



Countrywide mapping and assessment of organic carbon saturation in the topsoil using machine learning-based pedotransfer function with uncertainty propagation

Gábor Szatmári^a, László Pásztor^a, Annamária Laborczi^{a,*}, Gábor Illés^b, Zsófia Bakacsi^a, Dóra Zacháry^{c,d}, Tibor Filep^{c,d}, Zoltán Szalai^{c,d,e}, Gergely Jakab^{c,d,e}

^a Institute for Soil Sciences, Centre for Agricultural Research, Budapest, Hungary

^b Department of Ecology and Sylviculture, Forest Research Institute, University of Sopron, Sopron, Hungary

^c Geographical Institute, Research Centre for Astronomy and Earth Sciences, Budapest, Hungary

^d MTA Centre of Excellence, Hungary

^e Department of Environmental and Landscape Geography, Eötvös Loránd University, Budapest, Hungary

ARTICLE INFO

Keywords:

Soil organic carbon
Organic carbon saturation
Cubist
Pedotransfer function
Digital soil mapping
Uncertainty propagation

ABSTRACT

Stakeholders and policymakers have been becoming more and more interested not just in the potential organic carbon (SOC) saturation level of soils but also in spatially explicit information on the degree of SOC deficit, which can support future policy and sustainable management strategies, and carbon sequestration-associated spatial planning. Thus the objective of our study was to develop a cubist-based pedotransfer function (PTF) for predicting and mapping the saturated SOC content of the topsoils (0–30 cm) in Hungary and then compare the resulting map with the actual SOC map to determine and assess the degree of SOC deficit. It was assumed that topsoils covered by permanent forests can be practically considered as saturated in SOC. Using the monitoring points of the Hungarian Soil Information and Monitoring System located in forests as reference soil profiles, we developed a cubist-based PTF. The transparent model structure provided by cubist allowed to show that not just the physicochemical properties of soils (e.g., texture, and pH) but also environmental conditions, such as topography (e.g., slope, altitude, and topographical position) and climate (e.g., long-term mean annual temperature, and evaporation), characterizing landscape are important factors in predicting the level of SOC saturation. Our results also pointed out that there is SOC deficit on large part of the country (~80%) showing high spatial variability. It was also revealed that the most considerable potential for additional SOC sequestration can be found related to soils with medium to high actual SOC content.

1. Introduction

Soil organic carbon (SOC) is widely known as a key property regarding fertility, soil health, and various ecosystem services. The SOC content may rapidly change in time and space, thus it varies in a wide range in soils due to environmental and anthropogenic conditions. However, the degree of possibly stabilized SOC is limited, which is often called saturation. There is an increasing demand on the determination, assessment and mapping of the potential saturation level of various soils as enrichment in SOC improves soil quality, provides an effective and economical solution for greenhouse gas control of the atmosphere (Lal, 2004a, 2004b) and contributes to achieving land degradation neutrality

(Keesstra et al., 2018; Stavi and Lal, 2015). Beyond the complex molecular interactions and processes between the mineral phase and organic matter content of the soil, which are mostly unknown (Inagaki et al., 2020), the term SOC saturation is suitable and applicable also for policymakers and stakeholders (O'Rourke et al., 2015). It, therefore, provides useful indicator for sustainable management. Accordingly, an adequate and proper map based database of potential SOC saturation is a necessity to take the four per mille (Minasny et al., 2017) objectives into practice. However, the theoretical saturation value of a certain soil is still not easy to determine, no universal prediction method is accepted (Sanderman et al., 2010). The difficulty is that many synergic and independent properties - such as climatic, geological, topographic, and

Abbreviations: SOC, soil organic carbon; SOM, soil organic matter; PTF, pedotransfer function; ML, machine learning.

* Corresponding author.

E-mail address: laborczi@rissac.hu (A. Laborczi).

<https://doi.org/10.1016/j.catena.2023.107086>

Received 10 May 2022; Accepted 10 March 2023

Available online 29 March 2023

0341-8162/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

land use conditions of the field and the mineralogical and grain size distribution of the soil - affect SOC storage and saturation (Briedis et al., 2018; Cotrufo et al., 2019; Fekete et al., 2021).

The actual SOC content reflects the dynamic equilibrium between the organic matter input and mineralization by the microbiome, highlighting the role of environmental properties rather than organic matter composition (Schmidt et al., 2011). If the organic matter production differs by orders from mineralization, the SOC content changes rapidly. The inhibited mineralization by water saturation, for example, causes organic matter accumulation such as histic horizons even under deficient biomass production (Aide et al., 2020). As under the above circumstances, SOC cannot be saturated, thus saturation concept should be limited only to mineral soils. Nevertheless, many mineral soils are also oversaturated based on different saturation models (Wiesmeier et al., 2014).

The stabilization of organic matter is thus related to its protection against microbial availability. This protection is reported as binding directly to the mineral surface (Chenu et al., 2019; Kögel-knabner et al., 2008) or by the shelter effect of soil aggregates (Six et al., 2002). In early studies, the ratio of the fine fraction (particles < 20 or 50 µm) was considered as the primary driving force of potential SOC saturation (Hassink, 1997). Besides, measurements prove that there are differences among land uses and soil depth in the fine fraction related SOC content (Chen et al., 2019, 2018). On average, only half of the SOC was associated with the fine fraction (Wiesmeier et al., 2014), thus Barré et al. (2017) stated that the SOC holding capacity of the fine fraction is irrelevant for the SOC saturation value of the bulk soil. The increasing number of investigations provided a wider dataset for identifying various indicators. In many studies, the mineral composition of the clay fraction was found to be the most important property determining the SOC concentration and, therefore, the saturation level. Especially the role of the sesquioxides or the extractable amount of aluminum was determined important (Rasmussen et al., 2018). In other studies, the specific surface area, the amount of exchangeable cations or the base saturation determined the SOC capacity mostly (Juhos et al., 2021).

Nonetheless, these indicators do not take particulate organic matter (POM) into account. According to the traditional approach, POM is of raw plant residues without relevant mineral protection, thus does not contribute to the saturation. In sandy soils, however, POM is the main fraction of soil organic matter (SOM) (Hanegraaf et al., 2009) and may also persist in the long run (Witzgall et al., 2021). Moreover, the SOC stock increase due to forestation or tillage shift is mainly triggered by the increasing POM content (Cardinael et al., 2015; Chimento et al., 2016). Thus the theoretical concept of saturation can only hardly be interpreted directly into practice. However, some recent papers still keep the traditional approach considering only the mineral phase associated OM as sequestered and the indicator of saturation degree (Chen et al., 2018). For instance, Chen et al. (2019) estimated SOC storage potential for the bulk soil and limited the term “SOC saturation” only to the mineral phase associated OM. In the present study, SOC saturation is meant as the highest available SOC content of the bulk soil independently of pools (Barré et al., 2017; Chenu et al., 2019; Six et al., 2002).

Nowadays, Barré et al. (2017) recommended two approaches for SOC saturation estimation: (i) predicting storage kinetics and estimate the final capacity value for a given land use scenario, and (ii) reference establishment predicting the highest SOC concentration. Both approaches are applicable only by using empirical observations of SOC content and storage and model-driven approaches. Thus, the application of the highest measured SOC content as the possible saturation limit is still an axiom. Barré et al. (2017) suggested the highest 10% of SOC contents for a given agroclimatic region under forest to be considered as the potential saturation value. However, the pedological and topographical heterogeneity for such an agroclimatic zone triggers low spatial representability of the highest 10% (Suleymanov et al., 2021). On the other hand, the created SOC saturation models are suitable only for those conditions which were considered during the model

construction, thus the results cannot be extrapolated. Feng et al. (2013) suspected the linear regression SOC saturation models for large datasets being dubious due to the presence of unsaturated points and resulting in oversaturated predictions. Thus, both the measured data and model-driven approaches should be considered parallel to achieve the best SOC saturation prediction in a countrywide scale investigation.

Since the late 1980s, pedotransfer functions (PTFs) have been developed to estimate soil properties, functions and services from other more easily and/or cheaply measurable soil properties to satisfy specific or complex demands on information relating to soils (Bouma, 1989; Van Looy et al., 2017). In the last few years, a number of novel approaches, techniques and thoughts have been adapted, for example, the application of machine learning (ML) techniques in developing PTFs (Makó et al., 2017; Shiri et al., 2017; Xiangsheng et al., 2016). Data-driven predictive models given by ML algorithms has proved to be highly efficient, however, we should note that resulting data-driven models are hardly interpretable (i.e., not really transparent for users). A further novelty is to use not just soil properties as predictors in developing PTFs but also further environmental covariates (Gupta et al., 2021; Szabó et al., 2019) characterizing the soil forming factors. Besides, a number of papers also pointed out that if there is spatially exhaustive information on the predictors available, then PTFs can also be used for predictive mapping (Gupta et al., 2021; Tóth et al., 2017; Zhang et al., 2018). However, no map is error free (Heuvelink, 2018) and thus the quantification and assessment of spatial uncertainty has become an essential part of every predictive mapping procedure in soil science (Heuvelink and Webster, 2022; Szatmári and Pásztor, 2019). This is of great importance, as it can help end-users to make appropriate and responsible decisions based on the resulting maps to achieve the desired goals.

The objective of our study is to develop a machine learning-based PTF for predicting the saturated SOC content of the topsoils (0–30 cm) in Hungary with a resolution of 100 m, and then quantify and assess its uncertainty using error propagation analysis. The resulting map is then compared with the actual SOC content map of Hungary (Szatmári et al., 2019b; Szatmári and Pásztor, 2019) in order to assess the degree of SOC deficit in the country.

2. Concept for determining and mapping SOC saturation

The approach we used in this study is based on the principle that topsoils under permanent forest land cover due to the unlimited plant residuum input can be practically considered as saturated in SOC. About 30% of the territory of Hungary is covered by forest on a wide range of soil types, including Luvisols, Podzols, Arenosols, Cambisols, and even Chernozems, among others. Theoretically, some permanent grassland sites may be saturated as well, their involvement, however, is neglected owing to the less reliable land-use history or the role of occasional hydric conditions that might happen on these lowland sites or the effect of grazing.

The quasi SOC saturated forest topsoils have high SOC concentration variance, and thus they are not suitable for a general reference group. Besides, various landscape positions or changes in topographical properties must also affect the way and threshold of SOC saturation for the local scale. Accordingly, only topsoils with similar physicochemical properties (e.g., texture, pH) and environmental conditions (e.g., topography, climate) are comparable directly in terms of SOC saturation. Our concept, therefore, focuses on the measured soil properties and environmental conditions and aims to find relationship between them (i.e., fit a PTF), which can be used to predict saturated SOC content at unvisited sites. Using the developed PTF, the theoretical (potential) SOC saturation value will be predicted across Hungary. Comparing the saturated SOC content with the actual SOC content, the saturation deficit can be determined and assessed.

3. Materials and methods

3.1. Reference soil profiles and harmonization

In this study, we used the soil profiles of the Hungarian Soil Information and Monitoring System (SIMS) (Fig. 1), which is a countrywide soil monitoring system providing geographically referenced biological, physical and chemical information on the status and temporal change of Hungarian soils. It consists of 4,859 soil horizons belonging to 1,236 soil profiles. The following three types of soil profiles are distinguished in SIMS (Fig. 1):

- (i) “T” points (n = 865), which have been placed on agricultural lands (i.e., arable lands, pastures and orchards),
- (ii) “E” points (n = 183), which have been placed in forests,
- (iii) “S” points (n = 188), which have been placed in so-called “hot spot” regions representing different types of environmental hazards (e.g., mining sites, landfills or industrial areas)

Soil profiles placed in forests (i.e., “E” points) have been specifically designed for providing spatio-temporal information on soils in the context of forest ecosystems and therefore they are appropriate for the objective of this study. This means 183 soil profiles (Fig. 1) with 618 soil horizons in total. According to the concept on determining SOC saturation presented in Section 2, we could assume that the topsoils represented by these profiles are practically saturated in SOC and therefore they can be used as reference soil profiles for developing an empirical, data-driven predictive model between SOC saturation, and further soil properties and environmental covariates.

In this study, we used the following chemical and physical soil data associated with the soil horizons of “E” soil profiles: soil organic carbon (SOC) content [unit: %], pH [unit: -], calcium carbonate content [unit: %], as well as sand (>50 μm), silt (50–2 μm), and clay (<2 μm) content [unit: %]. In the following, we will consider data on SOC in these soil

horizons as saturated. The role of the other soil properties listed previously in SOC saturation is frequently discussed in the literature. For instance, a number of papers demonstrated the specific role of the fraction of fine particles (i.e., silt and clay content) in SOC saturation (Chen et al., 2018; Hassink, 1997). Others highlighted the role of specific surface area of the soil (Beare et al., 2014). The presence of cations such as calcium or sodium also has an effect on SOC saturation by creating complexes with SOM molecules that make soil organic matter stable and more resistant to mineralization (Juhos et al., 2021) or more mobile for leaching or decomposition, respectively (Wong et al., 2010). Furthermore, pH provides indispensable information on the acidity/alkalinity of the soil, which basically determines the quality and the degree of polymerization of SOM molecules in soil (Rasmussen et al., 2018).

Since the depth and thickness of soil horizons varies from soil profile to soil profile, there was a need to harmonize all soil data for the topsoil (0–30 cm). We used the mass-preserving spline technique (Bishop et al., 1999; Malone et al., 2011) for modelling the vertical distribution of each of the soil properties listed above at each reference soil profile. The fitted splines were used to derive the values of soil properties for the topsoil at each reference soil profiles. The splined, so-called harmonized values were used in further modelling.

3.2. Spatially exhaustive information on soil and environmental covariates

In Table 1, we summarized the spatially exhaustive information on soil and environmental covariates used in this study. Under the aegis of DOSoReML.hu (<https://www.dosoremi.hu>), a number of soil related properties, functions and services at various depths (e.g., soil depths specified by the GlobalSoilMap initiative (Arrouays et al., 2014)) have been mapped for Hungary using advanced digital soil mapping techniques (Pásztor et al., 2020, 2018, 2017, 2015). In this study, we used the maps of soil properties listed in Section 3.1 as spatially exhaustive

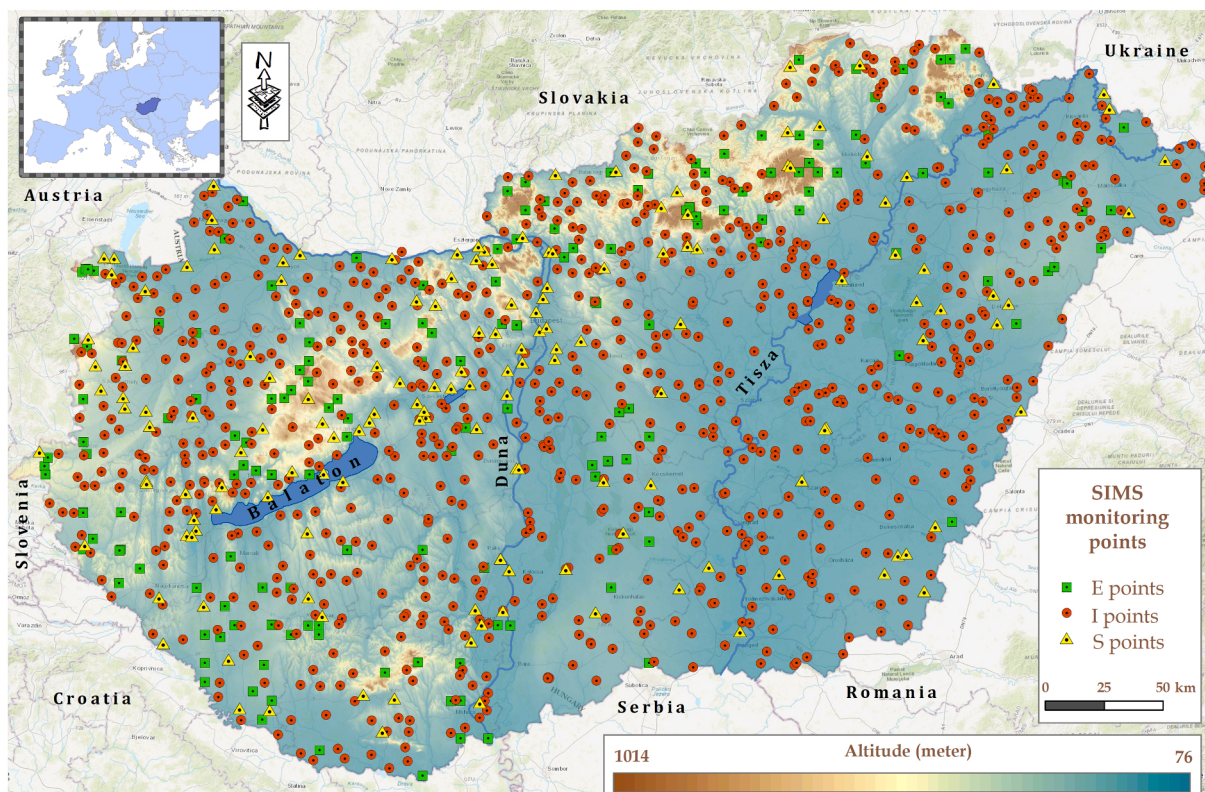


Fig. 1. Location of the soil profiles of the Hungarian Soil Information and Monitoring System (SIMS) (n = 1,236).

Table 1

Summary of the spatially exhaustive information on soil and environmental covariates used in this study. Abbreviation: DEM: digital elevation model, LS factor: slope length-gradient factor, MRVBF: multiresolution valley bottom flatness, MRRTF: multiresolution ridge top flatness, and SOC: soil organic carbon.

Factor	Covariates	Unit	Reference/Source
Soil	Calcium carbonate	[%]	Pásztor et al. (2020)
	pH	[-]	Szatmári et al. (2020)
	Clay	[%]	Laborczi et al. (2019)
	Silt	[%]	Laborczi et al. (2019)
	Sand	[%]	Laborczi et al. (2019)
	Actual SOC*	[%]	Szatmári and Pásztor (2019)
Topography	Elevation	[m]	DEM
	Slope	[%]	DEM
	Profile curvature	[-]	DEM
	Total curvature	[-]	DEM
	Topographic position index	[m]	DEM
	Topographic roughness index	[-]	DEM
	Surface area	[m ²]	DEM
	MRVBF	[-]	DEM
	MRRTF	[-]	DEM
	LS factor	[-]	DEM
	Topographic wetness index	[-]	DEM
	SAGA wetness index	[-]	DEM
	Vertical distance to channel network	[m]	DEM
	Horizontal distance to channel network	[m]	DEM
	Channel network base level	[m]	DEM
	Diurnal anisotropic heating	[-]	DEM
Mass balance index	[-]	DEM	
Stream power index	[-]	DEM	
Climate	Mean annual precipitation	[mm]	Szentimrey and Bihari (2007)
	Mean annual temperature	[°C]	Szentimrey and Bihari (2007)
	Mean annual evapotranspiration	[mm]	Szentimrey and Bihari (2007)
	Mean annual evaporation	[mm]	Szentimrey and Bihari (2007)

*Actual SOC content was only used for assessing SOC deficit.

information on soil. We should note that all these maps have been directly compiled for the topsoil (0–30 cm) with a spatial resolution of 100 m and quantified prediction uncertainty.

In addition to soil information, we also used spatially exhaustive environmental covariates, which may affect SOC saturation. Topography and its attributes were characterized by a countrywide digital elevation model (DEM) (source: Bashfield and Keim, 2011) and a number of geomorphometric parameters derived from that DEM, respectively. The derived geomorphometric parameters are also summarized in Table 1. Their role in SOC saturation is also decisive. For example, LS factor integrates the length and steepness of a slope and therefore provides quantified information on soil erosion susceptibility. Topographic wetness index is in relation with soil moisture associated with topographic position and therefore affects microbiological activity and SOM decomposition. Vertical and horizontal distance to channel network also provide information on the potential water-affectedness of soils.

We characterized climatic conditions by the data layers of long-term mean annual precipitation, temperature, evapotranspiration and evaporation provided by the Hungarian Meteorological Service (Szentimrey and Bihari, 2007). The role of climate in point of SOC saturation is also important because climatic conditions quantitatively determine mineralization of organic carbon in soils. Due to the various data sources, we resampled each of the environmental covariates into a common geographic reference system with a resolution of 100 m.

3.3. Developing pedotransfer function

A regression matrix was created, which contained the location of the reference soil profiles in its rows, as well as the harmonized values of soil properties and the extracted values of environmental covariates associated with these soil profiles in its columns. This regression matrix and the vector of the harmonized values of saturated SOC content were the basis for developing a machine learning-based PTF between saturated SOC content (as response variable), as well as the other soil properties and environmental covariates (as predictors).

Our aim was to apply a machine learning (ML) technique that not just performs outstandingly but also provides fairly easily interpretable predictive model. The *cubist* algorithm suits for these prerequisites since it is a frequently applied ML algorithm with high performance (Adhikari et al., 2014, 2013; Minasny and McBratney, 2008; Viscarra Rossel et al., 2015) and provides transparent model structure (Malone et al., 2017; Miller et al., 2015; Viscarra Rossel and Webster, 2012). In brief, cubist, which is an amalgamation of several methodologies developed mostly in the 1990s by Quinlan (1993, 1992, 1987), is a machine learning technique for generating rule-based predictive models (Kuhn and Johnson, 2013). Using the vector of the response variable's data and the regression matrix of the predictors, cubist partitions the response data into subsets within which their characteristics are similar with respect to the predictors used. A series of rules defines these subsets and these rules are arranged in a hierarchy. Each rule has a specific multivariate linear regression model and takes the form of "IF-THEN-ELSE", where.

- (i) "IF" defines the conditions for the subset,
- (ii) "THEN" contains a specific multivariate linear regression model fitted to the subset,
- (iii) "ELSE" stands for the next rule in the hierarchy.

When prediction of the response variable is targeted, it is checked whether the new values of predictors meet the rule's conditions. If they meet the conditions, then the rule's specific multivariate linear regression model is appropriate for predicting the value of the response variable. If they do not meet the conditions, then the next rule in the hierarchy is visited and its conditions are used for checking the new values. It is repeated till the new values meet one of the rules' conditions in the hierarchy and thus its specific multivariate linear regression model is appropriate for predicting the value of the response variable.

3.4. Quantification and propagation of uncertainty

In Hungary, we have made great efforts in recent years to provide prediction uncertainty to each digital soil mapping product in order to meet the end-users' demand for reliable spatial or spatio-temporal soil information (Pásztor et al., 2020; Szatmári et al., 2021, 2020, 2019b; Szatmári and Pásztor, 2019). Therefore, the spatial uncertainty of the digital soil property maps presented in Table 1 had been quantified and validated earlier (Pásztor et al., 2020) using geostatistical and/or machine learning approaches (Szatmári and Pásztor, 2019). Additionally, the covariance between the interpolation errors of these maps had also been quantified using multivariate geostatistics, as soil properties often show spatial interdependency with each other (Webster and Oliver, 2007), which affects the reliability of uncertainty quantifications especially when uncertainty propagation is also targeted (Heuvelink, 1998). Here we should note that we had no quantified information on the uncertainty of the environmental covariates used in this study. Hence, we were not able to directly account for their uncertainty.

We used a first-order Taylor expansion to propagate uncertainty through the fitted cubist-based PTF, as several studies demonstrated its applicability with PTFs (e.g., Román Dobarco et al., 2019b, 2019a; Styc et al., 2021; Styc and Lagacherie, 2021). Uncertainty propagation analysis can be formulated mathematically as follows (Heuvelink, 1998):

Eq. 1.

$$Y = g(\mathbf{Z})$$

where g is the fitted cubist-based PTF, Y is the saturated soil organic carbon content as the output of the PTF and $\mathbf{Z} = [Z_1, Z_2, \dots, Z_n]^T$ is the vector of the PTF's input variables. The objective is to determine the uncertainty in Y , given the PTF g and the inputs \mathbf{Z} and their associated uncertainty (Heuvelink, 2018). The uncertainty in Y is identified by the variance of Y , which is given by (Heuvelink, 1998):

Eq. 2.

$$\sigma_Y^2 \cong \sum_{i=1}^n \sum_{j=1}^n \rho_{ij} \sigma_i \sigma_j \frac{\delta g}{\delta z_i}(\boldsymbol{\mu}) \frac{\delta g}{\delta z_j}(\boldsymbol{\mu})$$

where n is the number of the PTF's input variables, ρ_{ij} is the correlation coefficient between the uncertainties in z_i and z_j , σ_i and σ_j are the uncertainties in z_i and z_j , respectively, $\frac{\delta g}{\delta z_i}(\boldsymbol{\mu})$ and $\frac{\delta g}{\delta z_j}(\boldsymbol{\mu})$ are the first derivatives of g with respect to z_i and z_j , respectively, and $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_n]^T$ is the vector of the mean values of the n input variables. We should note that if there is no correlation between the input variables (i.e., $\rho_{ij} = 0$), then σ_Y^2 is simply a summation of parts, each attributed to one of the inputs. This partitioning property allows to analyse how much each input contributes to the final uncertainty (Heuvelink, 2018, 1998). This is of great importance when deciding which of the inputs' uncertainty should be diminished to reduce the final uncertainty.

4. Results

4.1. Cubist-based PTF

Table 2 summarizes the descriptive statistics of the harmonized data on saturated SOC content and further soil properties used as predictors in this study. The development of the cubist-based PTF was complemented with 5-times repeated 10-fold cross-validation. According to the results of cross-validation, an R-squared value of 0.56 was obtained, i.e., the developed PTF explained almost 60% of the total variation of the saturated SOC data. Table 3 presents the model structure of the PTF, which not just summarizes the conditions and the specific multivariate linear regression models fitted to the different subsets of saturated SOC content data partitioned in model fitting but also shows the hierarchy via the number of rules (Table 3, first column). More detailed information on the fitted PTF function can be found in the Supplementary Material. Fig. 2 presents the spatially exhaustive soil and environmental covariates, which were found to be informative at the significance level of 0.05 based on the fitted cubist-based PTF (Table 3).

Fig. 3 (left map) presents spatial information on which of the rules and the associated multivariate linear regression model is appropriate for predicting the saturated SOC content on a particular area in Hungary. We should note that the rules identified notable geographical regions:

- (i) Rule 1 covers the sandy areas of Hungary with low SOC content,

Table 2

Summary statistics of saturated soil organic carbon content (SOC_{sat}) and further soil properties harmonized for the topsoil (0–30 cm) at the reference soil profiles ($n = 183$). Abbreviation: SD: standard deviation.

Soil properties	Unit	Minimum	Maximum	Mean	Median	SD
SOC _{sat}	[%]	0.16	6.87	2.17	1.72	1.41
pH	[-]	3.73	8.09	5.72	5.35	1.18
Calcium carbonate	[%]	0.00	20.96	4.83	4.05	3.83
Sand	[%]	2	99	49	44	29
Silt	[%]	0	80	35	35	21
Clay	[%]	0	56	17	16	12

Table 3

The model structure of the cubist-based pedotransfer function used for predicting the saturated soil organic carbon (SOC_{sat}) content in the topsoil (0–30 cm). Note that a significance level of 0.05 was used to select the informative covariates. Annotation: See the Supplementary Material for more detailed information on the fitted pedotransfer function.

Rule	Condition	Specific multivariate linear regression model
1	IF <i>Sand</i> > 61.69	THEN $SOC_{sat} = 0.6 + 0.113 \text{ Clay} - 0.0008 \text{ Sand} - 0.02 \text{ Temperature} + 0.01 \text{ pH}$
2	IF <i>pH</i> ≤ 6.24 and <i>Sand</i> ≤ 61.69 and <i>Slope</i> ≤ 13.96	THEN $SOC_{sat} = 7.6 - 0.0087 \text{ Evaporation} + 0.022 \text{ LS factor} + 0.0004 \text{ Altitude} - 0.0013 \text{ Sand} - 0.03 \text{ Temperature} + 0.02 \text{ pH}$
3	IF <i>pH</i> > 6.24 and <i>Sand</i> ≤ 61.69	THEN $SOC_{sat} = 31.6 - 0.0217 \text{ Evapotranspiration} - 0.8 \text{ pH} - 0.139 \text{ Slope}$
4	IF <i>pH</i> ≤ 6.24 and <i>Sand</i> ≤ 61.69 and <i>Slope</i> > 13.96	THEN $SOC_{sat} = 0.8 + 0.0966 \text{ Sand} + 0.27 \text{ Topographic position index}$

- (ii) Rule 2 refers to the hilly and mountainous regions of Hungary with acidic soils,
 (iii) Rule 3 relates to non-sandy lowlands with neutral or alkali soils,
 (iv) Rule 4 is for the most elevated areas of Hungary with mostly acidic soils.

Increasing rule numbers predict an increasing SOC saturation value along the rules (Table 3). Rule 1 covers the sandy areas of Hungary (Fig. 3), which have the lowest SOC storage capacity due to the lack of the fine fraction. The clay content and the pH are directly connected, whereas the sand content and the mean annual temperature are inverse relationships with the predicted SOC saturation value. Rule 2 relates the acidic areas with lower sand content and slope steepness < 14% (Fig. 3). In this group, higher clay content, altitude, pH, slope angle, and length indicate higher predicted SOC saturation, while the higher evaporation, sand content and mean annual temperature triggers lower SOC saturation values. Rule 3 predicts the SOC saturation values for non-sandy sites with neutral or alkali conditions (Fig. 3). This is the only group where the sand content is not a predictor variable. In contrast, evapotranspiration, pH and slope steepness all decrease the SOC saturation values. Rule 4 covers the acidic sites with high slope steepness (Fig. 3), where both sand content and topographic position index increase the saturated SOC content.

4.2. Applicability of the developed PTF

We should note that the fitted cubist-based PTF was elaborated on sampling points located in permanent forest in Hungary and therefore it is worth investigating whether there are areas where the applicability of the PTF is not recommended. As it was presented in Table 3, the fitted PTF is a set of specific multivariate linear regression models, each having a range within which the given regression model is reasonable to use for predicting the saturated SOC content. However, out of this range the applicability of the given regression model is definitely not recommended due to extrapolation. Therefore, we used the specific multivariate linear regression models (Table 3) to identify areas throughout the entire area of Hungary where extrapolation of the fitted PTF should be expected (Fig. 3, right map). It was found that for 7,922.7 km² the fitted cubist-based PTF cannot be used because of extrapolation. Considering the total area of Hungary (93,030 km²) this means about 8%. Although the areas, where the application of the fitted PTF is not recommended, show scattered pattern throughout the country, notable regions of Hungary can be identified. Almost 90% of the areas

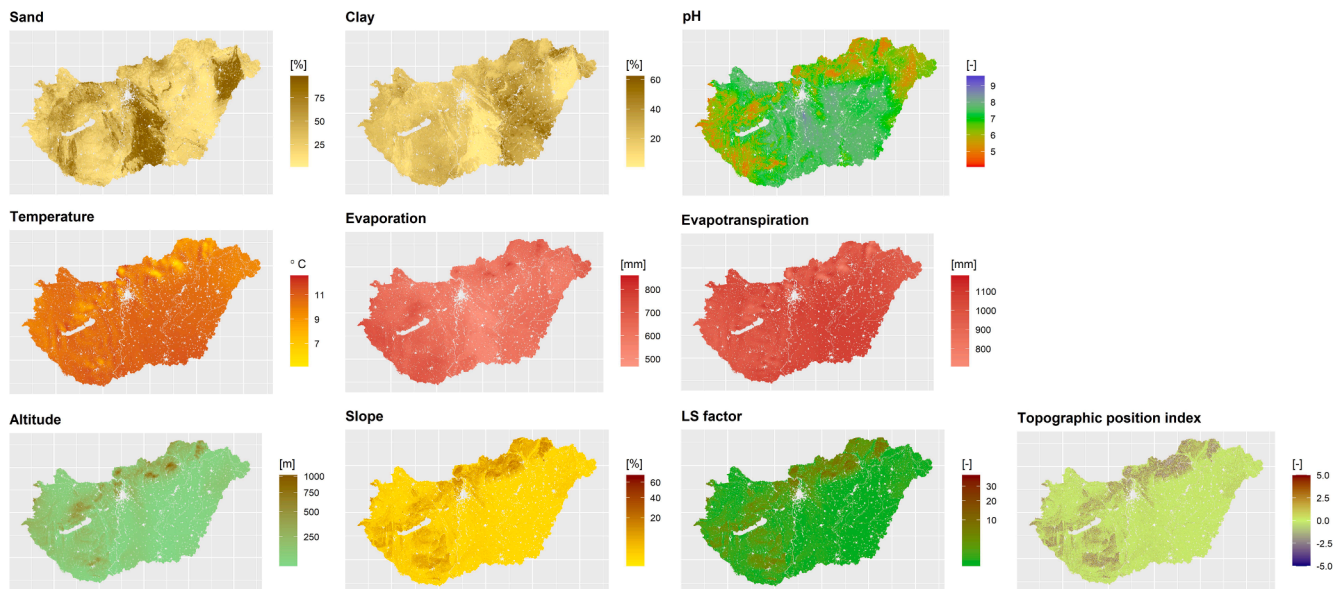


Fig. 2. Spatially exhaustive soil (first row) and environmental covariates (second row: climatic variables, and third row: topographical variables) found to be informative for predicting saturated soil organic carbon content in Hungary. Annotation: Settlements and open water bodies were left blank.

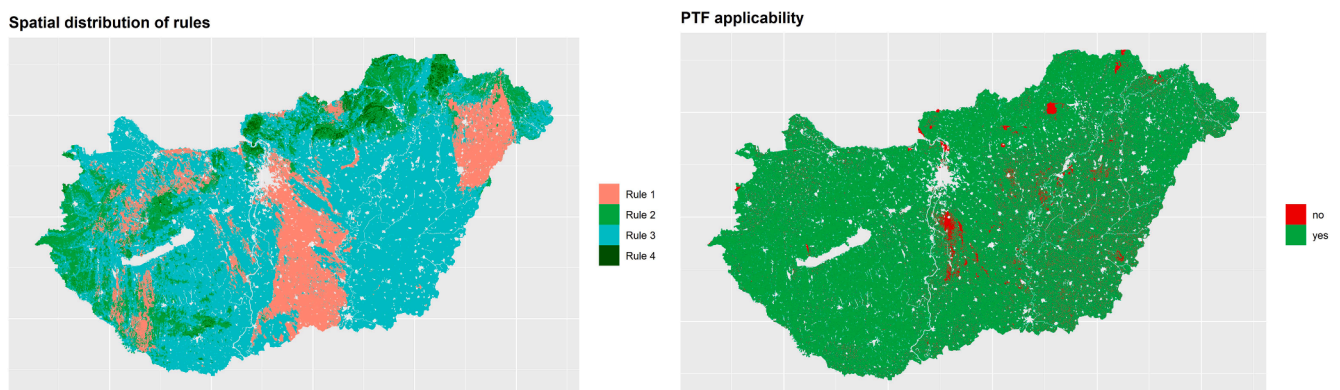


Fig. 3. Spatial distribution of the rules defined by the cubist-based pedotransfer function (left map) and the applicability of the pedotransfer function over Hungary (right map). Annotation: Conditions and multivariate linear regression models associated with the rules can be found in Table 3. Note that settlements and open water bodies were left blank.

characterized by salt-affected soils (e.g., Danube-Tisza Interfluvium, Hortobágy) are proved to be not suitable for the application of the fitted PTF, which can be attributed not just to their extreme sodicity, alkalinity or salinity but also to the fact that their formation is highly influenced by shallow groundwater, permanent or temporary waterlogging. Additionally, areas characterized by highly or moderately water-affected soils are also highlighted along the main rivers (e.g. Danube, Tisza, Dráva) and in the eastern part of the country, which is completely acceptable as the concept introduced in Section 2 is referred to mineral soils and not to water-affected soils, where the balance between SOM accumulation and mineralization is strongly affected by surplus water. This also true for the peatlands, which can be easily identified on the map too (e.g. southeastern part of Lake Balaton).

4.3. Spatial prediction of saturated SOC content with uncertainty propagation

Fig. 4.A presents the spatial distribution of saturated SOC content over Hungary as a result of the application of the fitted cubist-based PTF (Table 3) on spatially exhaustive soil and environmental covariates (Fig. 2). Next to this map we also presented its variance (Fig. 4.B), as one

of the main results of the uncertainty propagation analysis performed. This map can be interpreted as the uncertainty in the saturated SOC content spatial predictions. We found that higher uncertainty is associated with areas where extrapolation of the fitted PTF is expected (Fig. 3, right map). We should also highlight that the final uncertainty associated with the saturated SOC content predictions takes into consideration not just the uncertainty associated with the fitted PTF but also the prediction uncertainty of the digital soil mapping products listed in Table 1. As was mentioned in Section 3.4, the partitioning property allows to analyze how much each input contributes to the final uncertainty. Thus, we determined the contributions and relative contributions of the input soil maps (Fig. 4.C and E) and the cubist-based PTF (Fig. 4.D and F) to the uncertainty in the saturated SOC content spatial predictions (Fig. 4.B). As we had no information about the uncertainty of the environmental covariates (Fig. 2, second and third row), we could not determine their contributions to the final uncertainty. However, their uncertainty contributes to the uncertainty in the fitted PTF (Fig. 4. D and F). This is because the cubist-based PTF was elaborated on their uncertain data and not on measured ones as in the case of soil inputs, where we used actual measurements from SIMS. Hence, uncertainty in the environmental covariates (Fig. 2, second and third row) is also

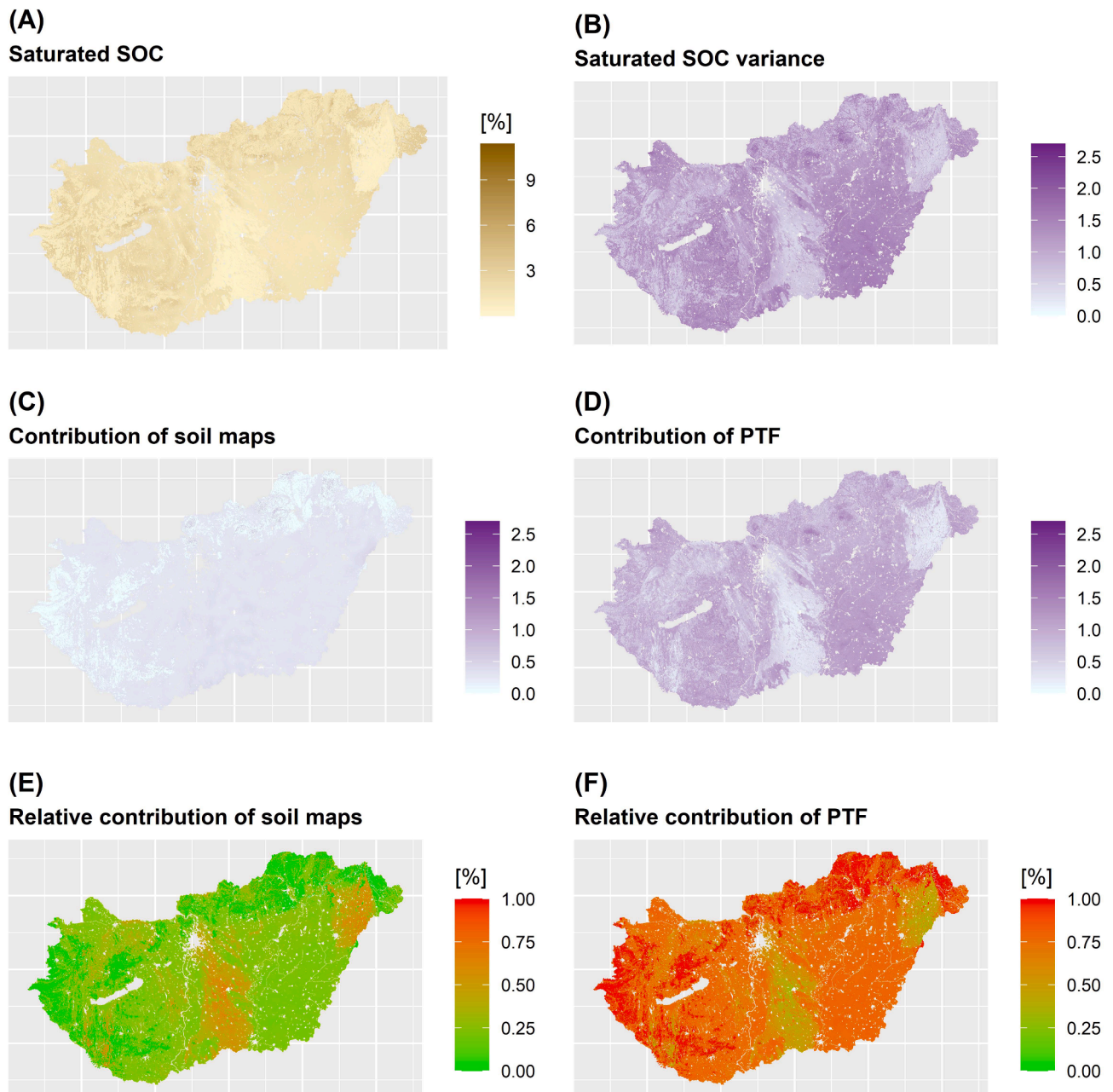


Fig. 4. Saturated soil organic carbon content for the topsoil (0–30 cm) (A) and its variance as a result of uncertainty propagation analysis (B). Contributions and relative contributions of the input soil maps (C and E) and the cubist-based pedotransfer function (D and F) to the uncertainty in saturated soil organic carbon content spatial predictions (B). Abbreviations: SOC: soil organic carbon, and PTF: pedotransfer function. Annotation: Settlements and open water bodies were left blank.

represented in the PTF uncertainty. By comparing the relative contributions (Fig. 4.E and F), it was found that the contribution of PTF uncertainty to the final uncertainty is much larger than the contribution of the soil maps uncertainty.

4.4. SOC deficit in Hungary

We compared the saturated SOC content map with the actual SOC map of Hungary (Fig. 5.A) (Szatmári and Pásztor, 2019). In brief, the actual SOC map also refers to the topsoil (0–30 cm) and has been created by using a combination of geostatistics and advanced ML techniques. A detailed description on the soil data, environmental covariates and digital soil mapping methodology used for predicting the spatial distribution of the actual SOC content can be found in Szatmári and Pásztor (2019). By subtracting the saturated SOC map from the actual SOC map,

we compiled a difference map presenting the SOC deficit for the topsoil in Hungary (Fig. 5.C). On the map, negative values (i.e., red colors) present areas with SOC deficit, whereas positive values (i.e., blue colors) show areas where soils are saturated or oversaturated in SOC. According to the compiled difference map, not just large parts of the country (ca. 80%) can be characterized by SOC deficit but it also shows high spatial variability across Hungary. Note that the uncertainty in the spatial predictions of saturated SOC content (Fig. 4.B) is an order of magnitude larger than the uncertainty in the spatial predictions of actual SOC (Fig. 5.B). Because of this the uncertainty in the SOC deficit map (Fig. 5.C) must be also large, which does not allow to delimit areas with statistically significant SOC deficit in the country.

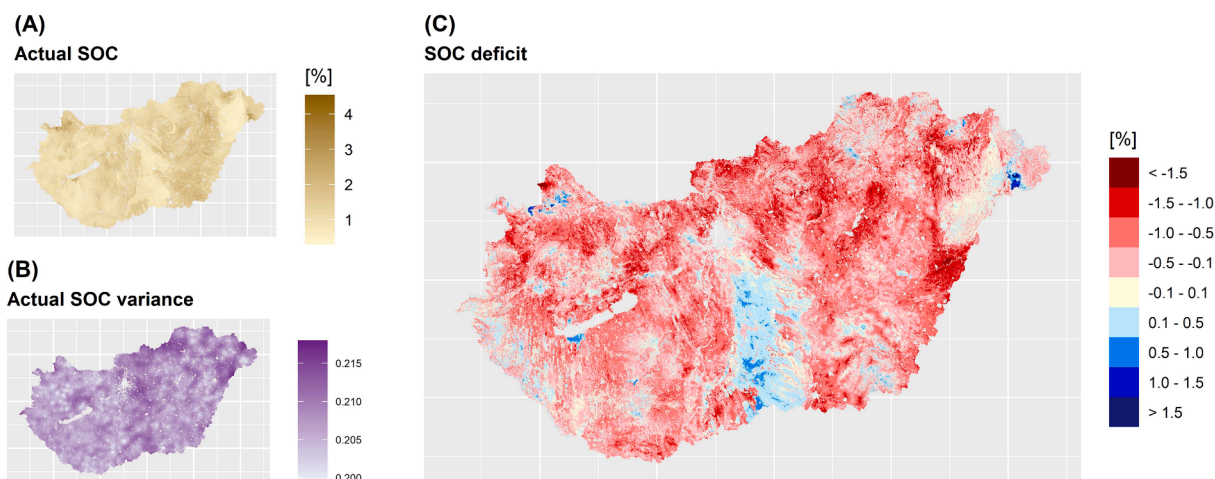


Fig. 5. Spatial distribution of the actual soil organic carbon (SOC) content (A) and the associated prediction uncertainty (B), and the map of SOC deficit (C). Annotation: Settlements and open water bodies were left blank.

5. Discussion

5.1. Interpretation and discussion of the results

The sand content was the most important predictor, which was included in three out of the four regression models. In two cases, it was inversely proportional with the SOC saturation value, even though in Rule 4, on the steepest parts sand content was proportional with SOC saturation (Table 3). This may result from the increased infiltration due to the coarser texture resulting in a lower amount of runoff and soil/carbon loss (Nagy et al., 2020). The other general predictor is pH which regulates the predicted SOC saturation value in the Rules 1–3 (Table 3). It has a positive effect on sandy soils and under acidic and neutral conditions. In contrast, in the alkali range, pH has a very strong inverse relationship with SOC saturation. This may indicate the mitigated SOC saturation potential under sodic conditions (Rasmussen et al., 2018; Wong et al., 2010), even though in the current case, the spatial prediction is unreliable for sodic soils due to extrapolation (Fig. 3, right map). Regarding the climatic effects, both evaporation and evapotranspiration decrease SOC saturation on non-sandy and non-hilly fields. Mean annual temperature mitigates SOC saturation as was reported by several studies (Wiesmeier et al., 2014; Zhang et al., 2015). Topography provided ambivalent results as slope steepness decreased SOC saturation values on the flat areas, but topographic position index and the LS factor were in direct linkage with the saturation. Taking the spatial resolution of the present study into account, the erosion and soil redistribution processes (Ghosh et al., 2021) may result in a more detailed spatial pattern (Centeri et al., 2014).

We should highlight that the compiled map of saturated SOC content (Fig. 4.A) should be interpreted together with the PTF's applicability map (Fig. 3, right map), as it was found that for 8% of the area of Hungary extrapolation of the fitted PTF should be expected. Therefore it is better to disregard those areas from the assessment, where the PTF's applicability is not recommended. As a general trend in Hungarian topsoils, the highest SOC deficits (Fig. 5.C) occur in areas with medium to high actual SOC content (Fig. 5.A). Base saturated Chernozems and Kastanozems storing more than two percent of SOC have the capacity to sequester an additional one percent of SOC mainly on the plains of Hungary (Fig. 5.C). The mountainous parts of the country as well as the acidic fields have lower capacity, however, it still means 0.1–0.5% additional SOC storage potential.

5.2. Limitations and further research

The estimated SOC saturation deficit value is rather theoretical with practical applicability. The saturated forest sites used as references have no tillage but practically unlimited organic matter input. On cropfields, however, a considerable amount of organic matter is gathered as yield and straw thus, the OM input is limited. Moreover, tillage increases SOM mineralization. Consequently, the theoretically calculated saturation value is hardly available for the practice without land use change (Stewart et al., 2007). Ensuring food security, most cropfields must be kept as arable land in the long run, accordingly loading the total amount of calculated SOC deficit is unlikely. Nevertheless, the application of conservation agriculture technics is believed to relevantly increase the SOC content under cropfields (Gelybó et al., 2022; Jakab et al., 2022; Madarász et al., 2021). Thus the amount of SOC saturation deficit may be an efficient indicator to identify hot spots where cultivation shift would benefit most, independently of the theoretical amount of SOC saturation.

On the other hand, we also should note that considerable parts of the country have been indicated as saturated or oversaturated (Fig. 5.C). The highest values of oversaturation are mostly related to peatlands (e. g., southwestern part of Lake Balaton) and other water affected areas. As it has been already mentioned in the Introduction, on these sites, the temporal or former water coverage inhibited the mineralization of soil organic matter, which process cannot be captured by the developed PTF (Fig. 3, right map) since the approach used in this study was based solely on forest sites. Further reason of (over-)saturation is related to the areas covered by Arenosols, especially in the Danube-Tisza Interfluve (Fig. 5. C). These soils can be characterized by basically low SOC concentration, where low changes may trigger oversaturation. Under the above conditions, the role of particulate organic matter or fresh plant debris may cause increased variability in Arenosols resulting in oversaturation.

In this study, we used the monitoring points of SIMS located in permanent forests (i.e., the “E” points) as reference soil profiles, which is a coherent dataset because the description of soil profiles, field sampling and laboratory analyses have been carried out according to a predefined protocol based on the Hungarian Standards. We should note that only a limited number of soil profiles located in permanent forests are available ($n = 183$), therefore more effort should be made in the future to collect additional soil samples from permanent forests across Hungary. The age of these permanent forests may also affect the current saturation degree if they are afforested in the recent past thus also, the age of forests would be useful to include in further studies.

The uncertainty associated with the saturated SOC content spatial

predictions was larger than the uncertainty of the actual SOC content by an order of magnitude (Fig. 4.B and 5.B), which made it impossible to delimit areas with statistically significant SOC deficit in Hungary. Note that this is not unique in the literature, a number of papers demonstrated that large prediction uncertainty can put restraint to digital soil assessment (Heuvelink et al., 2020; Szatmári et al., 2021, 2019b). However, the methodology presented in this study and the compiled maps provided useful preliminary results on identifying and delimiting areas with SOC deficit in Hungary, which should be fine-tuned in the future with probably more additional observations on saturated SOC content. As it was pointed out by the results of uncertainty propagation analysis, this large amount of uncertainty can be mostly attributed to the fitted PTF (Fig. 4.D and F). This is of great importance, as it basically gives the direction on how to design the additional sampling campaign in order to successfully improve the cubist-based PTF and at the same time reduce its uncertainty. Besides, the cubist algorithm also proved to be an efficient ML technique not just for developing PTF but also for making the resulting PTF more transparent for users, an issue frequently addressed in recent studies (Szatmári et al., 2020; Wadoux et al., 2020). The transparency of the model structure can help us not just to better understand soil and environmental conditions affecting SOC saturation but also to optimize the number and location of additional sampling points making the additional soil survey more cost-effective. In the last few years, a number of studies have addressed the issue of optimizing sampling design with known model structure and demonstrated that uncertainty can be reduced (Ließ, 2015; Szatmári et al., 2019a, 2015; Wadoux et al., 2019).

6. Conclusions

Our objective was to develop a cubist-based PTF, which can be used for predicting and mapping the saturated SOC content of the topsoils in Hungary and then compare the resulting map with the actual SOC map in order to assess the degree of SOC deficit in the country. The fitted cubist-based PTF proved to be efficient in predicting SOC saturation across Hungary, furthermore, the cubist algorithm gave a fairly transparent model structure, which helped to better understand the factors and driving forces of SOC saturation. We have highlighted that not just the physicochemical properties of soils, but also environmental conditions, such as topography and climate, characterizing landscape are important factors in predicting the level of potential SOC saturation. Our results also pointed out that there is SOC deficit on large part of the country (ca. 80%), which shows high spatial variability. Besides, the most considerable potential for additional SOC sequestration was found related to soils with medium to high actual SOC content.

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.

Acknowledgments

This research was funded by the National Research, Development and Innovation Office (NKFIH; grant numbers: K-123953 and K-131820), and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences (G. Szatmári).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.catena.2023.107086>.

References

- Adhikari, K., Hartemink, A.E., Minasny, B., Bou Kheir, R., Greve, M.B., Greve, M.H., 2014. Digital mapping of soil organic carbon contents and stocks in Denmark. *PLoS One* 9. <https://doi.org/10.1371/journal.pone.0105519>.
- Adhikari, K., Kheir, R.B., Greve, M.B., Bocher, P.K., Malone, B.P., Minasny, B., McBratney, A.B., Greve, M.H., 2013. High-Resolution 3-D Mapping of Soil Texture in Denmark. *Soil Sci. Soc. Am. J.* 77, 860. <https://doi.org/10.2136/sssaj2012.0275>.
- Arrouays, D., McBratney, A.B., Minasny, B., Hempel, J.W., Heuvelink, G.B.M., MacMillan, R.A., Hartemink, A.E., Lagacherie, P., McKenzie, N.J., 2014. The GlobalSoilMap project specifications. *Glob. Basis Glob. Spat. Soil Inf. Syst. - Proc. 1st Glob. Conf.* <https://doi.org/10.1201/b16500-4>.
- Barré, P., Denis, A., Basile-Doelsch, I., Bispo, A., Lauric, C., Claire, C., Tiphaine, C., Delphine, D., Thomas, E., Sylvain, P., Angers, D.A., Cécillon, L., Chenu, C., Chevallier, T., Derrien, D., Eglin, T.K., Pellerin, S., 2017. Ideas and perspectives: Can we use the soil carbon saturation deficit to quantitatively assess the soil carbon storage potential, or should we explore other strategies? *Biogeosciences Discuss.* 1–12. <https://doi.org/10.5194/bg-2017-3951>.
- Bashfield, A., Keim, A., 2011. Continent-wide DEM Creation for the European Union, in: 34th International Symposium on Remote Sensing of Environment - The GEOSS Era: Towards Operational Environmental Monitoring.
- Beare, M.H., McNeill, S.J., Curtin, D., Parfitt, R.L., Jones, H.S., Dodd, M.B., Sharp, J., 2014. Estimating the organic carbon stabilisation capacity and saturation deficit of soils: A New Zealand case study. *Biogeochemistry* 120, 71–87. <https://doi.org/10.1007/s10533-014-9982-1>.
- 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](https://doi.org/10.1016/S0016-7061(99)00003-8).
- Bouma, J., 1989. *Using Soil Survey Data for Quantitative Land Evaluation*. Springer, New York, NY, pp. 177–213.
- Briedis, C., de Moraes Sá, J.C., Lal, R., Tivet, F., Franchini, J.C., de Oliveira Ferreira, A., da Cruz Hartman, D., Schimiguel, R., Bressan, P.T., Inagaki, T.M., Romaniw, J., Gonçalves, D.R.P., 2018. How does no-till deliver carbon stabilization and saturation in highly weathered soils? *Catena* 163, 13–23. <https://doi.org/10.1016/j.catena.2017.12.003>.
- Cardinael, R., Chevallier, T., Barthès, B.G., Saby, N.P.A., Parent, T., Dupraz, C., Bernoux, M., Chenu, C., 2015. Impact of alley cropping agroforestry on stocks, forms and spatial distribution of soil organic carbon - A case study in a Mediterranean context. *Geoderma* 259–260, 288–299. <https://doi.org/10.1016/j.geoderma.2015.06.015>.
- 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. *Types, Management Practices and Environmental Benefits, Soil Carbon*.
- Chen, S., Arrouays, D., Angers, D.A., Chenu, C., Barré, P., Martin, M.P., Saby, N.P.A., Walter, C., 2019. National estimation of soil organic carbon storage potential for arable soils: A data-driven approach coupled with carbon-landscape zones. *Sci. Total Environ.* 666, 355–367. <https://doi.org/10.1016/j.scitotenv.2019.02.249>.
- Chen, S., Arrouays, D., Angers, D.A., Martin, M.P., Walter, C., 2018. Soil carbon stocks under different land uses and the applicability of the soil carbon saturation concept. *Soil Tillage Res.* 0–1. <https://doi.org/10.1016/j.still.2018.11.001>.
- Chenu, C., Angers, D.A., Barré, P., Derrien, D., Arrouays, D., Balesdent, J., 2019. Increasing organic stocks in agricultural soils: Knowledge gaps and potential innovations. *Soil Tillage Res.* 188, 41–52. <https://doi.org/10.1016/j.still.2018.04.011>.
- Chimento, C., Almagro, M., Amaducci, S., 2016. Carbon sequestration potential in perennial bioenergy crops: the importance of organic matter inputs and its physical protection. *GCB Bioenergy* 8, 111–121. <https://doi.org/10.1111/gcbb.12232>.
- Cotrufo, M.F., Ranalli, M.G., Haddix, M.L., Six, J., Lugato, E., 2019. Soil carbon storage informed by particulate and mineral-associated organic matter. *Nat. Geosci.* 12, 989–994. <https://doi.org/10.1038/s41561-019-0484-6>.
- Fekete, I., Berki, I., Lajtha, K., Trumbore, S., Francioso, O., Gioacchini, P., Montecchio, D., Várbró, G., Béni, Á., Makádi, M., Demeter, I., Madarász, B., Juhos, K., Kotrocó, Z., 2021. How will a drier climate change carbon sequestration in soils of the deciduous forests of Central Europe? *Biogeochemistry* 152, 13–32. <https://doi.org/10.1007/s10533-020-00728-w>.
- Feng, W., Plante, A.F., Six, J., 2013. Improving estimates of maximal organic carbon stabilization by fine soil particles. *Biogeochemistry* 112, 81–93. <https://doi.org/10.1007/s10533-011-9679-7>.
- Gelybó, G., Barcza, Z., Dencsó, M., Potyó, I., Kása, I., Horel, P., Kovai, K., Birkás, M., Kern, A., Hollós, R., Tóth, E., 2022. Effect of tillage and crop type on soil respiration in a long-term field experiment on chernozem soil under temperate climate. *Soil Tillage Res.* 216. <https://doi.org/10.1016/j.still.2021.105239>.
- Ghosh, A., Singh, A.K., Kumar, S., Manna, M.C., Jha, P., Bhattacharyya, R., Sannagoudar, M.S., Singh, R., Chaudhari, S.K., Kumar, R.V., 2021. Do moisture conservation practices influence stability of soil organic carbon and structure? *Catena* 199, 105127. <https://doi.org/10.1016/j.catena.2020.105127>.
- Gupta, S., Lehmann, P., Bonetti, S., Papritz, A., Or, D., 2021. Global Prediction of Soil Saturated Hydraulic Conductivity Using Random Forest in a Covariate-Based GeoTransfer Function (CoGTF) Framework. *J. Adv. Model. Earth Syst.* 13, e2020MS002242. <https://doi.org/10.1029/2020MS002242>.
- Hanegraaf, M.C., Hoffland, E., Kuikman, P.J., Brussaard, L., 2009. Trends in soil organic matter contents in Dutch grasslands and maize fields on sandy soils. *Eur. J. Soil Sci.* 60, 213–222. <https://doi.org/10.1111/J.1365-2389.2008.01115.X>.

- Hassink, J., 1997. The capacity of soils to preserve organic C and N by their association with clay and silt particles. *Plant Soil* 191, 77–87. <https://doi.org/10.1023/A:1004213929699>.
- Heuvelink, G.B.M., 2018. Uncertainty and Uncertainty Propagation in Soil Mapping and Modelling. Springer, Cham, pp. 439–461. https://doi.org/10.1007/978-3-319-63439-5_14.
- Heuvelink, G.B.M., 1998. Error propagation in environmental modelling with GIS. Taylor and Francis. <https://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://doi.org/10.1111/ejss.12998>.
- Heuvelink, G.B.M., Webster, R., 2022. Spatial statistics and soil mapping: A blossoming partnership under pressure. *Spat. Stat.* 100639 <https://doi.org/10.1016/J.SPASTA.2022.100639>.
- Inagaki, T.M., Possinger, A.R., Grant, K.E., Schweizer, S.A., Mueller, C.W., Derry, L.A., Lehmann, J., Kögel-Knabner, I., 2020. Subsoil organo-mineral associations under contrasting climate conditions. *Geochim. Cosmochim. Acta* 270, 244–263. <https://doi.org/10.1016/j.gca.2019.11.030>.
- Jakab, G., Vancsik, A., Filep, T., Madarász, B., Zacháry, D., Ringer, M., Ujházy, N., Szalai, Z., 2022. Soil organic matter characterisation using alkali and water extraction, and its relation to soil properties. *Geoderma Reg.* 28, e00469.
- Juhos, K., Madarász, B., Kotroczo, Z., Béni, Á., Makádi, M., Fekete, I., 2021. Carbon sequestration of forest soils is reflected by changes in physicochemical soil indicators – A comprehensive discussion of a long-term experiment on a detritus manipulation. *Geoderma* 385, 114918. <https://doi.org/10.1016/j.geoderma.2020.114918>.
- 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>.
- Kögel-knabner, I., Ekschmitt, K., Flessa, H., Guggenberger, G., Matzner, E., Marschner, B., Lützw, M.V., 2008. An integrative approach of organic matter stabilization in temperate soils: Linking chemistry, physics, and biology § 1090, 5–13. <https://doi.org/10.1002/jpln.200700215>.
- Kuhn, M., Johnson, K., 2013. Applied predictive modeling, Applied Predictive Modeling. Springer New York. <https://doi.org/10.1007/978-1-4614-6849-3>.
- Laborczi, A., Szatmári, G., Kaposi, A.D., Pásztor, L., 2019. Comparison of soil texture maps synthesized from standard depth layers with directly compiled products. *Geoderma* 352, 360–372. <https://doi.org/10.1016/j.geoderma.2018.01.020>.
- 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 Security. *Science (80-)* 304, 1623–1627.
- Lieb, M., 2015. Sampling for regression-based digital soil mapping: Closing the gap between statistical desires and operational applicability. *Spat. Stat.* 13, 106–122. <https://doi.org/10.1016/j.spasta.2015.06.002>.
- Madarász, B., Jakab, G., Szalai, Z., Juhos, K., Kotroczo, Z., Tóth, A., Ladányi, M., 2021. Long-term effects of conservation tillage on soil erosion in Central Europe: A random forest-based approach. *Soil Tillage Res.* 209, 104959 <https://doi.org/10.1016/J.STILL.2021.104959>.
- Makó, A., Tóth, G., Weynants, M., Rajkai, K., Hermann, T., Tóth, B., 2017. Pedotransfer functions for converting laser diffraction particle-size data to conventional values. *Eur. J. Soil Sci.* 68, 769–782. <https://doi.org/10.1111/ejss.12456>.
- Malone, B.P., McBratney, A.B., Minasny, B., 2011. Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. *Geoderma* 160, 614–626. <https://doi.org/10.1016/j.geoderma.2010.11.013>.
- Malone, B.P., Minasny, B., McBratney, A.B., 2017. Using R for Digital Soil Mapping, Progress in Soil Science. Springer International Publishing, Cham. <https://doi.org/10.1007/978-3-319-44327-0>.
- 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., Stolbovov, V., Stockmann, U., Sulaeman, Y., Tsui, C.C., Vágen, 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>.
- Minasny, B., McBratney, A.B., 2008. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. *Chemom. Intell. Lab. Syst.* 94, 72–79. <https://doi.org/10.1016/j.chemolab.2008.06.003>.
- Nagy, G., Lóczy, D., Czigány, S., Pirkhoffer, E., Fábíán, S.Á., Ciglić, R., Ferk, M., 2020. Soil moisture retention on slopes under different agricultural land uses in hilly regions of Southern Transdanubia. *Hungarian Geogr. Bull.* 69, 263–280. <https://doi.org/10.15201/HUNGEOBULL.69.3.3>.
- O'Rourke, S.M., Angers, D.A., Holden, N.M., McBratney, A.B., 2015. Soil organic carbon across scales. *Glob. Chang. Biol.* 21, 3561–3574. <https://doi.org/10.1111/gcb.12959>.
- 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>.
- Pásztor, L., Laborczi, A., Takács, K., Szatmári, G., Bakacsi, Z., Szabó, J., Illés, G., Laborczi, A., Takács, K., Szatmári, G., Bakacsi, Z., Szabó, J., Illés, G., 2018. DOSoReMI as the national implementation of GlobalSoilMap for the territory of Hungary. In: Arrouays, D., Savin, I., Leenaars, J.G.B., McBratney, A.B. (Eds.), *GlobalSoilMap - Digital Soil Mapping from Country to Globe*. CRC Press, pp. 17–22. <https://doi.org/10.1201/9781351239707-4>.
- Pásztor, L., Laborczi, A., Takács, K., 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., Szatmári, G., Fodor, N., Illés, G., Farkas-Iványi, K., Bakacsi, Z., Szabó, J., 2017. Compilation of Functional Soil Maps for the Support of Spatial Planning and Land Management in Hungary. In: Pereira, P., Brevik, E.C., Munoz-Rojas, M., Miller, B.A. (Eds.), *Soil Mapping and Process Modeling for Sustainable Land Use Management*. Elsevier Ltd, Amsterdam, pp. 293–317 <https://doi.org/10.1016/B978-0-12-805200-6.00009-8>.
- Quinlan, J.R., 1993. Combining Instance-Based and Model-Based Learning. Morgan Kaufmann.
- Quinlan, J.R., 1992. LEARNING WITH CONTINUOUS CLASSES. *World Scientific*.
- Quinlan, J.R., 1987. Simplifying decision trees. *Int. J. Man. Mach. Stud.* 27, 221–234. [https://doi.org/10.1016/S0020-7373\(87\)80053-6](https://doi.org/10.1016/S0020-7373(87)80053-6).
- Rasmussen, C., Heckman, K., Wieder, W.R., Keiluweit, M., Lawrence, C.R., Berhe, A.A., Blankinship, J.C., Crow, S.E., Druhan, J.L., Hicks Pries, C.E., Marin-Spiotta, E., Plante, A.F., Schädel, C., Schimel, J.P., Sierra, C.A., Thompson, A., Wagai, R., 2018. Beyond clay: towards an improved set of variables for predicting soil organic matter content. *Biogeochemistry* 137, 297–306. <https://doi.org/10.1007/s10533-018-0424-3>.
- Román Dobarco, M., Bourennane, H., Arrouays, D., Saby, N.P.A., Cousin, I., Martin, M.P., 2019a. Uncertainty assessment of GlobalSoilMap soil available water capacity products: A French case study. *Geoderma* 344, 14–30. <https://doi.org/10.1016/J.GEODERMA.2019.02.036>.
- Román Dobarco, M., Cousin, I., Le Bas, C., Martin, M.P., 2019b. Pedotransfer functions for predicting available water capacity in French soils, their applicability domain and associated uncertainty. *Geoderma* 336, 81–95. <https://doi.org/10.1016/J.GEODERMA.2018.08.022>.
- Sanderman, J., Farquharson, R., Baldock, J., 2010. Soil Carbon Sequestration Potential: A review for Australian agriculture A report prepared for Department of Climate Change and Energy Efficiency. CSIRO Rep.
- Schmidt, M.W.L., Torn, M.S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I.A., Kleber, M., Kögel-Knabner, I., Lehmann, J., Manning, D.A.C., Nannipieri, P., Rasse, D.P., Weiner, S., Trumbore, S.E., 2011. Persistence of soil organic matter as an ecosystem property. *Nature* 478, 49–56. <https://doi.org/10.1038/nature10386>.
- Shiri, J., Keshavarzi, A., Kisi, O., Karimi, S., Iturraran-Viveros, U., 2017. Modeling soil bulk density through a complete data scanning procedure: Heuristic alternatives. *J. Hydrol.* 549, 592–602. <https://doi.org/10.1016/j.jhydrol.2017.04.035>.
- Six, J., Feller, C., Denef, K., Ogle, S., Sa, J.C.D.M., Albrecht, A., 2002. Soil organic matter, biota and aggregation in temperate and tropical soils - Effects of no-tillage. *Agronomie* 22, 755–775. <https://doi.org/10.1051/agro:2002043>.
- 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>.
- Stewart, C.E., Paustian, K., Conant, R.T., Plante, A.F., Six, J., 2007. Soil carbon saturation: Concept, evidence and evaluation. *Biogeochemistry* 86, 19–31. <https://doi.org/10.1007/s10533-007-9140-0>.
- Styc, Q., Gontard, F., Lagacherie, P., 2021. Harvesting spatially dense legacy soil datasets for digital soil mapping of available water capacity in Southern France. *Geoderma Reg.* 24, e00353.
- Styc, Q., Lagacherie, P., 2021. Uncertainty assessment of soil available water capacity using error propagation: A test in Languedoc-Roussillon. *Geoderma* 391, 114968. <https://doi.org/10.1016/J.GEODERMA.2021.114968>.
- Suleymanov, A., Gabbasova, I., Suleymanov, R., Abakumov, E., Polyakov, V., Liebelt, P., 2021. Mapping soil organic carbon under erosion processes using remote sensing. *Hungarian Geogr. Bull.* 70, 49–64. <https://doi.org/10.15201/hungeobull.70.1.4>.
- Szabó, B., Szatmári, G., Takács, K., Laborczi, A., Makó, A., Rajkai, K., Pásztor, L., 2019. Mapping soil hydraulic properties using random-forest-based pedotransfer functions and geostatistics. *Hydrol. Earth Syst. Sci.* 23, 2615–2635. <https://doi.org/10.5194/hess-23-2615-2019>.
- Szatmári, G., Bakacsi, 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., Barta, K., Pásztor, L., 2015. An application of a spatial simulated annealing sampling optimization algorithm to support digital soil mapping. *Hungarian Geogr. Bull.* 64, 10.15201/hungeobull.64.1.4.
- Szatmári, G., László, P., Takács, K., Szabó, J., Bakacsi, Z., Koós, S., Pásztor, L., 2019a. Optimization of second-phase sampling for multivariate soil mapping purposes: Case study from a wine region, Hungary. *Geoderma* 352, 373–384. <https://doi.org/10.1016/j.geoderma.2018.02.030>.
- 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., 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., Pirkó, B., Koós, S., Laborczi, A., Bakacsi, Z., Szabó, J., Pásztor, L., 2019b. 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>.
- Szentimrey, T., Bihari, Z., 2007. Mathematical background of the spatial interpolation methods and the software MISH (Meteorological Interpolation based on Surface

- Homogenized Data Basis). In: in: *Proceedings from the Conference on Spatial Interpolation in Climatology and Meteorology*, pp. 17–27.
- T. Aide, M., Aide, C., Braden, I., 2020. Soil Genesis of Histosols and Gelisols with a Emphasis on Soil Processes Supporting Carbon Sequestration, in: *Environmental Change and Sustainability [Working Title]*. IntechOpen. <https://doi.org/10.5772/intechopen.94399>.
- Tóth, B., Weynants, M., Pásztor, L., Hengl, T., 2017. 3D soil hydraulic database of Europe at 250 m resolution. *Hydrol. Process.* 31, 2662–2666. <https://doi.org/10.1002/hyp.11203>.
- Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., Montzka, C., Nemes, A., Pachepsky, Y.A., Padarian, J., Schaap, M.G., Tóth, B., Verhoef, A., Vanderborght, J., van der Ploeg, M.J., Weihermüller, L., Zacharias, S., Zhang, Y., Vereecken, H., 2017. Pedotransfer Functions in Earth System Science: Challenges and Perspectives. *Rev. Geophys.* <https://doi.org/10.1002/2017RG000581>.
- 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>.
- Viscarra Rossel, R.A., Webster, R., 2012. Predicting soil properties from the Australian soil visible-near infrared spectroscopic database. *Eur. J. Soil Sci.* 63, 848–860. <https://doi.org/10.1111/j.1365-2389.2012.01495.x>.
- Wadoux, A.-M.-J.-C., Minasny, B., McBratney, A.B., 2020. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Rev.* 210, 103359 <https://doi.org/10.1016/j.earscirev.2020.103359>.
- Wadoux, A.M.J.C., Brus, D.J., Heuvelink, G.B.M., 2019. Sampling design optimization for soil mapping with random forest. *Geoderma* 355, 113913. <https://doi.org/10.1016/j.geoderma.2019.113913>.
- Webster, R., Oliver, M.A., 2007. *Geostatistics for environmental scientists*, 2nd ed. Wiley.
- Wiesmeier, M., Hübner, R., Spörlein, P., Geuß, U., Hangen, E., Reischl, A., Schilling, B., von Lütow, M., Kögel-Knabner, I., 2014. Carbon sequestration potential of soils in southeast Germany derived from stable soil organic carbon saturation. *Glob. Chang. Biol.* 20, 653–665. <https://doi.org/10.1111/gcb.12384>.
- Witzgall, K., Vidal, A., Schubert, D.I., Höschen, C., Schweizer, S.A., Buegger, F., Pouteau, V., Chenu, C., Mueller, C.W., 2021. Particulate organic matter as a functional soil component for persistent soil organic carbon. *Nat. Commun.* 2021 121 12, 1–10. <https://doi.org/10.1038/s41467-021-24192-8>.
- Wong, V.N.L., Greene, R.S.B., Dalal, R.C., Murphy, B.W., 2010. Soil carbon dynamics in saline and sodic soils: a review. *Soil Use Manag.* 26, 2–11. [10.1111/j.1475-2743.2009.00251.x](https://doi.org/10.1111/j.1475-2743.2009.00251.x).
- Xiangsheng, Y., Guosheng, L., Yanyu, Y., 2016. Pedotransfer Functions for Estimating Soil Bulk Density: A Case Study in the Three-River Headwater Region of Qinghai Province, China. *Pedosphere* 26, 362–373. [https://doi.org/10.1016/S1002-0160\(15\)60049-2](https://doi.org/10.1016/S1002-0160(15)60049-2).
- Zhang, K., Dang, H., Zhang, Q., Cheng, X., 2015. Soil carbon dynamics following land-use change varied with temperature and precipitation gradients: Evidence from stable isotopes. *Glob. Chang. Biol.* 21, 2762–2772. <https://doi.org/10.1111/gcb.12886>.
- Zhang, Y., Schaap, M.G., Zha, Y., 2018. A High-Resolution Global Map of Soil Hydraulic Properties Produced by a Hierarchical Parameterization of a Physically Based Water Retention Model. *Water Resour. Res.* 54, 9774–9790. <https://doi.org/10.1029/2018WR023539>.