, H., Cao, J., Hong, S., Wang, H., Dong, X., 2023. Quantifying changes in soil organic carbon density from 1982 to 2020 in Chinese grasslands using a random forest model. Front. Plant Sci. 14, 1–16. https://doi.org/10.3389/fpls.2023.1076902.

Copernicus Climate Change Service - Climate Data Store, 2019. Land cover classification gridded maps from 1992 to present derived from satellite observation [WWW Document]. Copernicus Clim. Chang. Serv. Clim. Data Store. https://doi.org/https://doi.org/10.24381/cds.006f2c9a.

Cornes, R.C., van der Schrier, G., van den Besselaar, E.J.M., Jones, P.D., 2018. An ensemble version of the E-OBS temperature and precipitation data sets. J. Geophys. Res. Atmos. 123, 9391–9409. https://doi.org/10.1029/2017JD028200.

Cressie, N.A.C., 1993. Statistics for Spatial Data. Wiley.

Cressie, N., 2006. Block kriging for lognormal spatial processes. Math. Geol. 38, 413–443. https://doi.org/10.1007/s11004-005-9022-8.

Csikós, N., Szabó, B., Hermann, T., Laborczi, A., Matus, J., Pásztor, L., Szatmári, G., Takács, K., Tóth, G., 2023. Cropland Productivity Evaluation: A 100 m Resolution Country Assessment Combining Earth Observation and Direct Measurements. Remote Sens. 2023, Vol. 15, Page 1236 15, 1236. https://doi.org/10.3390/RS15051236.

Davis, M.W., 1987. Production of conditional simulations via the LU triangular decomposition of the covariance matrix. Math. Geol. 19, 91–98. https://doi.org/10.1007/BF00898189/METRICS.

de Gruijter, J.J., McBratney, A.B., Minasny, B., Wheeler, I., Malone, B.P., Stockmann, U., 2016. Farm-scale soil carbon auditing. Geoderma 265, 120–130. https://doi.org/10.1016/j.geoderma.2015.11.010.

De Rosa, D., Ballabio, C., Lugato, E., Fasiolo, M., Jones, A., Panagos, P., 2024. Soil organic carbon stocks in European croplands and grasslands: how much have we lost in the past decade? Glob. Chang. Biol. 30, 1–15. https://doi.org/10.1111/gcb.16992.

Deutsch, C.V., 1997. Direct assessment of local accuracy and precision. In: Baafi, E.Y., Schofield, N.A. (Eds.), Geostatstics Wollongong '96. Kluwer Academic Publishers, pp. 115–125.

Deutsch, C.V., Journel, A.G., 1998. GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press.

Erdélyi, D., Kern, Z., Nyitrai, T., Hatvani, I.G., 2023. Predicting the spatial distribution of stable isotopes in precipitation using a machine learning approach: a comparative assessment of random forest variants. GEM - Int. J. Geomathematics 14, 1–19. https://doi.org/10.1007/S13137-023-00224-X/TABLES/1.

European Environment Agency, 2016. EU-DEM (raster) - version 1.0, Apr. 2016 [WWW Document]. URL https://sdi.eea.europa.eu/catalogue/srv/api/records/3473589f-0854-4601-919e-2e7dd172ff50.

Fodor, N., Pásztor, L., Németh, T., 2014. Coupling the 4M crop model with national geo-databases for assessing the effects of climate change on agro-ecological characteristics of Hungary. Int. J. Digit. Earth 7, 391–410. https://doi.org/10.1080/17538947.2012.689998.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press.

Goovaerts, P., 2001. Geostatistical modelling of uncertainty in soil science. Geoderma 103, 3–26. https://doi.org/10.1016/S0016-7061(01)00067-2.

Gräler, B., Pebesma, E., Heuvelink, G., 2016. Spatio-Temporal Interpolation using gstat. R J. 8, 204–218.

Guevara, M., Arroyo, C., Brunsell, N., Cruz, C.O., Domke, G., Equihua, J., Etchevers, J., Hayes, D., Hengl, T., Ibelles, A., Johnson, K., de Jong, B., Libohova, Z., Llamas, R., Nave, L., Ornelas, J.L., Paz, F., Ressl, R., Schwartz, A., Victoria, A., Wills, S., Vargas, R., 2020. Soil Organic Carbon Across Mexico and the Conterminous United States (1991–2010). Global Biogeochem. Cycles 34, no. https://doi.org/10.1029/2019GB006219.

Guevara, M., Federico Olmedo, G., Stell, E., Yigini, Y., Aguilar Duarte, Y., Arellano Hernández, C., Arévalo, G.E., Eduardo Arroyo-Cruz, C., Bolivar, A., Bunning, S., Bustamante Cañas, N., Omar Cruz-Gaistardo, C., Davila, F., Dell Acqua, M., Encina, A., Tacona, H.F., Fontes, F., Herrera, J.A.H., Roberto Ibelles Navarro, A., Loayza, V., Manueles, A.M., Mendoza Jara, F., Olivera, C., Osorio Hermosilla, R., Pereira, G., Prieto, P., Ramos, I.A., Carlos Rey Brina, J., Rivera, R., Rodríguez-Rodríguez, J., Roopnarine, R., Ibarra, A.R., Amaury Rosales Riveiro, K., Andrés Schulz, G., Spence, A., Vasques, G.M., Vargas, R.R., Vargas, R., 2018. No silver bullet for digital soil mapping: country-specific soil organic carbon estimates across Latin America. Soil 4, 173–193. https://doi.org/10.5194/soil-4-173-2018.

Gyalog, L., Síkhegyi, F., 2005. Magyarország földtani térképe, M=1:100 000 (Geological map of Hungary, M=1:100,000). Magyar Állami Földtani Intézet, Budapest.

Hartemink, A.E., McSweeney, K. (Eds.), 2014. Soil Carbon. Springer. https://doi.org/10.1007/978-3-319-04084-4.

Hatvani, I.G., Szatmári, G., Kern, Z., Erdélyi, D., Vreča, P., Kanduč, T., Czuppon, G., Lojen, S., Kohán, B., 2021. Geostatistical evaluation of the design of the precipitation stable isotope monitoring network for Slovenia and Hungary. Environ. Int. 146, 106263. https://doi.org/10.1016/j.envint.2020.106263.

Helfenstein, A., Mulder, V.L., Heuvelink, G.B.M., Okx, J.P., 2022. Tier 4 maps of soil pH at 25 m resolution for the Netherlands. Geoderma 410, 115659. https://doi.org/10.1016/J.GEODERMA.2021.115659.

Helfenstein, A., Mulder, V.L., Heuvelink, G.B.M., Hack-ten Broeke, M.J.D., 2024. Three-dimensional space and time mapping reveals soil organic matter decreases across anthropogenic landscapes in the Netherlands. Commun. Earth Environ. 5, 1–16. https://doi.org/10.1038/s43247-024-01293-y.

Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., De Jesus, J.M., Tamene, L., Tondoh, J.E., 2015. Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. PLoS One 10, 1–26. https://doi.org/10.1371/journal.pone.0125814.

Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., Gräler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 2018, e5518. https://doi.org/10.7717/peerj.5518.

Heuvelink, G.B.M., Kros, J., Reinds, G.J., De Vries, W., 2016. Geostatistical prediction and simulation of European soil property maps. Geoderma Reg. 7, 201–215. https://doi.org/10.1016/j.geodrs.2016.04.002.
Heuvelink, G.B.M., Pebesma, E., Gräler, B., 2017. Space-Time Geostatistics. Encycl. GIS 1919–1926. https://doi.org/10.1007/978-3-319-17885-1_1647.
Heuvelink, G.B.M., van Egmond, F.M., 2010. Space-time geostatistics for precision agriculture: a case study of NDVI mapping for a Dutch potato field. Geostat. Appl. Precis. Agric. 117–137. https://doi.org/10.1007/978-90-481-9133-8_5.
Heuvelink, G.B.M., Angelini, M.E., Poggio, L., Bai, Z., Batjes, N.H., van den Bosch, R., Bossio, D., Estella, S., Lehmann, J., Olmedo, G.F., Sanderman, J., 2020. Machine learning in space and time for modelling soil organic carbon change. Eur. J. Soil Sci. https://doi.org/10.1111/ejss.12998.
Heuvelink, G.B.M., Griffith, D.A., 2010. Space-time geostatistics for geography: a case study of radiation monitoring across parts of Germany. Geogr. Anal. 42, 161–179. https://doi.org/10.1111/J.1538-4632.2010.00788.X.
Heuvelink, G.B.M., 1998. Error propagation in environmental modelling with GIS, Error propagation in environmental modelling with GIS. Taylor and Francis. https://doi.org/10.4324/9780203016114.
Hidy, D., Barcza, Z., MarjanoviÄ, H., Sever, M.Z.O., Dobor, L., Gelybó, G., Fodor, N., Pintér, K., Churkina, G., Running, S., Thornton, P., Bellocchi, G., Haszpra, L., Horváth, F., Suyker, A., Nagy, Z., 2016. Terrestrial ecosystem process model Biome-BGCMuSo v4.0: summary of improvements and new modeling possibilities. Geosci. Model Dev. 9, 4405–4437. https://doi.org/10.5194/gmd-9-4405-2016.
Illés, G., Sutikno, S., Szatmári, G., Sandhyavitri, A., Pásztor, L., Kristijono, A., Molnár, G., Yusa, M., Székely, B., 2019. Facing the peat CO_2 threat: digital mapping of Indonesian peatlands—a proposed methodology and its application. J. Soils Sediments 19, 3663–3678. https://doi.org/10.1007/S11368-019-02328-0/FIGURES/11.
IUSS Working Group WRB, 2022. World Reference Base for Soil Resources. International soil classification system for naming soils and creating legends for soil maps (4th edition). Vienna, Austria.
Jakab, G., Szabó, J., Szalai, Z., Mészáros, E., Madarász, B., Centeri, C., Szabó, B., Németh, T., Sipos, P., 2016. Changes in organic carbon concentration and organic matter compound of erosion-delivered soil aggregates. Environ. Earth Sci. 75, 1–11. https://doi.org/10.1016/j.lrr.2016.07.003.
Jakab, G., Bede-Fazekas, Á., Vona, V., Madarász, B., Karlik, M., Zacháry, D., Filep, T., Dévény, Z., Centeri, C., Masoudi, M., Bidló, A., Al-Graiti, T., Szatmári, G., Vancsik, A., Király, C., Darabos, G., Angyal, Z., Szalai, Z., 2024. Beyond land use: understanding variations in topsoil bulk versus recalcitrant organic matter. Catena 244, 108232. https://doi.org/10.1016/J.CATENA.2024.108232.
Keesstra, S.D., Bouma, J., Wallinga, J., Tittonell, P., Smith, P., Cerdà, A., Montanarella, L., Quinton, J.N., Pachepsky, Y., Van Der Putten, W.H., Bardgett, R.D., Moolenaar, S., Mol, G., Jansen, B., Fresco, L.O., 2016. The significance of soils and soil science towards realization of the United Nations sustainable development goals. Soil 2, 111–128. https://doi.org/10.5194/SOIL-2-111-2016.
Keesstra, S., Mol, G., de Leeuw, J., Okx, J., Molenaar, C., de Cleen, M., Visser, S., 2018. Soil-related sustainable development goals: four concepts to make land degradation neutrality and restoration work. Land 7, 133. https://doi.org/10.3390/land7040133.
Knotters, M., Teuling, K., Reijneveld, A., Lesschen, J.P., Kuikman, P., 2022. Changes in organic matter contents and carbon stocks in Dutch soils, 1998–2018. Geoderma 414, 115751. https://doi.org/10.1016/J.GEODERMA.2022.115751.
Koós, S., Pirkó, B., Szatmári, G., Csathó, P., Magyar, M., Szabó, J., Fodor, N., Pásztor, L., Laborczi, A., Pokovai, K., Szabó, A., 2021. Influence of the shortening of the winter fertilization prohibition period in Hungary assessed by spatial crop simulation analysis. Sustainability 13, 417. https://doi.org/10.3390/SU13010417.
Kyriakidis, P.C., Journel, A.G., 1999. Geostatistical space-time models: a review. Math. Geol. 31, 651–684. https://doi.org/10.1023/A:1007528426688/METRICS.
Lal, R., 2004a. Soil carbon sequestration to mitigate climate change. Geoderma. https://doi.org/10.1016/j.geoderma.2004.01.032.
Lal, R., 2004b. Soil carbon sequestration impacts on global climate change and food secruity. Science (80-.) 304, 1623–1627.
Lal, R., Smith, P., Jungkunst, H.F., Mitsch, W.J., Lehmann, J., Nair, P.K.R., McBratney, A. B., de Moraes Sá, J.C., Schneider, J., Zinn, Y.L., Skorupa, A.L.A., Zhang, H.-L., Minasny, B., Srinivasrao, C., Ravindranath, N.H., 2018. The carbon sequestration potential of terrestrial ecosystems. J. Soil Water Conserv. 73, 145A–A152. https://doi.org/10.2489/jswc.73.6.145A.
Lark, R.M., Lapworth, D.J., 2012. Quality measures for soil surveys by lognormal kriging. Geoderma 173–174, 231–240. https://doi.org/10.1016/j.geoderma.2011.12.008.
Lefévre, C., Rekik, F., Alcantara, V., Wiese, L. (Eds.), 2017. Soil Organic Carbon: the Hidden Potential. FAO, Rome.
Malone, B., Hedley, C., Roudier, P., Minasny, B., Jones, E., McBratney, A., 2018. Auditing on-farm soil carbon stocks using downscaled national mapping products: examples from Australia and New Zealand. Geoderma Reg. 13, 1–14. https://doi.org/10.1016/J.GEODRS.2018.02.002.
McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma. https://doi.org/10.1016/S0016-7061(03)00223-4.
Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., Nauss, T., 2018. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. Environ. Model. Softw. 101, 1–9. https://doi.org/10.1016/J.ENVSOFT.2017.12.001.
Minasny, B., Berglund, Ö., Connolly, J., Hedley, C., de Vries, F., Gimona, A., Kempen, B., Kidd, D., Lilja, H., Malone, B., McBratney, A., Roudier, P., O'Rourke, S., Rudiyanto, Padarian, J., Poggio, L., ten Caten, A., Thompson, D., Tuve, C., Widyatmanti, W., 2019. Digital mapping of peatlands – A critical review. Earth-Sci. Rev. 196, 102870. https://doi.org/10.1016/J.EARSCIREV.2019.05.014.
Minasny, B., McBratney, A.B., Malone, B.P., Wheeler, I., 2013. Digital Mapping of Soil Carbon, in: Sparks, D.L. (Ed.), Advances in Agronomy. Elsevier, pp. 1–47. https://doi.org/10.1016/B978-0-12-405942-9.00001-3.
Minasny, B., Malone, B.P., McBratney, A.B., Angers, D.A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z.S., Cheng, K., Das, B.S., Field, D.J., Gimona, A., Hedley, C.B., Hong, S.Y., Mandal, B., Marchant, B.P., Martin, M., McConkey, B.G., Mulder, V.L., O'Rourke, S., Richer-de-Forges, A.C., Odeh, I., Padarian, J., Paustian, K., Pan, G., Poggio, L., Savin, I., Stolbovoy, V., Stockmann, U., Sulaeman, Y., Tsui, C.C., Vågen, T.G., van Wesemael, B., Winowiecki, L., 2017. Soil carbon 4 per mille. Geoderma 292, 59–86. https://doi.org/10.1016/j.geoderma.2017.01.002.
MSZ–08–0452:1980, 1980. A talaj szerves széntartalmának meghatározása (Determination of the organic carbon content of soil), Magyar Szabványügyi Testület.
Orton, T.G., Pringle, M.J., Page, K.L., Dalal, R.C., Bishop, T.F.A., 2014. Spatial prediction of soil organic carbon stock using a linear model of coregionalisation. Geoderma 230–231, 119–130. https://doi.org/10.1016/j.geoderma.2014.04.016.
Orton, T.G., Pringle, M.J., Allen, D.E., Dalal, R.C., Bishop, T.F.A., 2015. A geostatistical method to account for the number of aliquots in composite samples for normal and lognormal random variables. Eur. J. Soil Sci. 66, 1023–1032. https://doi.org/10.1111/ejss.12297.
Padarian, J., Stockmann, U., Minasny, B., McBratney, A.B., 2022. Monitoring changes in global soil organic carbon stocks from space. Remote Sens. Environ. 281, 113260. https://doi.org/10.1016/j.rse.2022.113260.
Pásztor, L., Laborczi, A., Takács, K., Szatmári, G., Dobos, E., Illés, G., Bakacsi, Z., Szabó, J., 2015. Compilation of novel and renewed, goal oriented digital soil maps using geostatistical and data mining tools. Hungarian Geogr. Bull. 64, 49–64. https://doi.org/10.15201/hungeobull.64.1.5.
Pásztor, L., Laborczi, A., Takács, K., Illés, G., Szabó, J., Szatmári, G., 2020. Progress in the elaboration of GSM conform DSM products and their functional utilization in Hungary. Geoderma Reg. https://doi.org/10.1016/j.geodrs.2020.e00269.
Paul, C., Bartkowski, B., Dönmez, C., Don, A., Mayer, S., Steffens, M., Weigl, S., Wiesmeier, M., Wolf, A., Helming, K., 2023. Carbon farming: Are soil carbon certificates a suitable tool for climate change mitigation? J. Environ. Manage. 330, 117142. https://doi.org/10.1016/J.JENVMAN.2022.117142.
Poggio, L., De Sousa, L.M., Batjes, N.H., Heuvelink, G.B.M., Kempen, B., Ribeiro, E., Rossiter, D., 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. Soil 7, 217–240. https://doi.org/10.5194/soil-7-217-2021.
Poggio, L., Gimona, A., 2014. National scale 3D modelling of soil organic carbon stocks with uncertainty propagation - An example from Scotland. Geoderma 232–234, 284–299. https://doi.org/10.1016/j.geoderma.2014.05.004.
Schillaci, C., Acutis, M., Lombardo, L., Lipani, A., Fantappiè, M., Märker, M., Saia, S., 2017. Spatio-temporal topsoil organic carbon mapping of a semi-arid Mediterranean region: the role of land use, soil texture, topographic indices and the in fl uence of remote sensing data to modelling. Sci. Total Environ. 601–602, 821–832. https://doi.org/10.1016/j.scitotenv.2017.05.239.
Schmidinger, J., Heuvelink, G.B.M., 2023. Validation of uncertainty predictions in digital soil mapping. Geoderma 437, 116585. https://doi.org/10.1016/J.GEODERMA.2023.116585.
Shrestha, D.L., Solomatine, D.P., 2006. Machine learning approaches for estimation of prediction interval for the model output. Neural Networks 19, 225–235. https://doi.org/10.1016/J.NEUNET.2006.01.012.
Snepvangers, J.J.J.C., Heuvelink, G.B.M., Huisman, J.A., 2003. Soil water content interpolation using spatio-temporal kriging with external drift. Geoderma 112, 253–271. https://doi.org/10.1016/S0016-7061(02)00310-5.
Stavi, I., Lal, R., 2015. Achieving zero net land degradation: challenges and opportunities. J. Arid Environ. 112, 44–51. https://doi.org/10.1016/j.jaridenv.2014.01.016.
Stefanovits, P., 1963. Magyarország Talajai, second ed. Akadémia Kiadó, Budapest.
Stockmann, U., Adams, M.A., Crawford, J.W., Field, D.J., Henakaarchchi, N., Jenkins, M., Minasny, B., McBratney, A.B., Courcelles, V.de.R.de., Singh, K., Wheeler, I., Abbott, L., Angers, D.A., Baldock, J., Bird, M., Brookes, P.C., Chenu, C., Jastrow, J.D., Lal, R., Lehmann, J., O'Donnell, A.G., Parton, W.J., Whitehead, D., Zimmermann, M., 2013. The knowns, known unknowns and unknowns of sequestration of soil organic carbon. Agric. Ecosyst. Environ. 164, 80–99. https://doi.org/10.1016/j.agee.2012.10.001.
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.
Szatmári, G., Pásztor, L., 2019. Comparison of various uncertainty modelling approaches based on geostatistics and machine learning algorithms. Geoderma 337, 1329–1340. https://doi.org/10.1016/j.geoderma.2018.09.008.
Szatmári, G., Pirkó, B., Koós, S., Laborczi, A., Bakacsi, Z., Szabó, J., Pásztor, L., 2019. Spatio-temporal assessment of topsoil organic carbon stock change in Hungary. Soil Tillage Res. 195, 104410. https://doi.org/10.1016/j.still.2019.104410.
Szatmári, G., Pásztor, L., Heuvelink, G.B.M., 2021. Estimating soil organic carbon stock change at multiple scales using machine learning and multivariate geostatistics. Geoderma 403, 115356. https://doi.org/10.1016/J.GEODERMA.2021.115356.
Szatmári, G., Pásztor, L., Laborczi, A., Illés, G., Bakacsi, Z., Zacháry, D., Filep, T., Szalai, Z., Jakab, G., 2023. Countrywide mapping and assessment of organic carbon saturation in the topsoil using machine learning-based pedotransfer function with uncertainty propagation. Catena 227, 107086. https://doi.org/10.1016/J.CATENA.2023.107086.
Tayebi, M., Rosas, J.T.F., Mendes, W.de.S., Poppiel, R.R., Ostovari, Y., Ruiz, L.F.C., Dos Santos, N.V., Cerri, C.E.P., Silva, S.H.G., Curi, N., Silvero, N.E.Q., Demattê, J.A.M.,

2021. Drivers of organic carbon stocks in different lulc history and along soil depth for a 30 years image time series. Remote Sens. 13, 1–32. https://doi.org/10.3390/rs13112223.

Tian, X., de Bruin, S., Simoes, R., Isik, M.S., Minarik, R., Ho, Y.-F., Sahin, M., Herold, M., Consoli, D., Hengl, T., 2024. Spatiotemporal prediction of soil organic carbon density for Europe (2000–2022) in 3D+T based on Landsat-based spectral indices time-series. Preprint. https://doi.org/https://doi.org/10.21203/rs.3.rs-5128244/v1.

Tóth, E., Gelybó, G., Dencső, M., Kása, I., Birkás, M., Horel, Á., 2018. Soil CO_2 emissions in a long-term tillage treatment experiment. Soil Manag. Clim. Chang. 293–307. https://doi.org/10.1016/B978-0-12-812128-3.00019-7.

Urbina-Salazar, D., Vaudour, E., Richer-de-Forges, A.C., Chen, S., Martelet, G., Baghdadi, N., Arrouays, D., 2023. Sentinel-2 and sentinel-1 bare soil temporal mosaics of 6-year periods for soil organic carbon content mapping in Central France. Remote Sens. 15, 1–24. https://doi.org/10.3390/rs15092410.

Vaysse, K., Heuvelink, G.B.M., Lagacherie, P., 2017. Spatial aggregation of soil property predictions in support of local land management. Soil Use Manag. 33, 299–310. https://doi.org/10.1111/sum.12350.

Veronesi, F., Schillaci, C., 2019. Comparison between geostatistical and machine learning models as predictors of topsoil organic carbon with a focus on local uncertainty estimation. Ecol. Indic. 101, 1032–1044. https://doi.org/10.1016/j.ecolind.2019.02.026.

Wackernagel, H. 2003. Multivariate Geostatistics, Multivariate Geostatistics. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-05294-5.

Wadoux, A.M..-C., Courteille, L., Arrouays, D., Carvalho De Gomes, L., Cortet, J., Creamer, R.E., Eberhardt, E., Greve, M.H., Grüneberg, E., Harhoff, R., Heuvelink, G. B.M., Krahl, I., Lagacherie, P., Miko, L., Mulder, V.L., Pásztor, L., Pieper, S., Richer-de-Forges, A.C., Sánchez-Rodriguez, A.R., Rossiter, D.G., Steinhoff-Knopp, B., Stöckhardt, S., Szatmári, G., Takács, K., Tsiafouli, M., Vanwalleghem, T., Wellbrock, N., Wetterlind, J., 2024. On soil districts. Geoderma (accepted manuscript).

Wadoux, A.M.J.C., Heuvelink, G.B.M., 2023. Uncertainty of spatial averages and totals of natural resource maps. Methods Ecol. Evol. 14, 1320–1332. https://doi.org/10.1111/2041-210X.14106.

Wadoux, A.M.J.C., Heuvelink, G.B.M., Lark, R.M., Lagacherie, P., Bouma, J., Mulder, V. L., Libohova, Z., Yang, L., McBratney, A.B., 2021. Ten challenges for the future of pedometrics. Geoderma 401, 115155. https://doi.org/10.1016/J.GEODERMA.2021.115155.

Wadoux, A.-M.-C., Román Dobarco, M., Malone, B., Minasny, B., McBratney, A.B., Searle, R., 2023. Baseline high-resolution maps of organic carbon content in australian soils. Sci. Data 10. https://doi.org/10.1038/s41597-023-02056-8.

Wang, J., Filippi, P., Haan, S., Pozza, L., Whelan, B., Bishop, T.F., 2024. Gaussian process regression for three-dimensional soil mapping over multiple spatial supports. Geoderma 446, 116899. https://doi.org/10.1016/J.GEODERMA.2024.116899.

Webster, R., Oliver, M.A., 2007. Geostatistics for Environmental Scientists, 2nd ed. Wiley.

Yang, R.M., Huang, L.M., Zhang, X., Zhu, C.M., Xu, L., 2023. Mapping the distribution, trends, and drivers of soil organic carbon in China from 1982 to 2019. Geoderma 429, 116232. https://doi.org/10.1016/J.GEODERMA.2022.116232.

Zhang, Z., Ding, J., Zhu, C., Wang, J., Ge, X., Li, X., Han, L., Chen, X., Wang, J., 2023a. Historical and future variation of soil organic carbon in China. Geoderma 436, 116557. https://doi.org/10.1016/j.geoderma.2023.116557.

Zhang, Z., Ding, J., Zhu, C., Wang, J., Li, X., Ge, X., Han, L., Chen, X., Wang, J., 2023b. Exploring the inter-decadal variability of soil organic carbon in China. Catena 230, 107242. https://doi.org/10.1016/j.catena.2023.107242.