

# Estimating soil organic carbon stock change at multiple scales using machine learning and multivariate geostatistics



Gábor Szatmári <sup>a,\*</sup>, László Pásztor <sup>a</sup>, Gerard B.M. Heuvelink <sup>b,c</sup>

<sup>a</sup> Department of Soil Mapping and Environmental Informatics, Institute for Soil Sciences, Centre for Agricultural Research, Budapest, Hungary

<sup>b</sup> Soil Geography and Landscape Group, Wageningen University, Wageningen, The Netherlands

<sup>c</sup> ISRIC – World Soil Information, Wageningen, The Netherlands

## ARTICLE INFO

Handling Editor: Kristin Piikki

**Keywords:**

Soil organic carbon  
Spatio-temporal assessment  
Geostatistical simulation  
Uncertainty assessment  
Change of support  
Random forest

## ABSTRACT

Many national and international initiatives rely on spatially explicit information on soil organic carbon (SOC) stock change at multiple scales to support policies aiming at land degradation neutrality and climate change mitigation. In this study, we used regression cokriging with random forest and spatial stochastic cosimulation to predict the SOC stock change between two years (i.e. 1992 and 2010) in Hungary at multiple aggregation levels (i.e. point support, 1 × 1 km, 10 × 10 km square blocks, Hungarian counties and entire Hungary). We also quantified the uncertainty associated with these predictions in order to identify and delimit areas with statistically significant SOC stock change. Our study highlighted that prediction of spatial totals and averages with quantified uncertainty requires a geostatistical approach and cannot be solved by machine learning alone, because it does not account for spatial correlation in prediction errors. The total topsoil SOC stock for Hungary was predicted to increase between 1992 and 2010 with 14.9 Tg, with lower and upper limits of a 90% prediction interval equal to 11.2 Tg and 18.2 Tg, respectively. Results also showed that both the predictions and uncertainties of the average SOC stock change were smaller for larger spatial supports, while spatial aggregation also made it easier to obtain statistically significant SOC stock changes. The latter is important for carbon accounting studies that need to prove in Measurement, Reporting and Verification protocols that observed SOC stock changes are not only practically but also statistically significant.

## 1. Introduction

Soil is the largest terrestrial pool of organic carbon (Stockmann et al., 2013). Soil organic carbon (SOC) directly or indirectly influences various soil-related functions and services, such as food production and climate change mitigation. Therefore, there is an increasing demand for information about how SOC stocks vary not just in space but also in time. This increasing demand is explicitly demonstrated by a number of recent global and continental initiatives, e.g. SoilGrids (Hengl et al., 2017; Hengl et al., 2014; Poggio et al., 2021), GlobalSoilMap (Arrouays et al., 2014), 4 per mille Soils for Food Security and Climate (Minasny et al., 2017), Global Soil Organic Carbon Mapping Campaign (Yigini et al., 2018), Soil organic carbon projections for Europe (Yigini and Panagos, 2016) and Australian SoilGrids (Viscarra Rossel et al., 2015). Besides, a large effort has been made by many countries in order to compile their own national, high resolution SOC stock map (e.g. Adhikari et al., 2014; Kempen et al., 2019; Liang et al., 2019; Padarian et al., 2017; Szatmári

et al., 2019; Vitharana et al., 2019).

More recently, the assessment of SOC stock change is in the spotlight and a number of international initiatives, studies and researches rest on it, e.g. achieving land degradation neutrality (Keesstra et al., 2018; Stavi and Lal, 2015) and mitigating climate change (Lal, 2004; Minasny et al., 2017; Stockmann et al., 2013). If detailed, spatio-temporally referenced observations on SOC stock are available, then an explicit space-time model can be built for studying and assessing the spatio-temporal variation of SOC stocks (e.g. Heuvelink et al., 2020). However, if the observations on SOC stock are only detailed in space and not so much in time (e.g. SOC stock was only measured at two points in time), then building an explicit space-time model is not possible. In such a case, a different approach must be used. A frequently adopted approach is to model and predict separately the spatial distribution of SOC stocks for the different years, after which the SOC stock change is computed by subtraction (e.g. Schillaci et al., 2017; Szatmári et al., 2019; Zhou et al., 2019). However, such procedure does not account for the correlation

\* Corresponding author.

E-mail address: [szatmari@rissac.hu](mailto:szatmari@rissac.hu) (G. Szatmári).

between the SOC stock prediction errors for the two years and will therefore fail to characterise the uncertainty about the SOC stock change reliably. Therefore, it is better to jointly model the spatial distributions of SOC stock for different years. In geostatistics, the cokriging approach can be used to jointly model the spatial distributions of interdependent variables (McBratney and Webster, 1983; Wackernagel, 2003). Cokriging gives both the variances and covariance of the prediction errors for each prediction location, which allows to compute the prediction error variance associated with the SOC stock change for each location. Knowledge about the uncertainty of SOC stock change is important to identify and delimit areas with a statistically significant change in SOC stock.

It is common that data on SOC are frequently found to be positively skewed (e.g. Liang et al., 2019; Orton et al., 2014; Szatmári and Pásztor, 2019). From a geostatistical point of view, this could pose difficulties for spatial modelling and inference. For example, a property of a variable with a positively skewed distribution is the proportional effect, meaning that variability is higher in areas with high values than in areas with low values, which may corrupt the experimental variogram (Chilès and Delfiner, 2012; Webster and Oliver, 2007). Furthermore, most kriging techniques (e.g. ordinary, simple, universal and regression kriging) are only optimal if the variable of interest is normally distributed. In the case of a multivariate normal distribution, the conditional distribution is also normal and the conditional mean is a linear function of the conditioning data. Therefore, in case of a variable with a positively skewed distribution a commonly adopted approach is to: (i) transform the positively skewed variable to normal scale; (ii) perform geostatistical modelling and prediction of the transformed variable; and (iii) back-transform the kriging results to the original, untransformed scale (Webster and Oliver, 2007).

The scale at which spatially explicit information on SOC stock change is needed may vary widely from application to application. For example, in case of a country, state or smaller region (county, landscape, etc.), one may need SOC maps at a relatively fine spatial scale ( $\leq 1$  ha) (Adhikari et al., 2019; Ellili et al., 2019; Schillaci et al., 2017; Szatmári et al., 2019), whereas for a global or continental assessment it is sufficient to know the SOC stock change at a coarser scale ( $\geq 1 \text{ km}^2$ ) (Lugato et al., 2015; Yigini et al., 2018). An extreme example is carbon accounting, in which only the spatial total or average SOC stock change for a given time period is required for a country as a whole. In such cases stakeholders may be interested in information on SOC stock change referring to an area or region that is larger than the support of the observations, an issue which is commonly referred to as change-of-support. In change-of-support, we are interested in the spatial total or average of a soil property over a region or “block” that may be irregularly shaped (Webster and Oliver, 2007). In geostatistics, this is achieved with block kriging, which not only derives a prediction of the spatial average but also quantifies the associated prediction uncertainty (Cressie, 2006; Lark and Lapworth, 2012; Webster and Oliver, 2007). However, block kriging may be challenging in case observations are transformed prior to geostatistical modelling and predictions are back-transformed afterwards (Cressie, 2006; Cressie, 1993; Orton et al., 2015). In such case one often resorts to spatial stochastic simulation, where the spatial average over a region is approximated by generating many simulated region values as spatial averages of simulated point values (Goovaerts, 2001). This method can also be applied in a cokriging case and thus be used to compute the change in the average SOC stock for regions between two years.

The objective of this study was to use cokriging and spatial stochastic simulation to predict the SOC stock change between two years, i.e. 1992 and 2010, in Hungary at multiple aggregation levels. Simultaneously, we also aimed to quantify the uncertainties associated with these predicted SOC stock changes in order to be able to identify and delimit areas with statistically significant change.

## 2. Materials and methods

### 2.1. Soil data

We derived soil data from the Hungarian Soil Information and Monitoring System (SIMS), which is a countrywide soil monitoring system with the aim of providing geographically referenced biological, physical and chemical information on the status and temporal change of Hungarian soils. It stores measurements of 4,859 genetic soil horizons belonging to 1,236 soil profiles (Fig. 1). The density of soil profiles is 0.013 profiles per  $\text{km}^2$  and all soil profiles are benchmark locations, which were selected by expert knowledge-based purposive sampling in 1992. According to the sampling protocol of SIMS, all soil pits are revisited after a certain time period and sampled at each genetic soil horizon by taking point samples. In practice, a new soil pit is dug a few meters from where the previous pit was, thus introducing short-distance variation in measured soil properties. Note that these differences have no systematic component and will average out when spatial aggregates are derived. The laboratory measurements of the soil samples are carried out according to the Hungarian standards, where SOC is determined by wet combustion.

SOC stock can be readily computed by the following soil properties: (i) soil organic carbon content [unit: wt. %]; bulk density (BD) [unit: g  $\text{cm}^{-3}$ ]; and (iii) proportion of coarse fragment (CF) (i.e. proportion of stones and gravels) [unit: vol. %]. These soil data for the years 1992 and 2010 were derived from the genetic soil horizons of SIMS and used in this study. BD and CF were only measured in 1992, nonetheless, they can be considered static that have negligible change over time. Therefore, we used the BD and CF measurements of 1992 for both years.

### 2.2. Environmental covariates

We used the same set of environmental covariates for statistical modelling of SOC stocks as that used in Szatmári et al. (2019). The covariates are listed in Table 1. A detailed description on the selection of the covariates and their preprocessing can be found in Szatmári et al. (2019). We used this set of 26 environmental covariates for modelling the trend component of SOC stock for both years.

### 2.3. Supports for spatial aggregation

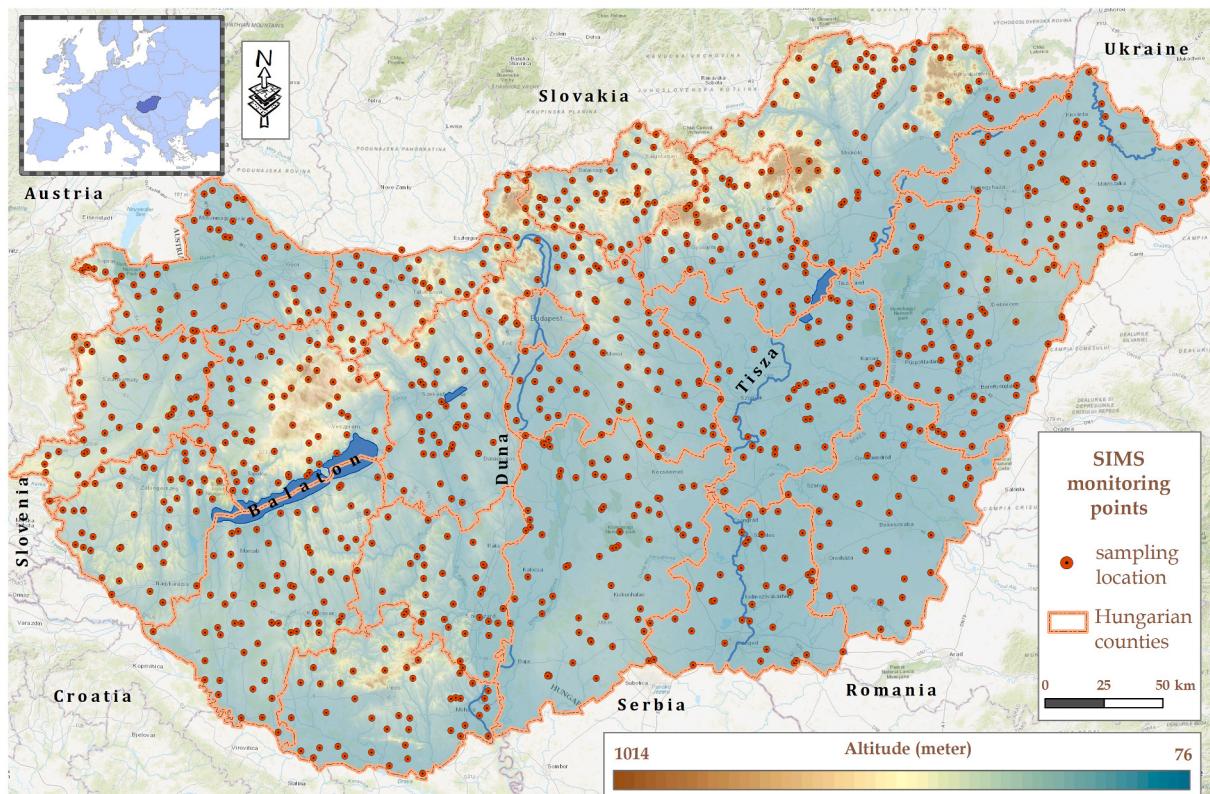
In this study we chose supports which are frequently used in Hungary and that are of interest to stakeholders. We used 1  $\text{km} \times 1 \text{ km}$  and 10  $\text{km} \times 10 \text{ km}$  blocks, which could be useful for countrywide crop simulation (Fodor et al., 2014) and terrestrial ecosystem process modelling (Hidy et al., 2016), in national ragweed (*Ambrosia artemisiifolia* L.) pollen forecasting (Csépe et al., 2020), in natural vegetation mapping and in national contributions to global/continental assessment of soil organic carbon stock (Yigini et al., 2018). In addition, we also aggregated to the Hungarian counties and entire Hungary (Fig. 1). The latter is important in carbon accounting, in which the spatial total SOC stock change has to be reported by a country.

### 2.4. Spatial modelling of SOC stocks and SOC stock change

Fig. 2 presents a flowchart summarizing the methods and approaches used in this study. The various steps shown in the flowchart are explained in the subsections below. We used a direct mapping approach (Miller et al., 2015), also referred to as “calculate-then-model” (Orton et al., 2014) or “calculate first, interpolate later” (Heuvelink and Pebesma, 1999) for modelling the spatial distribution of SOC stocks in 1992 and 2010. This means that SOC stocks are computed at the level of soil profiles first, and subsequently used for spatial modelling.

#### 2.4.1. Computing SOC stock at the level of soil profiles

First, we computed SOC stocks for each genetic soil horizon at each



**Fig. 1.** Monitoring sites (i.e. soil profiles) of the Hungarian Soil Information and Monitoring System (SIMS) on top of a digital elevation model of Hungary, and showing borders between the counties of Hungary.

soil profile for 1992 and 2010 using soil data listed in Section 2.1. We used the following equation

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

where  $SOC_{\text{stock}}$  [tons  $\text{ha}^{-1}$ ] is the soil organic carbon stock,  $SOC$  [wt. %] is the SOC content,  $BD$  [ $\text{g cm}^{-3}$ ] is the bulk density,  $CF$  [vol. %] is the proportion of coarse fragments,  $(\text{Top} - \text{Bottom})$  [cm] is the thickness of the given soil horizon, and 100 is used for unit conversion. After computing SOC stock for each soil horizon, we used a mass preserving spline (MPS) technique (Bishop et al., 1999) for modelling the vertical distribution of SOC stock at each soil profile. The fitted MPSs were used to derive the SOC stock for the depth interval 0–30 cm at each soil profile.

#### 2.4.2. Spatial modelling

As indicated in the Introduction, SOC stock data frequently show positively skewed distribution and the same was found in Hungary by Szatmári and Pásztor (2019). Therefore, we transformed the SOC stock values by taking their square root and spatial modelling was conducted on the transformed, normal scale.

In general, the spatial variation of soil properties can be described and modelled in terms of a deterministic component and a stochastic component. Thus, we can write

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

$$\begin{bmatrix} Z_{1992}(\mathbf{u}) \\ Z_{2010}(\mathbf{u}) \end{bmatrix} = \begin{bmatrix} m_{1992}(\mathbf{u}) \\ m_{2010}(\mathbf{u}) \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{1992}(\mathbf{u}) \\ \boldsymbol{\varepsilon}_{2010}(\mathbf{u}) \end{bmatrix}$$

where  $\mathbf{Z}$  is a vector of transformed SOC stocks for the years 1992 and 2010,  $\mathbf{m}$  is a vector of the deterministic part describing SOC stock spatial variation that can be explained from covariates for the years 1992 and 2010,  $\boldsymbol{\varepsilon}$  is a vector of the stochastic residuals that can be spatially auto-

correlated as well as cross-correlated for the years 1992 and 2010, and  $\mathbf{u}$  is a vector of geographical coordinates.

In both years, we separately modelled the deterministic part using random forest (RF; Breiman, 2001). Before fitting RF between the transformed SOC stock values and the environmental covariates listed in Table 1, we fine-tuned the hyperparameter  $mtry$ , which is the number of input covariates selected randomly at each split. Repeated 5-fold cross-validation was used for evaluating the performance of RF fitting in both years. We used the fine-tuned values of  $mtry$  for fitting the final RFs, which were then used to predict the deterministic component for all locations of a 100 m × 100 m grid covering Hungary in both years.

Since we wanted to account for correlation between the stochastic residuals for both years, we used regression cokriging (coRK), where RF predictions for the years 1992 and 2010 were taken to be  $m_{1992}(\mathbf{u})$  and  $m_{2010}(\mathbf{u})$  of Eq. 2, respectively. Geostatistical modelling of  $\boldsymbol{\varepsilon}$  of Eq. 2 was obtained by first subtracting the RF predictions from the computed SOC stock values for all sampling locations for both years. Next we computed experimental variograms and a cross-variogram from these residuals and fitted a linear model of coregionalization (LMC). We used the LMC to ensure a statistically valid model (Goovaerts, 1997).

As we already pointed out in the Introduction, there is no analytical geostatistical solution to predict a spatial average over a region or block in case a non-linear transformation is applied prior to geostatistical modelling. We therefore used sequential Gaussian cosimulation to generate possible realizations of the geostatistical coRK model. We generated 500 stochastic realizations for both years. These realizations were conditional to the SOC stock observations from both years and reproduced the joint spatial variability.

We added the RF predictions to the stochastic realizations before we transformed the results back to the original, positively skewed scale, by taking the square. The corresponding realizations of SOC stock change were determined by taking the difference of the paired back-transformed realizations of the two years (i.e. subtracting the SOC

**Table 1**  
Summary of environmental covariates.

scorpan factors	Name	Resolution	Unit	Type
Soil Climate	Soil type map of Hungary	100 m		Categorical
	Long-term mean annual evaporation	100 m	[mm]	Continuous
	Long-term mean annual evapotranspiration	100 m	[mm]	Continuous
	Long-term mean annual precipitation	100 m	[mm]	Continuous
	Long-term mean annual temperature	100 m	[°C]	Continuous
Organisms	CORINE Land Cover from 1990	100 m		Categorical
	CORINE Land Cover from 2012	100 m		Categorical
Relief (or topography)	Digital elevation model (DEM)	100 m	[m]	Continuous
	Channel network base level	100 m	[m]	Continuous
	Diurnal anisotropic heating	100 m	[·]	Continuous
	Horizontal distance to channel network	100 m	[m]	Continuous
	LS factor	100 m	[·]	Continuous
	Mass balance index	100 m	[·]	Continuous
	Multiresolution ridge top flatness index	100 m	[·]	Continuous
	Multiresolution valley bottom flatness index	100 m	[·]	Continuous
	Multi-scale topographic position index	100 m	[m]	Continuous
	Profile curvature	100 m	[·]	Continuous
	SAGA wetness index	100 m	[·]	Continuous
	Slope	100 m	[°]	Continuous
	Stream power index	100 m	[·]	Continuous
	Surface area	100 m	[m <sup>2</sup> ]	Continuous
	Terrain ruggedness index	100 m	[·]	Continuous
	Topographic position index	100 m	[m]	Continuous
	Topographic wetness index	100 m	[·]	Continuous
Parent material	Total curvature	100 m	[·]	Continuous
	Vertical distance to channel network	100 m	[m]	Continuous
Geological map of Hungary	1:100,000			Categorical

stock simulated values of 1992 from the simulated values of 2010).

### 2.5. Spatial aggregation

The back-transformed realizations at point support and 100 m × 100 m resolution were the basis for predicting the spatial average of SOC stocks at the different supports listed in Section 2.3 for both years. For predicting the spatial average SOC stock change, we used the realizations of SOC stock changes computed in Section 2.4. The spatial average over a block at a given support is obtained by first averaging the simulated values falling in that block for each realization. This yields 500 average values for that block, which can be interpreted as 500 realizations of the block average. Second, we take the average of the 500 block averages to obtain a prediction for the spatial average over that block. This was done for predicting (i) the spatial average of SOC stocks for both years and (ii) the spatial average of SOC stock change.

For identifying blocks with statistically significant SOC stock change we counted how many of the 500 simulated block values of SOC stock change had a positive or negative change for each block. If the number of positive (negative) changes in SOC stock was larger than 475 (i.e. 95% of the 500 simulated block values), then we concluded that the SOC stock had statistically significantly increased (decreased) for that block.

### 2.6. Quantification of uncertainty

It is routine in digital soil mapping to quantify the prediction uncertainty. Thus, we also computed the 90% prediction interval (PI) in this study. The 90% PI reports the range of values within which the true value is expected to occur 9 times out of 10. The width of the 90% PI is a useful measure of the prediction uncertainty. We used the back-transformed realizations for quantifying the prediction uncertainty, not just for the two years but also for the SOC stock change between the two years. At the point support, the 90% PI at a prediction location was identified by the 0.05 and 0.95 quantiles of the 500 realizations at that location. In the case of spatial aggregation, the 90% PI for a given block was identified by the 0.05 and 0.95 quantiles of the 500 block values of that block.

### 2.7. Validation

We had no SOC stock observations at larger support and we therefore restricted the statistical validation to the point support predictions. We used 10-fold cross-validation to evaluate the performance of point support predictions and uncertainty quantifications. Common validation measures, i.e. bias (or mean error, ME) and root mean square error (RMSE), were computed:

$$ME = \frac{1}{n} \sum_{i=1}^n (P_i - O_i) \quad (3)$$

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

where  $n$  is the number of observations;  $P_i$  and  $O_i$  are the predicted and observed SOC stock or SOC stock change for observation location  $i$ , respectively. We also computed Lin's concordance correlation coefficient (CCC; Lin, 1989):

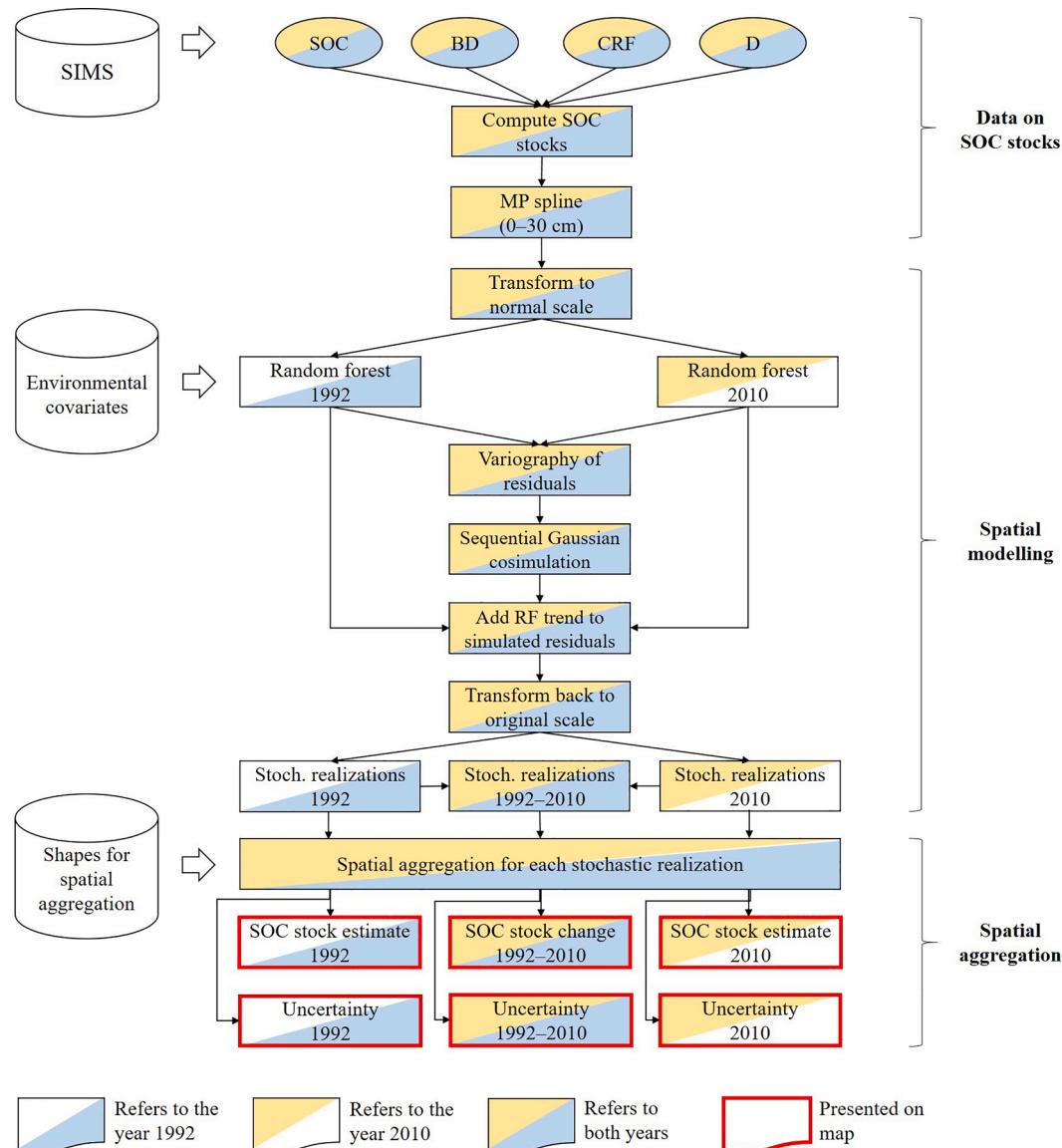
$$\rho_c = \frac{2 \sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sum_{i=1}^n (O_i - \bar{O})^2 + \sum_{i=1}^n (P_i - \bar{P})^2 + n(\bar{O} - \bar{P})^2} \quad (5)$$

where  $\bar{P}$  and  $\bar{O}$  are the mean of the predictions and observations.

Uncertainty quantifications were validated by accuracy plots and G statistics. An accuracy plot presents the observed fraction of validation data falling within symmetric prediction intervals of varying width. If the quantified uncertainty is a realistic assessment of the 'true' uncertainty, then the observed and expected fraction should be similar. For instance, in such case about 90% of the validation data would fall within the 90%. If the observed fraction is lower than the expected fraction, then the uncertainty has been underestimated. If it is higher, the uncertainty has been overestimated (i.e. too conservative). Ideally, an accuracy plot follows the 1:1 line. The G statistics measures the overall closeness of the accuracy plot to the 1:1 line by

$$G = 1 - \int_0^1 |\bar{\xi}(p) - p| dp \quad (6)$$

where  $\bar{\xi}(p)$  and  $p$  are the observed and the expected fraction for a  $p$ -width symmetric PI, respectively. Ideally, G is equal to 1. As it follows from Eq. 6, we did not distinguish between over- and underestimation of uncertainty by applying differing weights as recommended by Deutsch (1997). Indeed, we think that applying different weights on over- and underestimation could harm the objectivity of G statistics. Some studies also used non-weighted G statistics for validating uncertainty quantifications (e.g. Szatmári et al., 2020; Wadoux et al., 2018).



**Fig. 2.** Flowchart summarizing the methods and approaches used in this study. Abbreviations: SIMS: Hungarian Soil Information and Monitoring System, SOC: soil organic carbon, BD: bulk density, CRF: proportion of coarse fraction, D: depth, MP spline: mass-preserving spline, and RF: random forest.

### 3. Results

#### 3.1. Spatial modelling and spatial prediction at point support

Table 2 summarizes descriptive statistics of the computed SOC stocks

**Table 2**

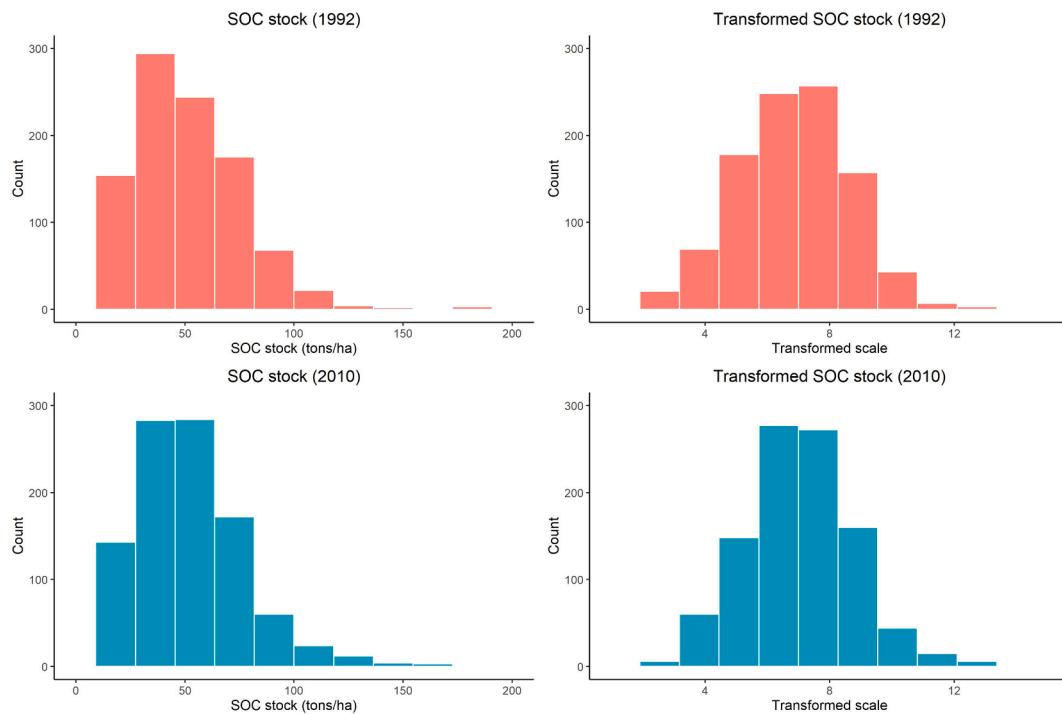
Summary statistics of soil organic carbon (SOC), bulk density (BD), proportion of coarse fragments (CF) and the computed soil organic carbon stock in the topsoil (0–30 cm) for the years 1992 and 2010.

Year	Soil properties	Unit	Min	Max	Mean	Median	Std. dev.
1992	BD	[g cm <sup>-3</sup> ]	0.70	1.91	1.43	1.44	0.14
	CF	[vol. %]	0.0	99.0	2.5	0.0	11.9
	SOC	[wt. %]	0.04	8.28	1.27	1.15	0.80
	SOC stock	[tons ha <sup>-1</sup> ]	0.0	218.6	48.7	46.1	27.3
2010	SOC	[wt. %]	0.16	4.12	1.30	1.18	0.66
	SOC stock	[tons ha <sup>-1</sup> ]	7.4	164.0	52.3	49.1	24.8

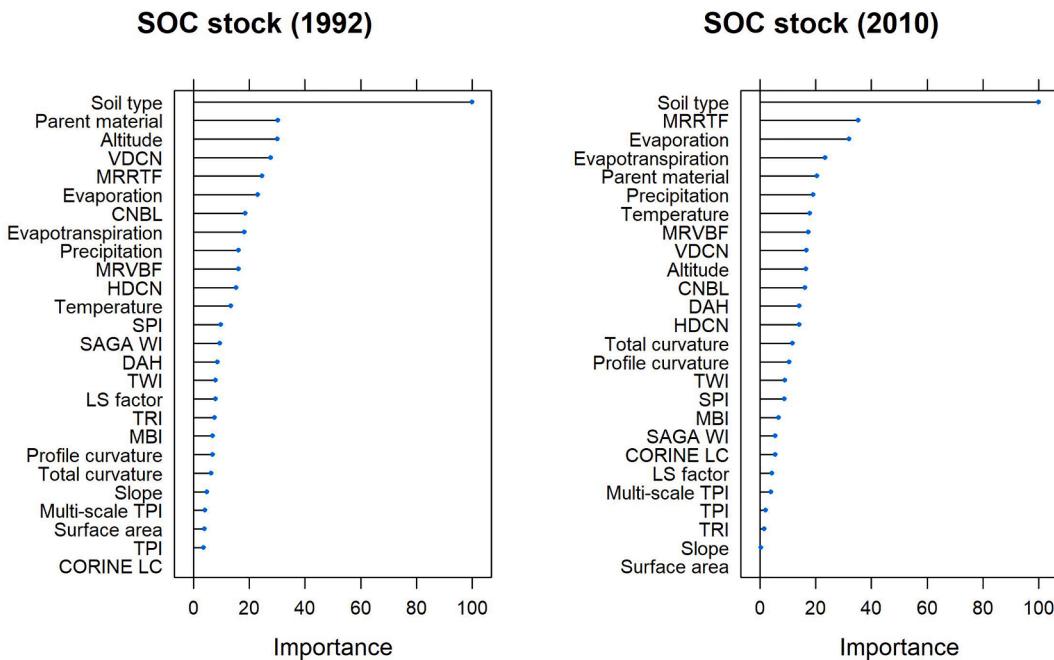
for the years 1992 and 2010, whereas Fig. 3 shows the histograms of the original (i.e. untransformed) and transformed data on SOC stock for both years. The transformed data on SOC stock showed a fairly normal distribution in both cases (Fig. 3, right column).

Fine-tuning of RF fittings showed that *mtry* values of 9 and 15 were found to be optimal for the years 1992 and 2010, respectively. The fitted RF models explained 41% and 38% of the total variance of SOC stocks in 1992 and 2010, respectively. The importance of the environmental covariates in both RF models are presented in Fig. 4. In this study, soil type proved to be the most important covariate for both years, which is not surprising because the genetic type of soils (according to the Hungarian soil classification system) depends on the amount and quality of soil organic matter, among others. We also found that parent material, geomorphometric parameters (e.g. altitude above sea level, ridge top flatness and vertical distance to channel network) and climate attributes (e.g. evaporation and evapotranspiration) were informative covariates in both RF models.

Fig. 5 presents the experimental direct- and cross-variograms of the residual and the fitted LMC. The model has a nested structure, where the first structure models discontinuity at the origin (i.e. nugget effect), the



**Fig. 3.** Histograms of the original (left column) and transformed (right column) data on soil organic carbon (SOC) stock for the years 1992 and 2010.



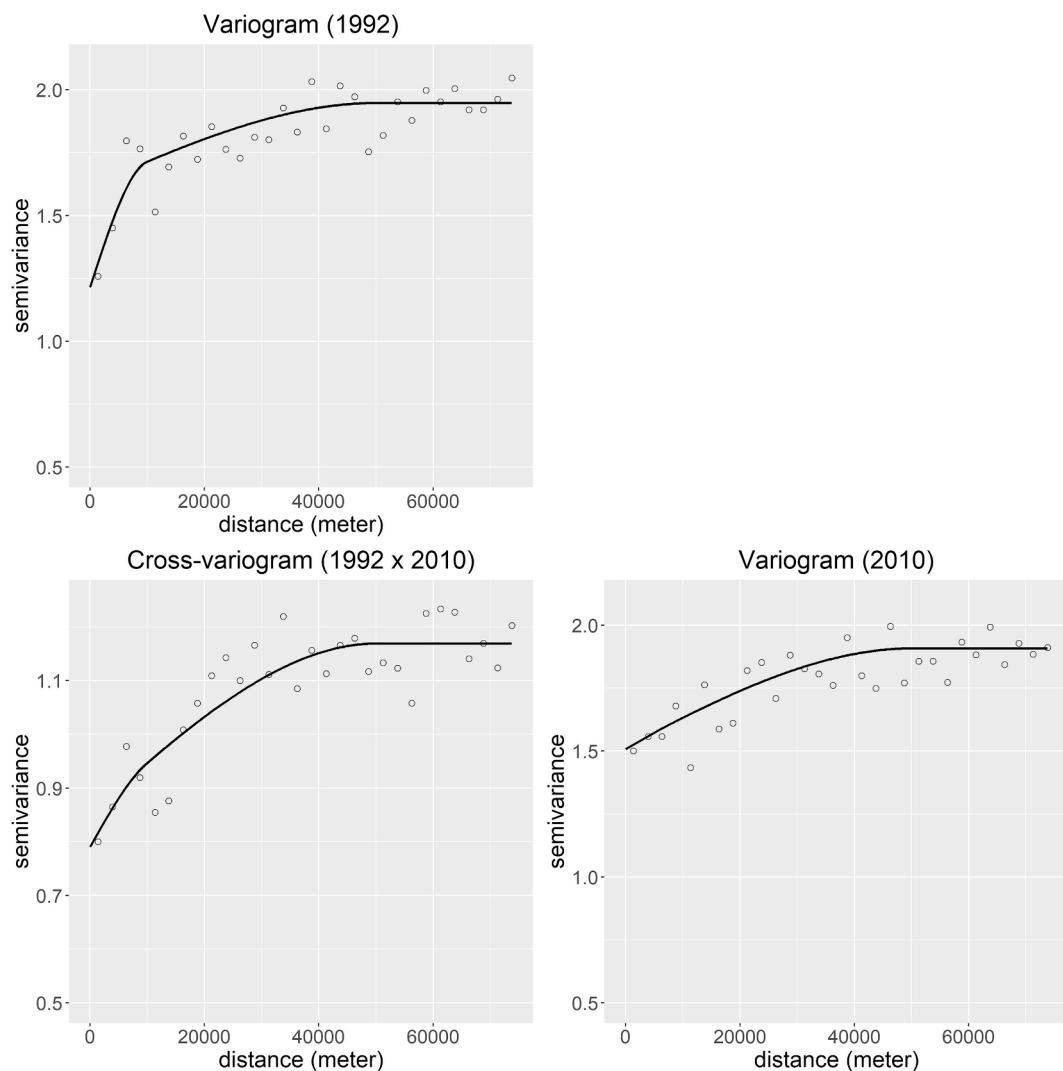
**Fig. 4.** Importance of environmental covariates in the fitted random forest models for the years 1992 and 2010. Abbreviations: CNBL: channel network base level, DAH: diurnal anisotropic heating, HDCN: horizontal distance to channel network, CORINE LC: CORINE land cover, MBI: mass balance index, MRRTF: multiresolution index of the ridge top flatness, MRVBF: multiresolution index of valley bottom flatness, SOC: soil organic carbon, SPI: stream power index, TPI: topographic position index, TRI: terrain ruggedness index, TWI: topographic wetness index, VDCN: vertical distance to channel network, and WI: wetness index.

second models a short scale variability with a range value of 10 km, and the third describes a larger scale of spatial continuity with a range value of 50 km. Table 3 summarizes the shapes and parameters of the LMC.

Fig. 6 presents the point-support predictions of SOC stocks and SOC stock change between the two years. It also shows the quantified prediction uncertainty (i.e. the lower and the upper limit of the 90% PI) at point support.

### 3.2. Performance of spatial predictions and uncertainty quantifications at point support

Table 4 summarizes the performance of the point-support predictions of SOC stocks and SOC stock change by the computed values of ME, RMSE and CCC. The ME results show that the predicted SOC stocks and SOC stock change are fairly unbiased. Fig. 7 presents the validation



**Fig. 5.** Experimental direct- and cross-variograms (open circles) and fitted linear model of coregionalization (solid lines).

**Table 3**  
Shapes and parameters of the fitted linear model of coregionalization.

	Model type	Partial sill	Range [m]
SOC stock 1992	Nugget	1.215	0
	Spherical	0.400	10,000
	Spherical	0.331	50,000
SOC stock 2010	Nugget	1.507	0
	Spherical	0.009	10,000
	Spherical	0.390	50,000
SOC stock 1992–2010	Nugget	0.790	0
	Spherical	0.062	10,000
	Spherical	0.316	50,000

of the uncertainty quantifications by the compiled accuracy plots and the G statistics. The accuracy plots approximately follow the 1:1 line, proving that the accuracy of uncertainty quantifications is acceptable. This is also supported by the G statistics, which are fairly close to the expected value. Based on the results of cross-validation it seems that the performance of spatial predictions and uncertainty quantifications was realistically assessed.

### 3.3. Spatial aggregation

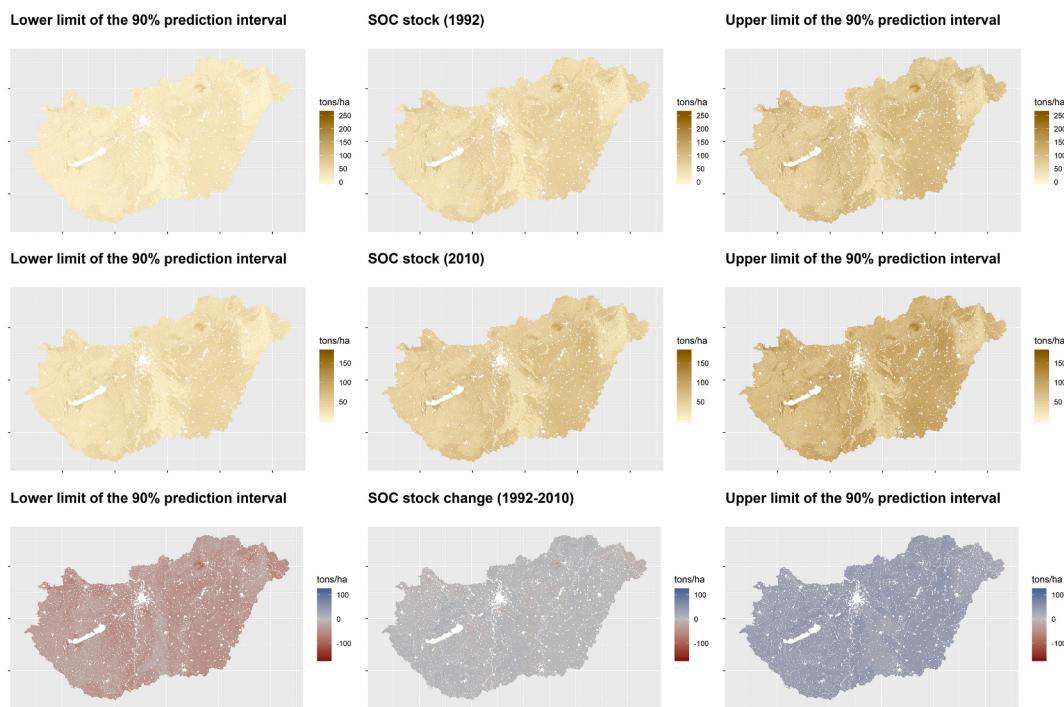
The resulting maps of spatial aggregation of SOC stocks for the two

years can be found in the [supplementary material \(SM\)](#). [Fig. 8](#) presents the results of the spatial aggregation of SOC stock change for all supports considered. In addition, it also shows the associated prediction uncertainty expressed by the width of the 90% PI and presents the map of statistically significant SOC stock change. Note that the area of significant change is larger for larger spatial supports. [Table 5](#) summarizes descriptive statistics of the point-support predictions and the spatially aggregated SOC stocks and SOC stock change at all supports used in this study. Note that [Table 5](#) does not include the results on spatial aggregation of the entire country because there is just a single value for each case, as shown in [Fig. 8](#).

## 4. Discussion

### 4.1. Interpretation and discussion of the results

In [Fig. 8](#) we observed that the larger the support, the closer to zero the spatially aggregated predictions of SOC stock change. This is also confirmed by [Table 5](#) in which the range values decrease as the support size increases. This can be explained by the fact that positive and negative values of SOC stock change neutralize each other in the course of spatial aggregation. This could be important from a practical point of view because it means that the absolute value of SOC stock change (i.e. regardless whether it is a positive or negative change) decreases as the size of area for which spatial prediction is required increases. This is



**Fig. 6.** Point-support predictions of soil organic carbon (SOC) stocks and SOC stock change for Hungary between the years 1992 and 2010 and their prediction uncertainty.

**Table 4**

10-fold cross-validation of point support predictions. Abbreviations: ME: mean error, RMSE: root mean square error, and CCC: Lin's concordance correlation coefficient.

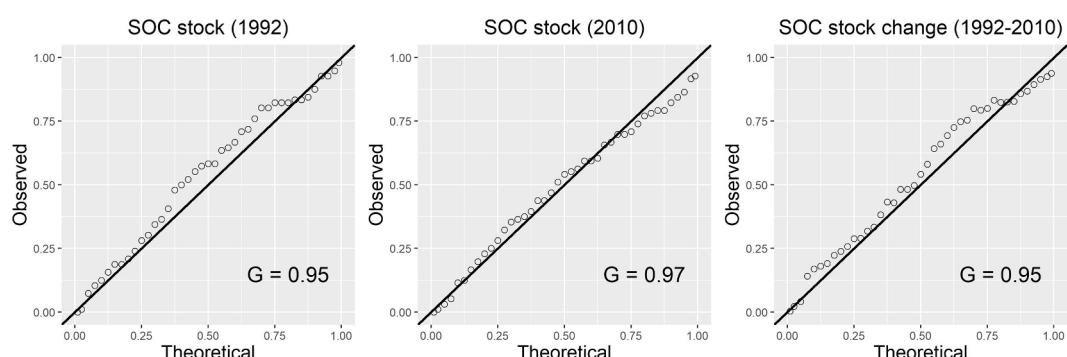
	ME [tons ha <sup>-1</sup> ]	RMSE [tons ha <sup>-1</sup> ]	CCC [-]
SOC stock 1992	0.24	17.59	0.60
SOC stock 2010	-3.28	20.13	0.45
SOC stock change	-3.57	18.11	0.44

nicely demonstrated especially in the case of the two largest aggregation levels (i.e. entire Hungary and the Hungarian counties) in Fig. 8, where the absolute value of SOC stock change is only a few tons ha<sup>-1</sup>.

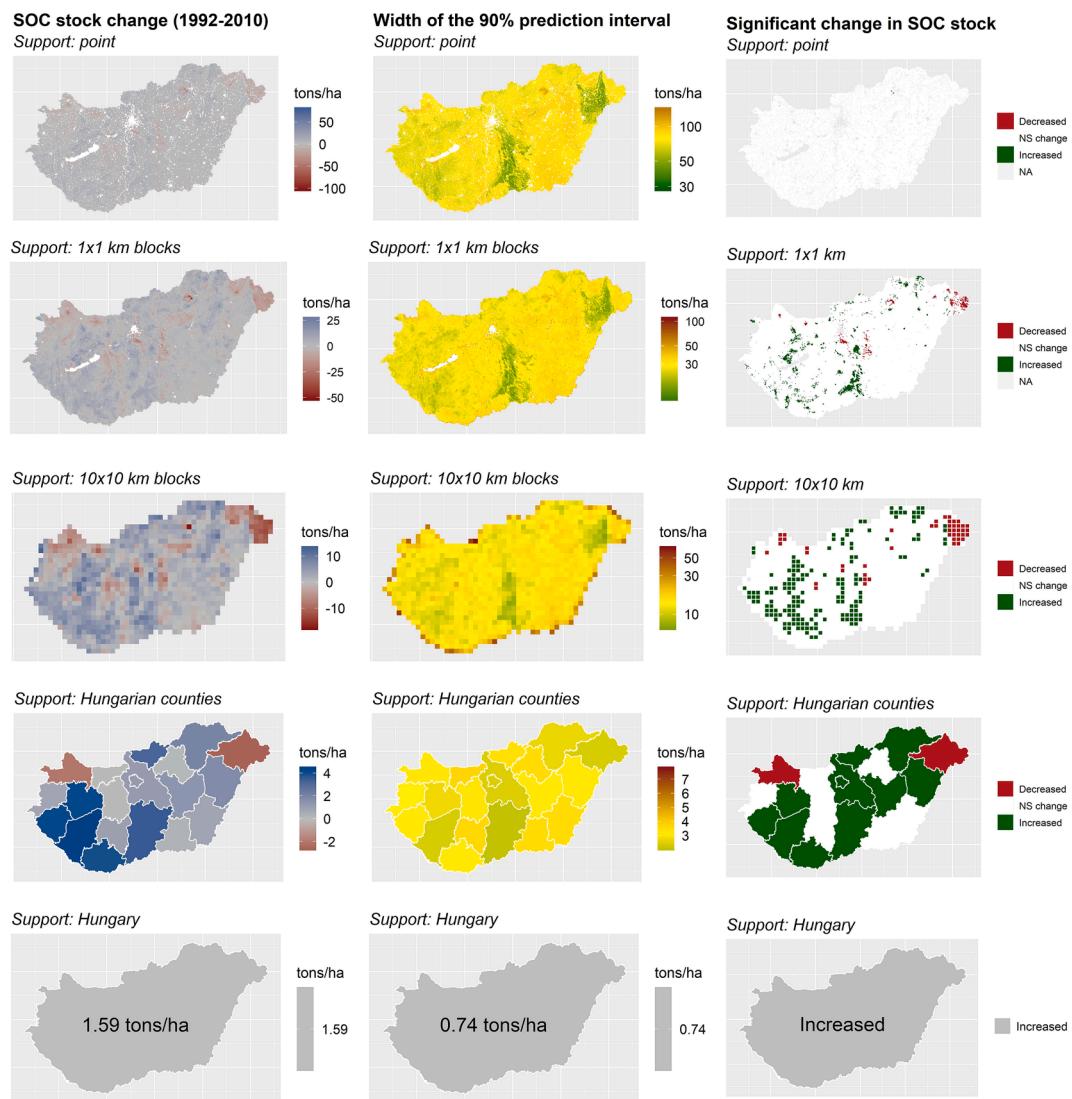
Another important observation from Fig. 8 and SM is that the larger the support, the smaller the prediction uncertainty, both for SOC stocks and SOC stock change. The reason for this is that within-block negative and positive interpolation errors partially cancel out when values are aggregated, a phenomenon well-known in geostatistics and also reflected in the fact that block kriging variances are typically much smaller than point kriging variances (Webster and Oliver, 2007). Furthermore, as was shown in Fig. 8, parallel to increasing the support size the area

that has a statistically significant change in SOC stock increases. Thus, decreasing prediction uncertainty by spatial aggregation made it possible to identify and delimit larger areas with statistically significant decrease or increase in SOC stock. This is an important message for users of SOC stock change maps, because we can rule out chance effects only for areas that have a statistically significant change. A broader discussion on this specific topic is given in Section 4.2.

Based on the spatially aggregated prediction of SOC stock change at the highest aggregation level (i.e. entire Hungary) (Fig. 8), the overall amount of SOC stock increase between 1992 and 2010 was 14.9 Tg C (1 teragram = 10<sup>12</sup> g) for entire Hungary with a 90% PI of [11.2 Tg C, 18.2 Tg C]. This means that there was a substantial enrichment in SOC stock in Hungary for the respective period, which was statistically significant, since the lower limit of the 90% PI (i.e. 11.2 Tg C) is much higher than zero. This can be explained by the fact that Hungary experienced substantial afforestation in the past three decades, which significantly increased the total area of forests. This is in line with the findings of Szatmári et al. (2019). At the level of Hungarian counties, we could also show that there was a significant change in SOC stock in 12 out of the 19 counties in Hungary. Moreover, 10 out of the 12 counties showed significant enrichment. This could be mostly attributed to the increasing



**Fig. 7.** Validation of uncertainty quantifications by accuracy plots and G-statistics. Abbreviation: SOC: soil organic carbon.



**Fig. 8.** Spatial aggregation of soil organic carbon stock (SOC) change between 1992 and 2010. Spatial prediction of SOC stock change (left column), the associated prediction uncertainty (middle column) and the map of significant change in SOC stock (right column). Abbreviation: NS change: no significant change.

**Table 5**

Summary statistics of spatial predictions at all supports except the entire country. Abbreviation: CV: coefficient of variation.

	Support	Min	Max	Range	Mean	Median	Variance	CV	n
SOC stock (1992)	Point	4.3	194.9	190.6	51.3	53.0	229.5	0.295	8,569,029
	1x1 km	10.5	146.4	135.9	51.4	53.0	194.2	0.271	93,068
	10x10 km	18.5	100.5	82.0	51.6	52.7	146.3	0.235	1,030
	Counties	40.5	63.5	23.0	51.1	49.5	46.2	0.133	20
SOC stock (2010)	Point	10.4	138.5	128.1	52.7	54.1	185.4	0.258	8,673,817
	1x1 km	18.3	101.7	83.4	52.8	53.6	152.7	0.234	93,013
	10x10 km	22.7	83.9	61.2	53.0	53.2	117.2	0.204	1,030
	Counties	40.8	64.4	23.6	52.3	51.9	42.2	0.124	20
SOC stock change (1992–2010)	Point	-101.7	79.0	180.7	1.6	1.7	46.4	4.292	8,475,606
	1x1 km	-51.0	26.5	77.5	1.5	1.7	25.3	3.361	92,920
	10x10 km	-17.6	13.1	30.7	1.5	1.9	16.8	2.817	1,030
	Counties	-2.6	4.4	7.0	1.4	1.1	4.3	1.521	20

area of forests, since most efforts made on afforestation took place in these counties. The remaining two counties showed significant loss in SOC stock, which may be explained by cultivation of pastures (county in the northwest) and degradation of woodlands (county in the northeast).

#### 4.2. General discussion

A commonly adopted approach in DSM is to add kriged residuals to the trend predictions obtained with multiple linear regression or machine learning, in order to improve the accuracy of the predictions (Keskin and Grunwald, 2018). However, we should emphasize that in this study we aimed for more than just improving the prediction

accuracy. Of foremost importance was that we wanted to quantify the prediction uncertainty of the spatial aggregates. Indeed, change-of-support problems with uncertainty quantification cannot be handled and solved without using geostatistics. Although several studies addressed the spatial distribution of observations in random forest fitting and prediction (Hengl et al., 2018; Møller et al., 2020; Sekulić et al., 2020), we should note that neither random forest nor other machine learning techniques are currently able to quantify spatial correlation. Therefore, there is a need to apply geostatistics or its combination with machine learning because geostatistics quantifies the spatial correlation (Matheron, 1963; Webster and Oliver, 2007), which is indispensable when the uncertainty of the spatially aggregated predictions is targeted.

In this study, we used the cokriging approach for jointly modelling the spatial distribution of SOC stocks for both years since it has the advantage of not just taking the spatial interdependency of SOC stocks between the two years into account but also exploiting this interdependency in spatial modelling (Webster and Oliver, 2007). In addition, cokriging yields coherent results with respect to the difference between SOC stock data observed in the two years (i.e. SOC stock change), which is very useful in further assessments. Due to these advantageous properties of the cokriging approach we advocate its use in cases when measurements are taken at only a limited number of times and therefore a full space-time model (Gräler et al., 2016) cannot be developed.

We used stochastic simulation instead of block kriging for predicting block means because there is no analytical geostatistical solution to predict a spatial average over a block in case a non-linear transformation is applied prior to geostatistical modelling. Furthermore, stochastic simulation is more flexible than block kriging because: (i) not just a linear average but also non-linear averages can be computed (e.g. geometric or harmonic mean, proportions above thresholds); and (ii) once the simulated point values have been generated, predictions for various block sizes and shapes can be derived at little computational cost (Goovaerts, 2001; Heuvelink and Pebesma, 1999).

A number of papers discussed that only a small area with a statistically significant change in SOC stock could be identified and delimited due to the relatively high prediction uncertainty (e.g. Heuvelink et al., 2020; Szatmári et al., 2019). Indeed, this is a key problem if, for example, a country wants to verify that its SOC conservation strategy carried out in a given period significantly increased SOC stock or not. Szatmári et al. (2019) and Heuvelink et al. (2020) extensively discussed the possibilities of reducing prediction uncertainty, e.g. collection of more observations and optimization of observation locations in the geographic and feature space, application of more and spatially (or spatiotemporally) detailed environmental covariates or improving the DSM technique. However, all these possibilities come at a cost and may have limited potential. This study highlighted that spatial aggregation substantially decreases prediction uncertainty and can effectively identify and delimit larger areas with significant decrease or increase in SOC stock. This is an important message for further studies on the spatio-temporal assessment of SOC stock change. The price paid is a loss of spatial detail, which means that spatial aggregation may not be a proper solution for local studies, precision agriculture and high-resolution in-situ management decision support. Moreover, it requires a geostatistical approach in which the spatial correlation of the prediction errors is quantified.

In case spatial averages or totals are required for blocks that cover many sampling locations, and when in addition these sampling locations were selected using probability sampling, an alternative to the model-based approach used in this study would be to apply design-based estimation (Brus, 2019; de Gruijter et al., 2016; de Gruijter et al., 2006). We could not use this approach in our study because the Hungarian SIMS is not a probability sample.

#### 4.3. Limitations and further research

In this study, we could only validate point predictions since we had no block-support SOC stock observations. In practice it may be difficult to collect block-support observations, although they may be estimated from averaging sufficient point samples within the block (Brus et al., 2011; Vaysse et al., 2017). Cross-validation statistics at point support are indicative of the quality of the block-support predictions, but not in all cases. For example, if the point-support predictions are unbiased then so will the block-support predictions. The block-support RMSE is likely substantially smaller than the point-support RMSE, but it is difficult to tell how much. Therefore, more effort should be made to collect validation data at block support, such as by using a probability sample of point observations within the block (Brus and de Gruijter, 1997; Brus and de Gruijter, 1993). Clearly, these validation data are not error-free since they are estimates based on point observations, and so one must assure that the sample size is sufficiently large to ensure that the estimation error is small compared to the error in the DSM map.

When spatial stochastic simulation is used, attention should be paid to how many realizations are needed to give reliable statistical inferences. In this study, we generated 500 realizations of SOC stocks and SOC stock change, which is likely sufficient to give reliable predictions both at point and at block supports. In the literature, some rule-of-thumb values have been published, e.g. it has been claimed that at least 100 realizations are needed to be able to give statistically sound inferences (Deutsch, 1997; Goovaerts, 2001). However, we should note that the number of realizations needed depends on what parameter is predicted. For instance, estimation of the mean of a distribution needs fewer realizations than estimation of the tails of a distribution. Therefore, we do not recommend the use of such rule-of-thumb values. Indeed, the literature of uncertainty propagation discuss this issue in more detail and provides approaches on how to determine the number of realizations needed to achieve stable results (Heuvelink, 1998).

We should note that we assumed in this study that there was no uncertainty about the trend (i.e. RF predictions). Thus, we only considered uncertainty about the stochastic residual in Eq. 2. Although cross-validation showed that uncertainty quantifications were appropriate (Fig. 7), it is advised to also take the uncertainty about the trend into account in spatial modelling in order to get a full picture about the overall uncertainty.

We used a direct mapping approach for modelling the spatial distribution of SOC stocks in 1992 and 2010, and then derived the SOC stock change between the two years since the cokriging approach yields coherent results as we pointed out in Section 4.2. However, there are further alternative trajectories or approaches (Heuvelink and Pebesma, 1999; Styc and Lagacherie, 2019) that may be interesting to check and compare with the approach used in this study, since Styc and Lagacherie (2019) observed important differences of performances between such inference trajectories.

#### 5. Conclusions

The objective of this study was to predict the SOC stock change between 1992 and 2010 in Hungary at multiple aggregation levels, and to quantify the uncertainty associated with these predicted SOC stock changes in order to be able to identify and delimit areas with statistically significant change. Our study highlighted that the sole application of machine learning techniques in spatial aggregation is not sufficient for this goal since these do not quantify the spatial correlation, while this is needed to quantify the uncertainty of spatially aggregated predictions. We also illustrated that by increasing the support the prediction uncertainty decreases, which is a well-known but frequently forgotten fact in geostatistics. Increasing the support thus allows to identify and delimit larger areas with statistically significant change in SOC stock. This is an important message for practitioners who not only need to estimate the SOC stock change resulting from a SOC conservation

strategy but also need to show that the effect of measures to increase the SOC storage is statistically significant.

## 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 K-131820. The work of Gábor Szatmári was supported by the Premium Postdoctoral Scholarship of the Hungarian Academy of Sciences (PREMIUM-2019-390). 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.2021.115356>.

## References

- Adhikari, K., Hartemink, A.E., Minasny, B., Bou Kheir, R., Greve, M.B., Greve, M.H., Hui, D., 2014. Digital mapping of soil organic carbon contents and stocks in Denmark. *PLoS One* 9 (8), e105519. <https://doi.org/10.1371/journal.pone.0105519>.
- Adhikari, K., Owens, P.R., Libohova, Z., Miller, D.M., Wills, S.A., Nemecek, J., 2019. Assessing soil organic carbon stock of Wisconsin, USA and its fate under future land use and climate change. *Sci. Total Environ.* 667, 833–845. <https://doi.org/10.1016/j.scitotenv.2019.02.420>.
- 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://dx.doi.org/10.1016/B978-0-12-800137-0.00003-0>.
- Bishop, T.F.A., McBratney, A.B., Laslett, G.M., 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma* 91 (1–2), 27–45. [https://doi.org/10.1016/S0016-7061\(99\)00003-8](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., 2019. Sampling for digital soil mapping: A tutorial supported by R scripts. *Geoderma* 338, 464–480. <https://doi.org/10.1016/j.geoderma.2018.07.036>.
- Brus, D.J., de Grujter, J.J., 1997. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with Discussion). *Geoderma* 80 (1–2), 1–44. [https://doi.org/10.1016/S0016-7061\(97\)00072-4](https://doi.org/10.1016/S0016-7061(97)00072-4).
- Brus, Dick J., DeGruijter, Jaap J., 1993. Design-based versus model-based estimates of spatial means: Theory and application in environmental soil science. *Environmetrics* 4 (2), 123–152. [https://doi.org/10.1002/\(ISSN\)1099-095X10.1002/env.v4:210.1002/env.3170040202](https://doi.org/10.1002/(ISSN)1099-095X10.1002/env.v4:210.1002/env.3170040202).
- 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>.
- Chilès, J.-P., Delfiner, P., 2012. Geostatistics: Modeling Spatial Uncertainty: Second Edition, 2nd ed, Geostatistics: Modeling Spatial Uncertainty: Second Edition. Wiley Blackwell. <https://dx.doi.org/10.1002/9781118136188>.
- Cressie, Noel, 2006. Block kriging for lognormal spatial processes. *Math. Geol.* 38 (4), 413–443. <https://doi.org/10.1007/s11004-005-9022-8>.
- Cressie, N.A.C., 1993. Statistics for Spatial Data. Wiley.
- Csépe, Z., Lelössy, Mányoki, G., Kajtor-Apatini, D., Udvárdy, O., Péter, B., Páldy, A., Gelybó, G., Szigeti, T., Pándics, T., Kofol-Seliger, A., Simčić, A., Leru, P.M., Eftimie, A.M., Sikoparija, B., Radisíć, P., Stjepanović, B., Hrga, I., Većenaj, A., Vucić, A., Peroš-Pucar, D., Škorić, T., Šćevková, J., Bastl, M., Berger, U., Magyar, D. 2020. The application of a neural network-based ragweed pollen forecast by the Ragweed Pollen Alarm System in the Pannonian biogeographical region. *Aerobiologia* (Bologna), 36, 131–140. <https://doi.org/10.1007/s10453-019-09615-w>.
- de Grujter, J.J., Bierkens, M.F.P., Brus, D.J., Knotters, M., 2006. Sampling for Natural Resource Monitoring. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://dx.doi.org/10.1007/3-540-33161-1>.
- de Grujter, 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>.
- Deutsch, C.V., 1997. Direct assessment of local accuracy and precision. In: Baafi, E.Y., Schofield, N.A. (Eds.), *Geostatistics Wollongong '96*. Kluwer Academic Publishers, pp. 115–125.
- Ellili, Y., Walter, C., Michot, D., Pichelin, P., Lemercier, B., 2019. Mapping soil organic carbon stock change by soil monitoring and digital soil mapping at the landscape scale. *Geoderma* 351, 1–8. <https://doi.org/10.1016/j.geoderma.2019.03.005>.
- Fodor, Nándor, Pásztor, László, Németh, Tamás, 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 (5), 391–410. <https://doi.org/10.1080/17538947.2012.689998>.
- Goovaerts, P., 2001. Geostatistical modelling of uncertainty in soil science. *Geoderma* 103 (1–2), 3–26. [https://doi.org/10.1016/S0016-7061\(01\)00067-2](https://doi.org/10.1016/S0016-7061(01)00067-2).
- Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press.
- Gräler, B., Pebesma, E., Heuvelink, G. 2016. Spatio-Temporal Interpolation using gstat. *R J.* 8, 204–218.
- Hengl, T., De Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Gonzalez, M.R., 2014. SoilGrids1km - Global soil information based on automated mapping. *PLoS One* 9. <https://doi.org/10.1371/journal.pone.0105992>.
- Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One* 12, e0169748. <https://dx.doi.org/10.1371/journal.pone.0169748>.
- 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://dx.doi.org/10.7717/peerj.5518>.
- Heuvelink, G.B.M. 1998. Error propagation in environmental modelling with GIS, Error propagation in environmental modelling with GIS. Taylor and Francis. <https://dx.doi.org/10.4324/9780203016114>.
- 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://dx.doi.org/10.1111/ejss.12998>.
- Heuvelink, Gerard B.M., Pebesma, Edzer J., 1999. Spatial aggregation and soil process modelling. *Geoderma* 89 (1–2), 47–65. [https://doi.org/10.1016/S0016-7061\(98\)00077-9](https://doi.org/10.1016/S0016-7061(98)00077-9).
- Hidy, Dóra, Barcza, Zoltán, Marjanović, Hrvoje, Ostrogović Sever, Maša Zorana, Dobor, Laura, Gelybő, Györgyi, Fodor, Nándor, Pintér, Krisztina, Churkina, Galina, Running, Steven, Thornton, Peter, Bellocchi, Gianni, Haszpra, László, Horváth, Ferenc, Suyker, Andrew, Nagy, Zoltán, 2016. Terrestrial ecosystem process model Biome-BGCMSu v4.0: Summary of improvements and new modeling possibilities. *Geosci. Model Dev.* 9 (12), 4405–4437. <https://doi.org/10.5194/gmd-9-4405-201610.5194/gmd-9-4405-2016-supplement>.
- 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>.
- Kempen, B., Dalsgaard, S., Kaaya, A.K., Chamuya, N., Ruipérez-González, M., Pekkarinen, A., Walsh, M.G., 2019. Mapping topsoil organic carbon concentrations and stocks for Tanzania. *Geoderma* 337, 164–180. <https://doi.org/10.1016/j.geoderma.2018.09.011>.
- Keskin, H., Grunwald, S., 2018. Regression kriging as a workhorse in the digital soil mapper's toolbox. *Geoderma* 326, 22–41. <https://doi.org/10.1016/j.geoderma.2018.04.004>.
- Lal, R., 2004. Soil carbon sequestration to mitigate climate change. *Geoderma* 123 (1–2), 1–22. <https://doi.org/10.1016/j.geoderma.2004.01.032>.
- 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>.
- Liang, Z., Chen, S., Yang, Y., Zhao, R., Shi, Z., Viscarra Rossel, R.A., 2019. Baseline map of soil organic matter in China and its associated uncertainty. *Geoderma* 335, 47–56. <https://doi.org/10.1016/j.geoderma.2018.08.011>.
- Lin, Lawrence I-Kuei, 1989. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics* 45 (1), 255. <https://doi.org/10.2307/2532051>.
- Lugato, Emanuele, Bampa, Francesca, Panagos, Panos, Montanarella, Luca, Jones, Arwyn, 2015. Potential carbon sequestration of European arable soils estimated by modelling a comprehensive set of management practices. *Glob. Chang. Biol.* 20 (11), 3557–3567. <https://doi.org/10.1111/gcb.12551>.
- Matheron, G. 1963. Principles of geostatistics. *Econ. Geol.* 58. <https://dx.doi.org/10.2113/gsecongeo.58.8.1246>.
- McBratney, A.B., Webster, R., 1983. Optimal interpolation and isarithmic mapping of soil properties. V. Co-regionalization and multiple sampling strategy. *J. Soil Sci.* 34, 137–162. <https://doi.org/10.1111/j.1365-2389.1983.tb00820.x>.
- Miller, B.A., Koszinski, S., Wehrhan, M., Sommer, M., 2015. Comparison of spatial association approaches for landscape mapping of soil organic carbon stocks. *Soil* 1, 217–233. <https://doi.org/10.5194/soil-1-217-2015>.
- 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., Sulamaan, 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>.
- Möller, A.B., Beucher, A.M., Pouladi, N., Greve, M.H., 2020. Oblique geographic coordinates as covariates for digital soil mapping. *Soil* 6, 269–289. <https://doi.org/10.5194/soil-6-269-2020>.

- 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://dx.doi.org/10.1111/ejs.12297>.
- 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>.
- Padarian, J., Minasny, B., McBratney, A.B., 2017. Chile and the Chilean soil grid: A contribution to GlobalSoilMap. *Geoderma Reg.* 9, 17–28. <https://doi.org/10.1016/j.geodrs.2016.12.001>.
- 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>.
- Schillaci, C., Acutis, M., Lombardo, L., Lipani, A., Fantappiè, M., Märker, M., Saia, S., 2017. Spatio-temporal topsoil organic carbon mapping of a semi-arid Mediterranean region: The role of land use, soil texture, topographic indices and the influence of remote sensing data to modelling. *Sci. Total Environ.* 601–602, 821–832. <https://doi.org/10.1016/j.scitotenv.2017.05.239>.
- Sekulić, Aleksandar, Kilibarda, Milan, Heuvelink, Gerard B.M., Nikolić, Mladen, Bajat, Branislav, 2020. Random forest spatial interpolation. *Remote Sens.* 12 (10), 1687. <https://doi.org/10.3390/rs12101687>.
- 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>.
- Stockmann, Uta, Adams, Mark A., Crawford, John W., Field, Damien J., Henakaarchchi, Nilusha, Jenkins, Meaghan, Minasny, Budiman, McBratney, Alex B., Courcelles, Vivien de Remy de, Singh, Kanika, Wheeler, Ichsanji, Abbott, Lynette, Angers, Denis A., Baldock, Jeffrey, Bird, Michael, Brookes, Philip C., Chenu, Claire, Jastrow, Julie D., Lal, Rattan, Lehmann, Johannes, O'Donnell, Anthony G., Parton, William J., Whitehead, David, Zimmermann, Michael, 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>.
- Styc, Q., Lagacherie, P., 2019. What is the best inference trajectory for mapping soil functions: An example of mapping soil available water capacity over languedoc roussillon (France). *Soil Syst.* 3, 1–17. <https://doi.org/10.3390/soilsystems3020034>.
- Szatmári, G., Bakacs, Z., Laborczi, A., Petrik, O., Pataki, R., Tóth, T., Pásztor, L., 2020. Elaborating Hungarian segment of the Global Map of Salt-affected Soils (GSSmap): National contribution to an international initiative. *Remote Sens.* 12, 1–19. <https://doi.org/10.3390/rs12244073>.
- 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, Gábor, Pirkó, Béla, Koós, Sándor, Laborczi, Annamária, Bakacs, Zsófia, Szabó, József, Pásztor, László, 2019. Spatio-temporal assessment of topsoil organic carbon stock change in Hungary. *Soil Till. Res.* 195, 104410. <https://doi.org/10.1016/j.still.2019.104410>.
- 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://dx.doi.org/10.1111/sum.12350>.
- Viscarra Rossel, R.A., Chen, C., Grundy, M.J., Searle, R., Clifford, D., Campbell, P.H., 2015. The Australian three-dimensional soil grid: Australia's contribution to the GlobalSoilMap project. *Soil Res.* 53, 845–864. <https://doi.org/10.1071/SR14366>.
- Vitharana, U.W.A., Mishra, U., Mapa, R.B., 2019. National soil organic carbon estimates can improve global estimates. *Geoderma* 337, 55–64. <https://doi.org/10.1016/j.geoderma.2018.09.005>.
- Wackernagel, H. 2003. Multivariate Geostatistics, Multivariate Geostatistics. Springer Berlin Heidelberg. 10.1007/978-3-662-05294-5.
- Wadoux, A.M.-C., Brus, D.J., Heuvelink, G.B.M., 2018. Accounting for non-stationary variance in geostatistical mapping of soil properties. *Geoderma* 324, 138–147. <https://doi.org/10.1016/j.geoderma.2018.03.010>.
- Webster, Richard, Oliver, Margaret A. (Eds.), 2007. Statistics in PracticeGeostatistics for Environmental Scientists. John Wiley & Sons, Ltd, Chichester, UK.
- Yigini, Y., Olmedo, G.F., Reiter, S., Baritz, R., Viatkin, K., Vargas, R. 2018. Soil organic carbon mapping: cookbook. FAO.
- Yigini, Y., Panagos, P., 2016. Assessment of soil organic carbon stocks under future climate and land cover changes in Europe. *Sci. Total Environ.* 557–558, 838–850. <https://doi.org/10.1016/j.scitotenv.2016.03.085>.
- Zhou, Y., Hartemink, A.E., Shi, Z., Liang, Z., Lu, Y., 2019. Land use and climate change effects on soil organic carbon in North and Northeast China. *Sci. Total Environ.* 647, 1230–1238. <https://doi.org/10.1016/j.scitotenv.2018.08.016>.