CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Agrobiology, Food, and Natural Resources

Department of Soil Science and Soil Protection

The Assessment of the Influence of Land-Use and Topography on the Spatial Distribution of Soil Organic Carbon using Digital Soil Mapping

by

Shahin Nozari

DISSERTATION THESIS

Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

Supervisor: Prof. Dr. Ing. Luboš Borůvka

January 2024

‘‘I have here only made a bouquet of other people’s flowers,

having brought nothing of my own but the string that binds them together‘‘

Michel de Montaigne

DECLARATION

I declare that I have elaborated on my dissertation work aimed at “The Assessment of the Influence of Land-Use and Topography on the Spatial Distribution of Soil Organic Carbon using Digital Soil Mapping”. This thesis is submitted in partial fulfillment of the requirements for the defense in the study program Exploitation and Conservation of Natural Resources at the Czech University of Life Sciences Prague, 2024. I declare that the thesis is an original work and has not been submitted anywhere for any degree or professional qualification.

Prague, 2024

In Prague, Date: 09.01.2024, Signature

ACKNOWLEDGMENTS

First and foremost, I would like to thank GOD for providing me the opportunity to continue my education and achieve my goals, and pursue them wholeheartedly.

I would like to express my deepest thanks to my mentor, the wise, fairly-minded, and honest, Prof. Dr. Luboš Borůvka, head of the Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food, and Natural Resources, Czech University of Life Sciences Prague in the Czech Republic, for his guidance, helpful encouragement, and continuous support during the practical stages of the experiments and dissertation writing.

Similarly, I wish to express my special gratitude and appreciation to Dr. Mohammad Reza Pahlavan Rad (Iran) for his sincere help, diligent effect, valuable ideas, and remarks during the research work.

I am grateful to my beloved parents and my kind fiance for their help and for supporting me during all stages of the research work, they have stood with me when I did not think that I could stand.

Lastly, I thank also all the colleagues and technicians of the Department of Soil Science and Soil Protection, Czech University of Life Sciences Prague, for their assistance and friendship.

ABSTRACT

This Ph.D. thesis investigates the intersection of digital soil mapping (DSM) with geographical information systems (GIS) and spatial statistics. Utilizing digitized data from legacy soil survey maps, the research investigates soil organic carbon (SOC) distribution in the Czech Republic, with a focus on the Liberec and Domažlice districts. Basically, this study aims to understand factors influencing SOC spatial distribution across various land-uses and topographic variables. It contributes to advancing DSM knowledge, refining local maps for Liberec and Domažlice regions, and improving the comprehension of SOC spatial patterns. The research also presents a spatial prediction structure for SOC using DSM and compares machine learning (ML) models.

Key objectives of this study included investigating the impact of ML model choice on SOC prediction accuracy, emphasizing the importance of training data quality, and indicating that the suitability of covariates such as Normalized Difference Vegetation Index (NDVI) and land cover is comparable to digital elevation models (DEMs). This study facilitates improvements in ML model selection and training data progress in DSM, fostering enhancements in local maps for areas with similar climates.

To achieve these objectives, in this study, the correlation between slope, elevation, and clay on SOC content in different land-uses was explored. Subsequently, the effects of region and elevation on predictors in SOC prediction models were evaluated. Following that, as the primary objective of this research, DSM and ML methods, including Random Forest (RF), Quantile Random Forest (QRF), and Cubist, were utilized to generate SOC maps at a 100 m spatial resolution, enabling the prediction of SOC distribution in the aforementioned districts in the Czech Republic. Additionally, infrared spectroscopy was used to assess the effects of land-use on soil organic matter (SOM) quality.

Analyzing the relationship between slope, elevation, and clay with SOC across different land-uses, the study found that the correlation between SOC and clay lacked statistical significance, despite both increasing with slope. Altitude positively correlated with SOC, though the correlation was only moderately strong. Multiple regression models revealed stronger SOC correlations in the Domažlice district, indicating a significant impact of land-use variations on SOC distribution. In summary, the integration of slope, elevation, and multiple regression analysis enhanced SOC predictions.

Examining the impacts of region and elevation on SOC prediction models revealed that edaphic series and soil classes are robust predictors at lower altitudes, while higher altitudes highlighted the significance of topography-related predictors. The meticulous selection of influential covariates was recognized as pivotal in DSM. Optimal prediction outcomes were noted in smaller, but consistent regions, such as specific natural forest areas, albeit with acknowledged model failures in certain natural forest areas. The study identified model limitations related to uncertainties arising from data harmonization, transformation, and standardization.

Finally, this research challenges the common reliance on DEMs by highlighting the equally crucial roles of other covariates, such as the NDVI and land cover. The results also showed that RF model consistently outperforms QRF and Cubist models in both districts, demonstrating better accuracy in predicting SOC. The RF model, leveraging diverse terrain covariates, exhibited greater R^2 values, smaller RMSE and MAE values, and a more uniform uncertainty distribution. Key predictors for SOC included land cover (vegetation) and elevation, with forest-dominated areas displaying the highest predicted SOC content. Notably, coniferous forests showed the greatest uncertainty in SOC predictions, likely due to a limited number of samples and increased covariate variability.

The results obtained from infrared spectroscopy highlighted the significant influence of land-use on SOM composition. Forest soils showed more pronounced variations, with the upper layer showing intensified aliphatic bands (3010–2800/cm) and higher acidity in SOM, evidenced by carboxylic band intensity, compared to grassland and cropland. Notably, grassland had distinct fulvic acids (FAs) compared to other land-uses, and in cropland soils, the aromaticity of humic acids (HAs) increased with depth. These findings underscore the crucial role of land-use in shaping SOM composition, emphasizing the necessity of considering diverse land-use factors for accurate predictions of SOC.

Recommendations for future work include exploring advanced ML approaches, such as deep learning, for improved SOC predictions, and conducting integrated environmental impact analyses to understand the joint effects of climate, vegetation patterns, and topographical factors on SOC distribution. Additionally, in-depth studies on microbial dynamics in SOC, longitudinal soil health assessments, and community-collaborative research initiatives are suggested to enhance understanding of regional land-use practices and environmental transformations.

Keywords: Digital Soil Mapping, Soil Organic Carbon, Machine-Learning, Statistical Analysis, Random Forest, Quantile Random Forest, Cubist, Geographical Information Systems, Diffuse Reflectance Infrared Spectroscopy with Fourier Transformation (DRIFT)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

Soil health and quality are crucial for human survival; however, the escalating population poses threats to soils (Krawczynski et al., 2015; Eleanor et al., 2013). Soil loss from agricultural areas is generally 10 to 40 times faster than the soil formation rate required for food security (Pimentel and Burgess, 2013), leading to observed soil organic carbon (SOC) loss of up to 50% (Parton et al., 2007). Carbon, in its organic and inorganic forms, plays a pivotal role in soil productivity (Victoria et al., 2012). The global SOC pool is estimated to surpass three times the carbon content in earthly vegetation or the atmosphere, making its contribution to climate change and atmospheric conditions significant (Zhu et al., 2018). Therefore, the demand for soil-related information is increasing worldwide, driven by concerns such as sustainable food production, climate change adaptation and mitigation, soil degradation, and land resource management (FAO and Global Soil Partnership, 2016). Despite the rising need for critical soil information for risk assessment and decision-making (Carré et al., 2007), the recognition of soil multifunctionality and decreasing the available resources for soil surveys have occurred. Thus, soil mappers strain after using legacy soil information and minimizing field surveys.

In the Czech Republic, soil-related information demands are partly satisfied with the accessible spatial soil information systems. Soil data of these spatial information systems are linked to point observations, signifying that end-user' particular interests in soil information can be satisfied by regionalizing these points, which can be performed efficiently by geostatistics. These days, evolving with technological advancements, digital soil mapping (DSM) and sampling optimization rely on prospering geostatistical approaches performed by soil mappers (Minasny et al., 2011). In the context of escalating global challenges, including overpopulation, food security, climate change, pollution, degradation, and resource depletion, DSM has emerged as a powerful tool supporting optimal environmental and agricultural management decisions by providing accessible soil information. Pedometrics, a branch of soil science within the pedological discipline that proposes to quantitatively identify, analyze and assess soil variations over space, has been applied to process soil information through the development of DSM (Burrough et al., 1994; Sanchez et al., 2009).

Digital Soil Maps (DSMs) were already in use in the 1970s (Webster and Burrough, 1972a, b); however, advances in remote sensing technology, geographical information systems (GIS), data mining and machine learning (ML) techniques, computer technology, and improved accessibility of spatial data sets have greatly facilitated the production of DSMs since the 2000s (McBratney et al., 2003; Scull et al., 2005; Minasny and McBratney, 2016). Also, technological advances have led to the ability to present DSMs in larger spatial dimensions and more differentiated resolutions (Minasny and McBratney, 2016). Consequently, DSMs have been enhanced in a variety of scales for many applications. At the global level, institutions such as the International Soil Reference and the Information Centre (ISRIC) have recently provided predictions of soil properties such as SOC for six standard depth intervals (Hengl et al., 2014). Also, numerous projects in Europe have used the DSM of the Joint Research Centre of the European Commission to predict soil erosion by wind (Borelli et al., 2014) and water (Panagos et al., 2015), as well as the prediction of total SOC stocks, to test potential climate and land cover changes (Yigini and Panagos, 2016). Additionally, DSM methods have often been used to focus on specific environmental issues at the regional and local levels. For example, to predict the spatial distribution of biological soil crusts to estimate soil resistance in semi-arid environments (Brungard et al., 2015), to monitor seasonal variations in soil salinity to better mitigate the consequences of salinization (Berkal et al., 2014), or to produce crop-specific suitability and contribution margin maps for crops (Harms et al., 2015; Kidd et al., 2015).

DSM's primary goal is to predict quantitative soil properties and visualize spatial variability through high-resolution mapping techniques (Selvaradjou et al., 2007). It involves generating and populating spatial soil information by combining field and laboratory observation methods with spatial and non-spatial soil inference systems. Therefore, DSM surpasses just the digitization of existing soil maps, extending to continuously mapped soil attributes. Alternative terms frequently used in the literature include soil-landscape modeling and predictive soil mapping.

DSM can generally be classified into three approaches: 1) pedotransfer functions, 2) geostatistical approaches, and 3) the State-Factor (Clorpt) approach. Pedotransfer functions entail mathematical functions where other soil properties predict a specific soil property, such as predicting hydraulic properties based on grain size distribution. Geostatistical approaches, introduced by Krige (1951) and Matheron (1963), involve using sampled and analyzed data to interpolate soil property maps. While the main principle of DSM is based on scoring and geostatistical approaches, almost all combinations of methods are possible as empirical quantitative soil-space prediction functions (Mulder et al., 2011).

This thesis investigates DSM approaches and features two case studies utilizing existing soil and auxiliary data to train various ML models. Therefore, SOC maps for two districts in the Czech Republic, Liberec and Domažlice districts, where high-resolution digital soil data are currently limited, were produced. DSM is expected to significantly enhance mapping efficiency, enabling more accurate and quantitative predictions of soil properties at each district that improve environmental and agricultural decision-making.

The study evaluates the relationship between basic soil properties (clay, silt, sand content, soil organic matter [SOM], etc.) and environmental variables of stands (relief parameters, geology, land-use, vegetation type, etc.) at selected districts using legacy soil data (systematic soil surveys, basal monitoring of soils, forest surveys, etc.) as the main source of soil data, supplemented by recent case studies. The study also develops and calibrates spatial prediction models using the binding relationships between soil properties and environmental covariates. Therefore, advanced computational methods and ML models including random forest (RF), quantile random forest (QRF), and cubist were used to evaluate spatial distribution of SOC and its role in soil characterization and ecosystem functioning. Finally, the reliability and uncertainty of the models employed for SOC prediction were evaluated.

Additionally, a study focused on the effect of different land-use on soil organic matter quality assess by means of diffuse reflectance spectroscopy with Fourier transformation is included. Different SOM quality under various land-use types leading to different resistance to decomposition can be one of the reasons for different SOC contents and stocks.

It should be noted that the exploration of SOM distribution in various land-uses sheds light on the complicated relationships between land-use practices and soil health. Additionally, investigating arable land, forests, and complex agricultural systems contributes valuable perspectives on how human activities and natural processes impact SOC content. This comprehensive study bridges precision SOC mapping through DSM with a holistic understanding of diverse land-use patterns, informing sustainable land management strategies.

1.1. Background

The largest terrestrial pool of bound carbon (C), which plays a central role in global C dynamics, is SOC. In this overview, mainly different aspects of SOC will be spatially evaluated concerning specific characteristics, input data, and models for SOC. Machine-learning (ML) is the self-adaptive method where a fitted pattern can then be used to set prediction targets for new data. Notwithstanding the evolving number of ML algorithms that have been extended, comparatively few studies have presented a comparison of some different learners; typically, model comparison studies are limited to comparing a few models. SOC is attracting increasing attention within both scientific and political platforms due to its influence on atmospheric CO_2 accumulation, which affects climate change (Selvaradjou et al., 2007). DSM could be an essential tool to support the understanding of the dynamics of SOC for both forest and agricultural soils. It increases the efficiency of the mapping process and allows a more detailed, accurate, and quantitative prediction of soil properties for different sites (Minasny et al., 2013). Also, more advanced approaches, such as soil anisotropy characteristics, and advances in computational methods, such as 3D DSM, have been developed that indicate both horizontal and vertical variability of soil properties (Minasny et al., 2013). Furthermore, in the face of ever-increasing global problems such as overpopulation, food security issues, climate change, environmental pollution and degradation, and depletion of natural resources, the DSM has become a powerful tool for optimal decision-making in environmental and agricultural management by providing relevant soil information (McBratney et al., 2003). Many researchers have studied the dynamics of SOC through DSM (e.g. Guevara et al., 2018), including several reviews (e.g. Minasny and McBratney, 2013). Unfortunately, detailed aspects (i.e., specific features, input data used, and models for spatial prediction) regarding DSM have not been logically compiled for SOC in forest and agricultural soils. Therefore, as an effort of this review, this will be a basic platform for such.

1.2. Research Hypotheses

SOC is related to the landscape position, and reliable models can describe this relationship and the effect of other environmental factors. SOC content and stocks can be spatially predicted using prediction models with reasonable accuracy.

1.3. Research Objectives

Based on the above-mentioned scientific hypotheses, the Ph.D. thesis has the following aims:

- Analysis of the relationship between SOC/SOM and environmental covariates.
- Identification of the most important environmental covariates for SOC prediction.
- Comparison of various prediction models and selection of the best model.
- Production of maps of soil properties (SOC) and related uncertainty for selected regions.

1.4. Conventional Soil Surveys and Their Limitations

In the primary mapping of soils, the applications are mostly concentrated on the agricultural use of soils (Soil Survey Staff, 1993). However, traditional mapping is still the largest source of information on soil; the development of new techniques used in this field brings a new and improved understanding of the spatial adaptation of soil and uses it for many other applications. Over the last 30 years, cartographers have sought to objectively measure, classify, and study the variability of soil cover to expand quantitative models (McBratney et al., 2000). Conventional soil surveys are based on landscape equations or concepts and happen to be one of the most important sources of spatial soil information in relation to terrain properties (Hudson, 1992). In a conventional soil survey, the soil mapper first delimits the area by the process of soil perception to create an environmental soil-landscape model. In extracting soil information from a given area, the soil-landscape model encapsulates the relationship between the soils at the site and the different land positions or units. The soil mapper primarily uses preconceived assumptions about what types of soils to assume in an area based on the available information about soil-environment relationships. The soil mapper later takes aerial photographs to identify patterns where soil-environment variables are an external expression on the landscape to correlate landscape features with soil boundaries.

Areas with similar soil-environmental characteristics share similar soil characteristics based on a kind of rule-based reasoning (Abraham, 2005). The soil expert manually draws the spatial extensions of different soils or soil combinations into the map by photo image analysis. The output results of the soil units are then demonstrated using polygons. Finally, the individual areas on the maps are then related to map units (Lark and Beckett, 1998) and are each formed with a polygon representing the spatial arrangement. The conventional process involves finding different soil-productive environments by visually interpreting geological maps, aerial photographs, and topographic maps. An attempt is made to map the unit as far as possible to a classification unit, which is then used in the map legend. If a particular landscape is included in a map or a classification unit, it is said to be a typical representative. Therefore, the polygonal approach often restricts an accurate description of land cover (Zhu, 2000) and reduces the ability to capture continuous changes in soil attributes.

Soil data are a necessary part of natural resource modeling. For example, a soil data layer is an essential source of information for modelers who simulate pollution potentials from agricultural areas without point sources. Some models, such as the Environmental Policy Integrated Climate model (EPIC), integrate soil, climate, economics, management, and other variables to simulate the effects of farming systems on the environment and productivity (Izaurralde et al. 2006). Nowadays, data from soil studies are used to estimate the potential for carbon sequestration and other soil conditions in the context of global change. A key component of the USDA Global Change research and development program is the role of soil in communicating the effects of agriculture and forestry on the global atmospheric composition of greenhouse gases. The polygon-based mapping practice is based on the discrete conceptual model (Zhu, 1997a), which limits the ability of the soil mapper to produce accurate soil maps. Several interests arise in standard soil surveys. First, the data are displayed as separate categories, with the requirements in the polygons being homogeneous (Hole, 1978; Zhu and Band, 1994). A remarkable amount of spatial generalization within the map unit occurs due to formations of subdominant soils that are too small to be determined at the spatial scale of the map (Hole and Campbell, 1985).

Consequently, the purity of the mapping units depends on the complexity of the terrain, the external expression of the boundaries, the surveying effort, and the mapping scale (Beckett, 1971). However, an increase in the proportion of "pure" mapping units has been envisaged to improve the exponential increase in the costs of developing the soil map (Bie et al., 1973). Additional controls may be associated with the lack of accuracy or imprecision of the boundaries of the map units when the variability (or lack thereof) of the topographic surface does not correspond in principle to the variability that may occur below ground (Hole, 1978). Furthermore, the soil changes are not unambiguous but rather they are fuzzy where the soil properties between two neighboring map units may be an intergrade of the soil attributes of the two units (Zhu and Band, 1994; Schaetzl and Anderson, 2005). The ultimate difficulty for conventional surveys stems from the soil experts themselves, where the description of the map units is based on mental models of soil-environment relationships that are rarely ever published (Thompson et al., 2012; Dewitte et al., 2013) (Figure 1.1). In contrast to these concerns, conventional soil maps have great value as a training data source for ML tools, where soil map units with several environmental variables is used to predict soils at other locations (Bui, 2003). These suggestions have been illustrated many times using decision trees (Bui and Moran, 2001, 2003; Grinand et al., 2008) and the algorithms of the random forest (RF) (Häring et al., 2012); and more recently, decision trees have also been applied to disaggregate complex map units (Odgers et al., 2014). Unfortunately, existing soil databases are neither complete nor accurate enough to support a comprehensive and credible use of soil information within the geospatial data infrastructure that is being developed worldwide. The main reason for this lack of spatial soil data is simply that conventional soil survey methods are relatively slow and expensive. Besides, we have also seen a global reduction in funding for soil research that began in the 1980s (Hartemink and McBratney 2008), resulting in a significant reduction in large-scale soil geospatial data collection and/or conventional soil surveying.

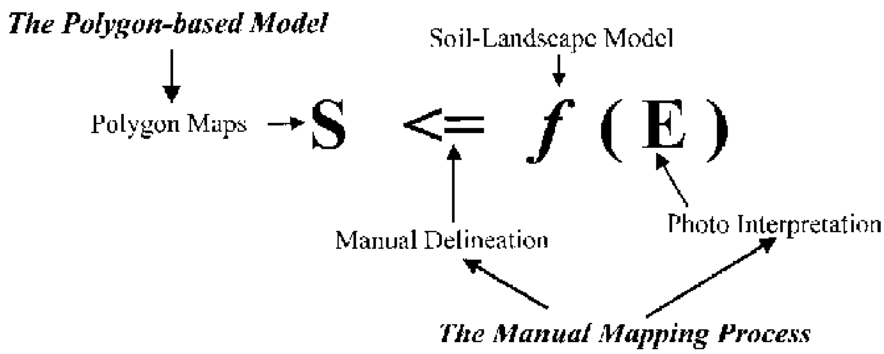


Figure 1. 1. Conventional soil mapping and its limiting factors (Zhu et al., 2001)

CHAPTER 2

Literature Review

This chapter provides a comprehensive literature review on soil organic carbon (SOC), digital soil mapping (DSM), and various DSM predictors, including soils, climate, organisms, relief, and parent materials, as well as ML models and geostatistical methods.

2.1. Soil Organic Carbon (SOC)

Soils are the essence of the Earth's "critical zone," the thin exterior veneer between the top of the tree canopy and the bottom of the groundwater aquifer, that human beings rely on. Soils develop by continually changing over the years and change at various rates. Along several pathways, as mineral material is released from the rock, decay is colonized by plants and soil biota. This colonization leads to the formation of soil organic matter (SOM) and soil structure and influences the carbon, nutrient, and water cycles. Soil carbon exists in two classes: organic and inorganic structures (Victoria et al., 2012). Meeting the food needs of an ever-growing world population can be achieved through the sustainable management of soil resources (Selvaradjou et al., 2007). The sustainable connotation of the SOM can be defined as follows: The SOM in the soil meets the needs and aspirations of food products in the present, which is coordinated with the development of population growth and resource use, without compromising the ability of future generations to meet their own food needs (Song et al., 2017).

Many of the processes that influenced the SOM in the last century were dominated by human management of vegetation, which in turn influenced the inputs and status of the SOM. Changes in vegetation cover, including those that occur in response to climate and land-use or management, influence the SOM by altering the rates, quality, and location of plant litter inputs to the soil. Every year, about 10 million ha of arable land is lost due to soil erosion, reducing the amount of land available for global food production (Faeth and Crosson, 1994). The loss of arable land is a serious problem, as the World Health Organization and Food and Agriculture Organization report that two-thirds of the world's population is undernourished (Pimentel and Burgess, 2013). There is a general pattern of soil carbon loss after the start of cultivation over

30 to 50 years with a loss of up to 50% of soil carbon (Parton et al., 2007). In 2016, it was estimated that more than 80% of all EU regions were affected by moderate to severe soil erosion in agricultural and natural grassland areas (European Commission - JRC).

Furthermore, SOC has been recognized as a key factor in soil fertility and environmental management. Globally, the pool of SOC has a significant impact on CO_2 concentration, which affects the rate of climate change and the state of the atmosphere (Zhu et al., 2018). SOC is one of the most important soil properties, and any change in its content and composition affects the physical, chemical, and biological properties of the soil. Improving SOC improves soil structure, increases the water and nutrient content of the soil, and reduces soil erosion and degradation so that higher plant productivity and better water quality can be expected in watersheds. Climatic, topographic, and management factors influence the content of SOC. At the local level, climatic factors do not play a major role in the quantity of SOC, while topographical factors are more important for the quantity and variability of SOC (Moghimi et al., 2015). SOC is the main component of SOM, which is formed by the biological, chemical, and physical decay of organic material entering the soil system from above-ground (e.g. leaf fall, crop residues, animal waste, and remains) and underground sources. The elemental composition of SOM varies, but the average values are about 50 percent carbon, 40 percent oxygen, and 3 percent nitrogen, with much lower amounts of phosphorus, potassium, calcium, magnesium, and micronutrients (Victoria et al., 2012). Land-use changes, especially the transformation into agricultural ecosystems, deplete the soil's carbon stocks. Therefore, degraded agricultural soils have a lower SOC stock than their potential capacity. Consequently, proper agricultural management, afforestation of agricultural soils, and management of forest plants can increase the stock of SOC through C sequestration (Lal, 2005). There is an interest in quantifying the capacity of different soil types and land management practices to support the increase in SOM and to understand how these changes will affect soil health, ecosystem services, and carbon sequestration in the medium and long term (Sarmadian et al., 2014).

2.2. Digital Soil Mapping

Soil spatial mapping is required for many environmental modeling and land management purposes. It involves the interpretation of the spatial matter and its properties. There has been progress in approaches to soil mapping and soil relationship to other natural environmental factors, and these approaches have been consistent concerning the development of techniques and activities (Němeček and Tomášek, 1983). Soil mapping is also governed by the requirements for which applications are to be obtained and the materials used. Soil mapping assistance in characterizing soil resources has introduced the DSM. The DSM approach is more appropriate for this purpose. The recent improvement in soil mapping has experienced a total change by amended technicality and spatial science knowledge. In the early mapping of soils, but not only them, these applications are primarily focused on the agricultural use of soils (Soil Survey Staff, 1993). Although traditional mapping is still the largest source of information on land, the development of new techniques applied in this field brings a unique and enhanced understanding of land's spatial arrangement and uses this knowledge in many other applications. DSM is a computer-aided procedure with an output in the form of digital maps of soil types and soil properties. There are usually geospatial and geostatistical tools that integrate information from soil attributes with data containing interdependent environmental parameters and remote sensing images (Dobos et al., 2006). This methodology involves the use of tools and approaches from a wide range of scientific environments, such as *scorpan* models, GIS, remote and proximal sensing, and computer programming, to place the spatial distribution of soils in a quantitative framework (McKenzie and Ryan, 1999; McBratney et al., 2003).

DSM has changed how soil resource assessment is approached around the world. New quantitative DSM products are appearing weekly with the associated uncertainty. Many methods and approaches have been expanded. We can map the whole world or a farmer's field. All this has been done since the turn of the millennium. DSM has evolved from a science-driven research phase in the early 1990s to a fully functional and operational process for spatial soil assessment and measurement. The increasing scale of DSM projects shows this development from small research areas to regional, national, and even continental and global parts. Recent advances in information technology and computational efficiency have been key factors in improving the DSM. These advances have motivated many inventions worldwide to create spatial databases to facilitate the collection, maintenance, dissemination, and use of spatial data. Green (1992) claimed that the combination of remote sensing within a GIS database reduced costs and time and increased the detailed information collected for soil investigation. The use of the digital elevation model (DEM) has been vital in extracting terrain attributes in terrain characterization (Dobos et al., 2000). The stages of this process are shown in Figure 2.1.

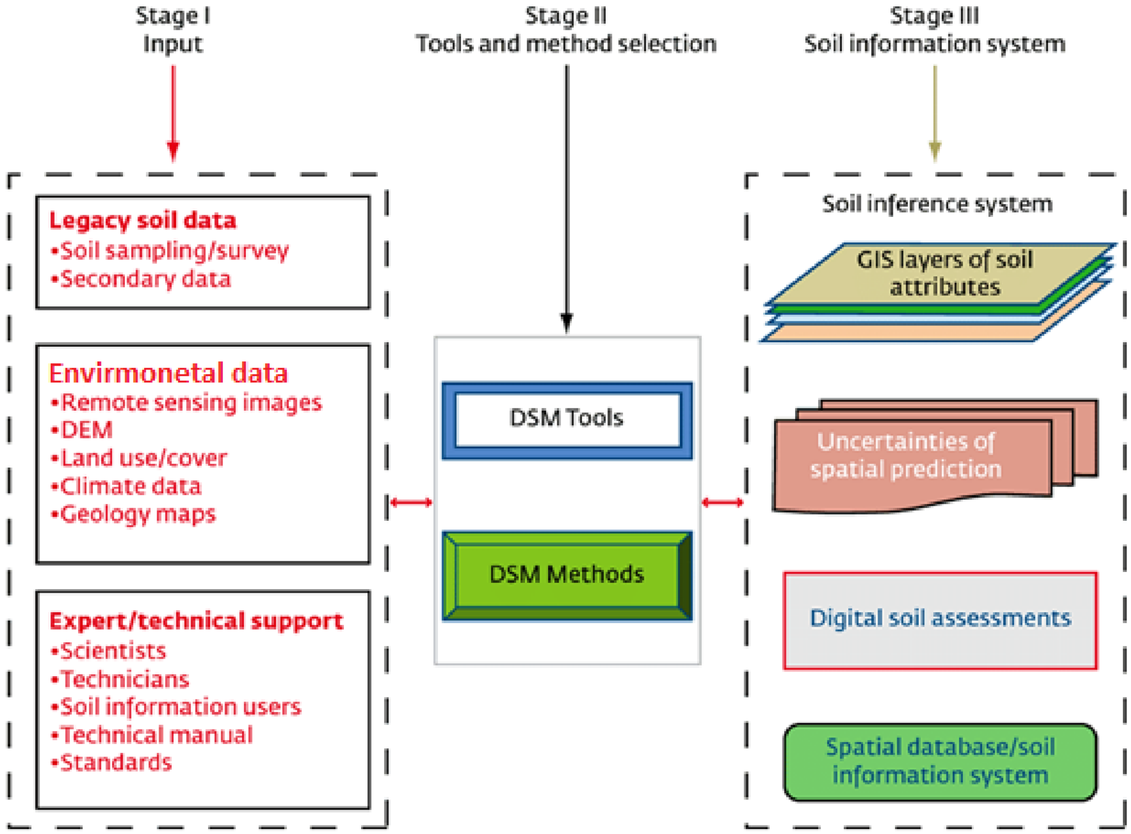


Figure 2. 1. The DSM approach (Omuto et al., 2012)

This approach has overcome some serious limitations of conventional soil surveys such as grid marking and morphological description of soil variability. The technique emphasizes the soil continuum, in which soil attributes at a given location also depend on its geographical location and the soil properties in adjacent areas, and then overcomes the limitations and coarseness of using large polygons as a means of explaining soil variability in the landscape, both geographically and in terms of attributes. DSM significantly develops the mapping process's proficiency and allows for a more accurate and quantitative prediction of soil properties at any location. Moreover, thoroughly presenting the heterogeneous relationship and quantifying the variation is not an easy task. DSM has become a powerful approach in assisting optimal environmental and agricultural management decisions.

Almost all combinations of methods are possible as experimental quantitative soil-spatial prediction functions, while the main tool of DSM is *scorpan* and ML methods. The auxiliary data, named environmental covariates, can be received from digital elevation models (DEM), remote sensing data (satellite or airborne images), proximal soil sensing, geology maps, geomorphology, and vegetation characteristics. Pedometric techniques are used in DSM, which predicts the spatial distribution of soil types and soil properties. The DSM models are divided into simple, intermediate, and complex models depending on their interpretability and the number of parameters required. Also, Brungard et al. (2015) illustrated that covariates selected by soil scientists familiar with the study area did not yield the most accurate models compared to covariates automatically selected by machine learning algorithms (Minasny and McBratney, 2010).

In contrast to the extended employment of Jenny's (1941) *clorpt* model, it is still essentially a conceptual model. With the growing possibilities of GIS software, coupled with geospatial data in digital format, the extensively employed *clorpt* model becomes unsuitable for modeling soils as a spatial phenomenon. McBratney et al. (2003) recognized the importance of a spatial component in soil formation theory and proposed the *scorpan* model. The *scorpan* model contains the five factors from Jenny's (1941) *clorpt* model, where cl (or c in the scorpan model) stands for climate, o for organisms, r for relief, p for the parent material at the spatial position (x,y), and the time as environmental covariate (t) is replaced by the age of the soil (a) in the *scorpan* model. Additionally, other soil information (s) is included as a predictor in the *scorpan* model. The DSM approach follows thus the general spatial prediction function *scorpan*:

$$S = f(s; c; o; r; p; a; n) + e$$

where S, which is the soil type or soil property to be predicted (for example SOC), is a function of soil (s), climate (c), organisms (o), relief (r), parent material (p), age (a), and spatial position (n); and where e is the error. These factors could be characterized by Landsat or other remote sensing spectral data or a digital elevation model. The topography has the potential to explain much of the variation of SOC. Terrain attributes, the most commonly used environmental predictors in the DSM, approximate water, material, and sediment flows through the landscape (McBratney et al., 2003). The subsequent sections present a brief overview of the soil-environment covariates used in the DSM literature.

2.3. DSM Predictors

2.3.1. Soils (S)

The secondary common *scorpan* factor used second most often was the soil information "s", which was applied by almost 40% of the reviewed studies in McBratney et al. (2003). Conventional soil survey data are generally used to train models or to build knowledge databases. Qi and Zhu (2003) noted that the soil-landscape relationships could be exploited and accepted for soil mapping and analysis. Hewitt (1993) mentioned that the rules for soil mapping, mental models, and soil-landscape relationships used independently by soil experts are generally not recorded. The extraction of scientific knowledge from old (legacy) data sources can help in cases where there are no specialists or soil-landscape relationships that are not recorded. Bui et al. (1999), Qi and Zhu (2003), Moran and Bui (2002), and Grinand et al. (2008) showed that soil-landscape relationships were extracted using classification trees and various other ML algorithms. Also, Qi and Zhu (2003) presented a technique for obtaining point data from soil survey polygons. In Lagacherie et al. (2006), a reference (or training) area approach with a small range was adopted to identify and formulate the soil pattern rules from which the soil survey was extended. These rule sets could then be applied to extrapolation targets. The s-factor involves the application of remote and proximal sensing data and the extraction of soil information from conventional soil maps. Soil samples are prepared from the field and brought to a laboratory to define the relationships between the properties of soil. Mulder et al. (2011) reported that remote sensing data are particularly valuable for mapping soil mineralogy, soil texture, soil moisture, organic carbon, iron and carbonate content, and salinity in bare soil.

2.3.2. Climate (C)

The most limited application of the environmental covariates of *scorpan* is climate (C). The local climate is mainly influenced by topography. Thus, topographic indices such as altitude and aspect can be used as a proxy for climate variables, since there is a correlation between altitude and the environmental lapse rate and a correlation between slope direction and temperature (Schaetzl and Anderson, 2005). Climate standard covariates include mean annual temperature, precipitation, and evapotranspiration, which can be obtained from satellite images (McBratney et al., 2003). The advantage of these covariates depends mainly on the size of the study object, whereby constant climatic conditions can be assumed if the study area becomes smaller. This prediction is always encumbered by some degree of uncertainty. *Scorpan* model turns out to be preferred due to its ability to predict different soil properties in the reference position compared to other forms of models. Measured soil attributes and auxiliary data are required to predict the soil units. The data should be correlated to monitor soil characteristics in the area under investigation. This correlation must exhibit spatial dependence with additional data. Using the *scorpan* model to predict classification classes is thus based on obtained dependency rules from the training set used for the site. The success of the model, whether it is a prediction of soil properties or units, depends on four conditions (Rokach, 2010):

- A sufficient number of additional data (in terms of the number of variables and the number of sampled points);
- Sufficient amount of data on the soil;
- The existence of a function that can describe the relationship between the soil and the additional data;
- A good correlation between the soil (or its properties) and the environment.

Geo-referenced soil observations are linked to environmental variables from the input data, showing the process of DSM in Figure 2.2.

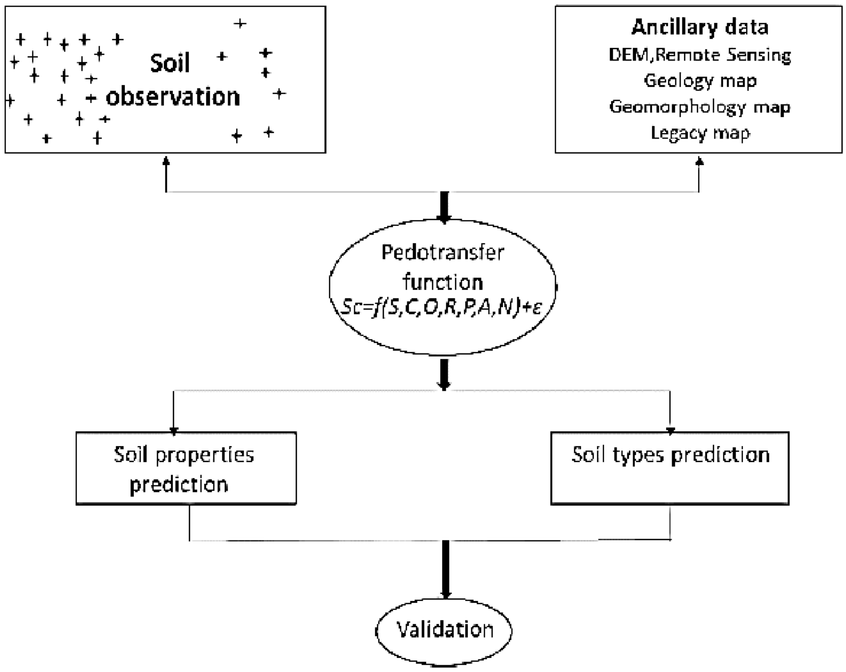


Figure 2. 2. Workflow of Digital Soil Mapping (Zeraatpishe, 2017)

2.3.3. Organisms (O)

Soil-environment layers representing organisms (o) were applied in 30% of the reviewed studies (McBratney et al., 2003). An important source of vegetation data can be satellite images, where vegetative indices have been established based on satellite band conditions (Mulder et al., 2011). The Normalized Difference Vegetation Index (NDVI) is such a case and is useful as a covariate in mapping SOC (Boettinger, 2010; Marchetti et al., 2010; Zhao and Shi, 2010). Other similar indicators extracted from the NDVI are the Soil Adjusted Vegetation Index (SAVI), the Transformed SAVI (TSAVI), the Modified SAVI (MSAVI), and the Global

Environmental Monitoring Index (GEMI) (Mulder et al., 2011). Remote sensing data could also be used to assess other vegetative properties such as the leaf area index (LAI), fractional canopy cover, the water content of plants, aboveground biomass, evapotranspiration, and vegetation height (Dorigo et al., 2007). The use of some of these vegetative indices has yet to be investigated in the DSM. Crop yields are the result of many factors such as soil-atmosphere interaction and plant's genetics, health, and susceptibility to pests and diseases. Therefore, crop data were also used as covariates for spatial prediction. Crop yield data (yield maps) can also be used as signs of soil properties, considering that plant growth is influenced by characteristics such as clay content, moisture content, and nutrient content (e.g. McBratney et al., 2000). Forest variables such as area, total gross volume, stand density, stand height, and aboveground biomass could provide insight into soil properties. Especially with the improvements in the LiDAR images, further forest inventory data could become available in the future (Woods et al., 2011; Treitz et al., 2012). Similarly, land and vegetation class data could be valuable data reference for the above-mentioned purposes.

2.3.4. Relief (R)

Following 132 papers reviewed by McBratney et al. (2003) on DSM, the "relief" (r) was considered the most commonly used factor in DSM studies of all seven *scorpan* aspects, as almost 80% of the reviews used digital elevation models (DEM) and other terrain derivatives calculated from them. Using of relief attributes and the derivation of the relationship between soil and relief are also of particular importance. The mapper's understanding is applied via the soil-landscape relationships to produce soil type maps in classical soil mapping (Hudson, 1992). The terrain models presented by Brough (1986), referred to as DEM, are an electronic model of the Earth's surface that can be stored and manipulated in a computer to provide many classes of data that can assist the soil surveyor in mapping and quantitatively describing landforms and soil variability. They can produce maps of slopes, aspects, slope gradient, and drainage network in catchments (Brough, 1986; Brabyn, 1997; Gallant, 2000). DEMs are particularly valuable because they are readily accessible and uniform in coverage. Some information, such as elevation, slope, and aspect maps, can be used in conjunction with photographs to improve their ability to study soil, as used by Dobos et al. (2000). Moreover, it has sometimes been demonstrated that a landscape classification can be produced promptly if only one DEM is used (MacMillan et al., 2000, 2003) and if soil attributes, particularly organic matter, are often associated with landform elements (Pennock et al., 1987). Furthermore, landform classes have been applied as environmental properties in DSM (Smith et al., 2010) and predictive ecosystem mapping (MacMillan et al., 2007). More recently, archived data on soil mapping are often available in sufficient quantities and at low cost (Green, 1992). The integration of remote sensing into a GIS database can reduce costs, shorten the time, and expand the detailed information prepared for soil investigation. Specifically, the use of the Digital Elevation Model (DEM) is important for deriving landscape properties that are used in the characterization of landforms (Brough, 1986; Dobos et al., 2000).

In recent times, DEMs and digital terrain data have also been applied for soil mapping and soil-specific predictions (Aksoy et al., 2009, Moore et al., Gessler et al., 1995, Dobos et al., 2000, 2001). The terrain model (DTM) is used in a non-specific sense for digital terrain and/or bathymetric (underwater equivalent of topography) data in all different digital forms, including mass points, fault lines, triangulated irregular networks (TINs), terrains, and DTMs. DEM usually mentions x/y coordinates and z values of the exposed, vegetation-less, and artificially highlighted earth landscape. The computerized elevation indicates that the information from DEM is used as part of the geographic surveys to talk about a part of the earth's surface. Terrain attributes such as slope or aspect are derived from elevation grids and tend to amplify systematic errors caused by data resolution and the DEM technique (Bolstad and Stowe, 1994; McKenzie et al., 2000). Predictions of environmental models based on a combination of DEM-derived areas contain an uncertainty component due to the unknown accuracy of the original elevation data. It is important to quantify this uncertainty, which has been adopted from DEM, and to consider its implications for the interpretation and use of model predictions. In a preliminary study, McSweeney et al. (1994) developed techniques for soil-terrain modeling, as a quantitative method for predicting soil variability patterns using observed patterns in environmental variables known to influence the variability of soil properties, such as topography and hydrological geology. Bell et al. (1992, 1994a) predicted and mapped the soil drainage classes using topographic information derived from a DEM, perennial stream, and ephemeral surface drainage pathways and geology. This quantitative mapping using environmental correlation involves the double development of prediction models for the study site. In routine soil studies, due to the complexity and wide range of spatial scales and soil formation processes operating on them, the development of models to predict spatial, quantitative, mechanical, and mathematical performance is an almost impossible task, but considerable effort has been made (Dietrich et al., 1995). However, a simplifying hypothesis is necessary, and approximate local models of pedogenesis with different levels of empirical evidence must be used. Soil-terrain modeling has also been used to model the spatial distribution of specific soil properties, including A-horizon thickness, organic matter content, extractable P, pH and sand and silt content (Moore et al., 1993), A-horizon thickness, and depth up to carbonates (Bell et al., 1994b) and A-horizon thickness and solum depth (Gessler et al., 1995).

Table 1. 1. Some primary terrain attributes (Gallant and Wilson, 2000)

Attribute	Definition	Significance
Altitude	Elevation	Climate, vegetation, potential energy
Upslope height	Mean height of the upslope area	Potential energy
Aspect	Slope azimuth	Solar insolation, evapotranspiration, flora and fauna distribution, and abundance
Slope	Gradient	Overland and subsurface flow velocity and runoff rate, precipitation, vegetation, geomorphology, soil water content, land capability class
Upslope slope	Mean slope of upslope area	Runoff velocity
Dispersal slope	Mean slope of the dispersal area	Rate of soil drainage
Catchment slope	Average slope over the catchment	Time of concentration
Upslope area	Catchment area above a short length of the contour	Runoff volume, steady-state runoff rate
Dispersal area	Area downslope from a short length of a contour	Soil drainage rate
Catchment area	Area draining to the catchment outlet	Runoff volume
Specific catchment area	Upslope area per unit width of the contour	Runoff volume, steady-rate runoff rate, soil-water content, geomorphology.
Flow path length	The maximum distance of water flow to a point in the catchment	Erosion rates, sediment yield, time of concentration
Upslope length	Mean length of flow paths to a point in the catchment	Flow acceleration, erosion rates
Dispersal length	Distance from one point in the catchment to the outlet	The impedance of soil drainage
Catchment length	Distance from the highest point to outlet	Overland flow attenuation
Profile curvature	Slope profile curvature	Flow acceleration, erosion/deposition rate, geomorphology
Plan curvature	Contour curvature	Converging/diverging flow, soil-water content, soil characteristics
Tangential curvature	Plan curvature multiplied by the slope	Provides an alternative measure of local flow convergence and divergence
Elevation percentile	The proportion of cells in a user-defined circle lower than the center	Relative landscape position, flora and fauna distribution, and abundance

2.3.5. Parent Material (P)

As far as source parent material (p) is concerned, a total of 25% of the studies examined by McBratney et al. (2003) contained a parent material layer; in 75% of the cases where a parent material map was used, geological maps were used, preferably as surface material maps. As a result, the transported parent materials are poorly illustrated and the parent material maps used for DSM are distorted in favor of residual parent materials (Smith et al., 2010). In addition, geological maps suffer from the same obstacles as conventional soil maps, which sometimes use complicated map units. Consequently, parent materials have sometimes been mapped as soil properties rather than environmental covariates for predicting soils (e.g. Bui and Moran, 2001; Lacoste et al., 2011; Lemercier et al., 2012). The existing soil survey identifies 6 different soil groups with 8 mineral parent material classes in the Liberec district and 5 distinct soil groups with 8 mineral parent material classes in the Domažlice district. Although the distribution of the organic parent materials is influenced by the topography, these parent materials are also strongly dependent on climatic and vegetative factors, which were considered in this study as well.

2.4. Machine-Learning Techniques

This sub-section aims to provide a brief introduction and a summarized overview of Random Forest (RF), one of the most renowned machine learning techniques used in mapping SOC, highlighting its relevance in DSM. RF learner is conceptually similar to tree-based learners and shares the same advantages; however, several decision trees are trained, and the results are based on predictions from an ensemble of individual trees (Breiman, 2001). For the RF learner, each tree is trained from a randomized bootstrap sample of the entire training set, and a subset of predictors used for the node partitioning rules is also randomly selected. However, the RF learner has been used early on for the analysis of large datasets in the bioinformatics literature (e.g. Díaz-Uriarte and Alvarez de Andrés, 2006; Qi, 2012; and Svetnik et al., 2003), DSM usage appears to become increasingly more prominent. Among the DSM applications of the RF learner were, similar to the decision trees, mapping of soil organic carbon (e.g. Grimm et al., 2008; Guo et al., 2015; Wiesmeier et al., 2011), mapping of soil texture (Ließ et al., 2012), as well as for classification purposes such as mapping of soil parent material classes (Heung et al., 2014) or updating and disaggregating conventional soil overview maps (Häring et al., 2012; and Rad et al., 2014).

Despite the similarities between single-tree-based learners and RF, few studies in DSM have compared the two, except Ließ et al. (2012), who compared them to predict particle size fractions using regression and found that RF scored better. Many prediction techniques have been extended by the DSM framework proposed by McBratney et al. (2003) to correlate auxiliary parameters and SOC. Minasny et al. (2013) present a comprehensive overview of the modeling of SOC. Generally, multiple and linear regression were used to describe the relationship between SOC and auxiliary variables. Some studies used generalized linear models, regression tree models, random forest, artificial neural networks, support vector machine regression, k-nearest neighbor, or genetic programming to establish the relationships between SOC content and other variables. However, such modeling methods can find non-linear relationships, which is more powerful for the digital mapping of SOC.

Unfortunately, the main disadvantage of machine learning methods is that they map the spatial variability of SOC at certain depths, whereas SOC usually varies continuously in a typical soil profile. Soil carbon content decreases rapidly with depth. Accordingly, to describe the vertical and lateral variation of SOC the variation can be modeled using the current soil depth to create a 3D map. Numerous efforts have been made to derive some functions of soil variation with depth. Although Bishop et al. (1999) indicated that equal-area square splines are more flexible and practical in-depth functions than other methods. Concerning the potential of soil depth functions and the ability of digital soil mapping, a combination of both methods seems to be the only way to predict the lateral and vertical variation of soil properties. Some researchers used splines to model the vertical distribution of SOC in the soil profiles and predicted SOC on a landscape scale using data mining tools and environmental variables as predictors. Mishra et al. (2009) applied a geographically weighted regression to map the SOC pool on a regional scale in the midwestern United States. They calculated the SOC pool in each soil horizon. Lorenzetti et al. (2014) compared 1:5,000,000 maps of Italian soil regions and digital soil maps with a grid resolution of 1 km to predict the world reference base (WRB) reference soil groups. They showed that the digital soil map has higher accuracy compared to conventional maps.

2.5. Statistical and Geostatistical Methods

Geostatistics is a branch of applied statistics (Goovaerts, 1999) that describes regular changes in natural objects, including soil. The usage of geostatistics in soil science certifies a quantitative description of soil spatial variation, improves the accuracy of estimating soil properties for data mixing and mapping, and forms the basis of a rational soil sampling design.

Theoretically, statistical models are limited to a large geographical area in a limited number of companies' field measurement data and estimate the concentration level of variables in the whole study area. The utilization of geostatistical interpolation models for spatial data on soil properties is significant for accurate agricultural soil management. The relationships between soil and environmental conditions are correlated through the fitting of a model using machine learning and/or geostatistical techniques, where the soil-environmental relationships are then used to predict the soil properties for areas that have not been sampled. (Hengl et al., 2014, 2015; McBratney et al., 2003).

Geostatistical models are also beneficial when the obligation of random distribution is not realistic because they allow eventual spatial patterns to be shown in the existing data (Peterson et al., 2010). Some geospatial approaches have been used for spatial prediction of soil organic carbon. These methods can be divided into three parts:

1. Techniques that use environmental correlation between the soil organic carbon and the environmental parameters (Martin et al., 2011),
2. Methods that use the spatial autocorrelations in the soil organic carbon observations approaches (Mishra et al., 2009), and
3. Hybrid approaches that use both environmental correlation and spatial autocorrelation (Martin et al., 2014).

Classical statistical or hybrid statistical methods without machine learning (ML) models (multiple linear regression (MLR), ordinary kriging (OK), and regression kriging (RK) exhibited worse prediction accuracy compared to the models that included ML (Tziachris et al., 2020). Geostatistical simulations based on multiple-point statistics can be considered as an advanced geostatistical approach. Geological structures can be accurately generated using object-based simulations. However, conditioning in these methods requires soft and good data to be compactly calculated. One drawback of pixel-based simulation methods is that they are based on variograms that show two-point statistics and therefore cannot produce complex and realistic geological structures. Therefore, models generated using these methods can not accurately represent any physics-based simulations (Tahmasebi, 2018). The principal component analysis is a usual starting point for examining and describing variation in data and according to many studies has been done with promising results (Borůvka and Kozák, 2001; Borůvka, 2010). Though R^2 is a valid statistic to evaluate the prediction accuracy of a model, a high R-squared model may not lead to accurate predictions. This is because the model could systematically and considerably over- and/or under-estimate the data at different points along the regression line. An over-fitted model could also lead to poor predictions (Muñoz and Felicísimo, 2004). As a result, evaluation of the models with other performance statistics is important, preferably based on an independent observation set, to provide further information on the prediction accuracy of the models.

CHAPTER 3

The Effects of Slope and Altitude on Soil Organic Carbon and Clay Content in Different Land Uses: A Case Study in the Czech Republic

Citation:

Nozari S., Borůvka L. (2023): The effects of slope and altitude on soil organic carbon and clay content in different land uses: A case study in the Czech Republic. Soil & Water Res., 18: 204–218. https://doi.org/10.17221/105/2022-SWR

Abstract: Soil organic carbon (SOC) and clay, as indicators of soil fertility, are mainly used to determine the ability of soil to retain water and store the nutrients that are necessary for plant growth. However, the distribution of SOC and clay is influenced by topography and land-use. In the present study, the relationships between SOC, clay, altitude, and slope in the topsoil of two different districts in the Czech Republic including the Liberec (71 samples) and Domažlice (67 samples) districts were investigated. To analyze the relationships between slope and SOC, linear regression was used. Results showed that SOC content increased when slope, clay, or altitude increased; however, there were no significant correlations between SOC and clay in both districts. Clay increased with decreasing slope, but clay and altitude were not correlated well in both areas. Then, study areas were divided into three land-use types including arable land, forest, and complex system of agriculture, parcels, and forests. Consequently, the correlations between SOC and slope and clay and slope were generally improved, indicating the importance of land-use on SOC and clay content. Additionally, using multiple regression with several topographic factors can provide a better prediction of SOC and clay content in each land-use for both districts, indicating the complex effects of topography on SOC and clay.

Keywords: ANOVA; coefficient of determination; correlation coefficient; linear regression; multiple regression; SOC

3.1. Introduction

The largest global stock of organic carbon on land is estimated in the soil at 2500 Pg to 2-m in-depth and is approximately twice as large as the atmospheric carbon stock (Adhikari et al. 2019). Soil organic carbon (SOC) is known as the main indication of soil quality and fertility because soil chemical properties such as pH and nutrients availability, soil physical properties such as structure and hydraulic conductivity, and soil biological activities such as microbial activity are substantially influenced by SOC (Nisha et al. 2007).

Slope and altitude are two important variables that affect the intensity and frequency of erosion, and subsequently, SOC and clay contents (Wei et al. 2017; Khan et al. 2019; Baltensweiler et al. 2020). Additionally, the relationship between SOC and clay is vital for the investigation of changes in SOC stocks. Many studies revealed that SOC level increases with increasing clay content (Zhong et al. 2018; Gruba & Socha 2019). This is because clay particles adsorb great amounts of SOC and clay soils are less aerated, so the decomposition of soil organic matter (SOM) is low (Hartati & Sudarmadji 2016). Inadequacy of information on variation of SOC and clay contents with changes in slope and altitude in different land-uses are major bottlenecks for predicting the SOC and clay contents.

This study aims to evaluate the relationship between slope, altitude, SOC, and clay content as well as the effects of land-use type on SOC stock. Moreover, similarities and/or differences in SOC and clay distribution in two different districts including Liberec and Domažlice districts, Czech Republic were investigated.

3.1.1. Literature review

Topographical factors, such as slope and altitude, and different land-uses substantially influence SOM content by affecting soil erosion and geological deposition processes as well as by controlling soil water, and subsequently, plant litter production and decomposition (Birkeland 1984; Thai et al. 2021).

It has been shown that increasing slope increases flow velocity leading to an increase in erosion intensity and frequency (Liu et al. 2015). In other words, a land with a steeper slope is expected to lose soil leading to SOC loss. However, Hontoria et al. (1999) reported a positive correlation between SOC and slope in Peninsular Spain. It indicates that in addition to slope itself, the position on the slope is also important, as at the upper parts erosion can cause SOC loss, while at the bottom of the slope, sedimentation can lead to SOC increase.

The behaviour and distribution of SOC on the slope is modified by clay. Clay can control SOC dynamics by protecting SOM from decomposition. It has been shown that SOC content increases with increasing clay and finer soil (containing more clay content) has higher SOC content (Gao et al. 2014). Several studies have indicated that the prediction of SOC loss depends on clay content. Olson et al. (2012) indicated that SOC can be predicted by flow dynamics when the clay content is low, however, SOC loss should be predicted by slope when clay content is high.

Altitude is another main factor that influences soil properties. Generally, SOC increases with increasing altitude (Griffiths et al. 2009). Altitude variation mainly affects climatic variables and vegetation types that have major impacts on SOC content (Zhu et al. 2010).

It has been also found that different land-use classes affect the soil quality indicators such as SOC and, to smaller extent, clay. This is because land-use is one of the main factors controlling soil capacity to retain water and nutrients as well as providing other ecosystem services (Wang et al. 2009; Ngatia et al. 2021). Xiaojun et al. (2013) also reported that land-use can affect SOC content even in a region with the same parent material and climate. They showed that different land-uses (with different plant covers and soil management practices) had substantially different SOC contents.

Other researchers also found that changing grassland to cropland led to a significant soil degradation through loss of fine soil particles, SOC, and nutrients. However, SOC gradually increased after returning to grassland (Zhao et al. 2005; Wei & Fang 2009). Wiesmeier et al. (2012) also indicated that grassland soils had considerably more SOC content compared to that of forest and cropland soils in southeast Germany. However, the differences between grassland and forest with cropland in this region were lower than the results observed from other researchers in central European countries (Gingrich et al. 2007; Martin et al. 2011; Wiesmeier et al. 2012).

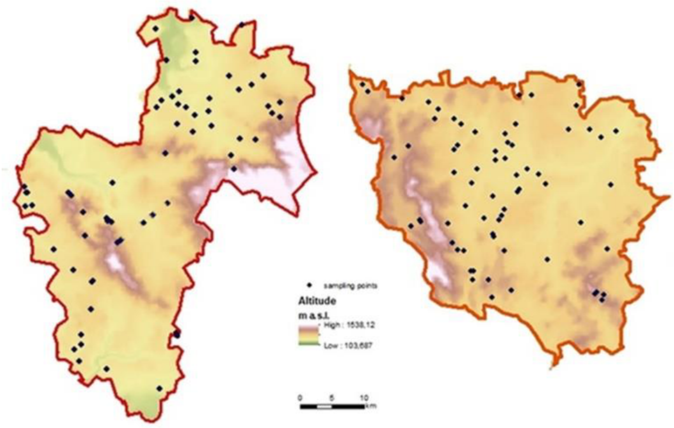
3.2. Materials and Methods

3.2.1. Study areas and sample collection

As shown in Figure 3.1, the study areas are Liberec (989 km^2) and Domažlice (1123 km^2) districts that are located in the north and west parts of the Czech Republic, respectively. Liberec district is covered by 42.4% forest and 47.2% agricultural land while Domažlice district is covered by 38.2% forest and 53.0% agricultural land, indicating that Domažlice district is covered mostly by agricultural land and by a slightly smaller proportion of forest area than Liberec district (Miko & Hošek 2009). The third category considered in this paper, the complex systems of agriculture, parcels, and forests, forms less than 6% of the area in each of the two districts.

Figure 3. 1. Location of the Liberec (red) and Domažlice (green) districts in the Czech Republic

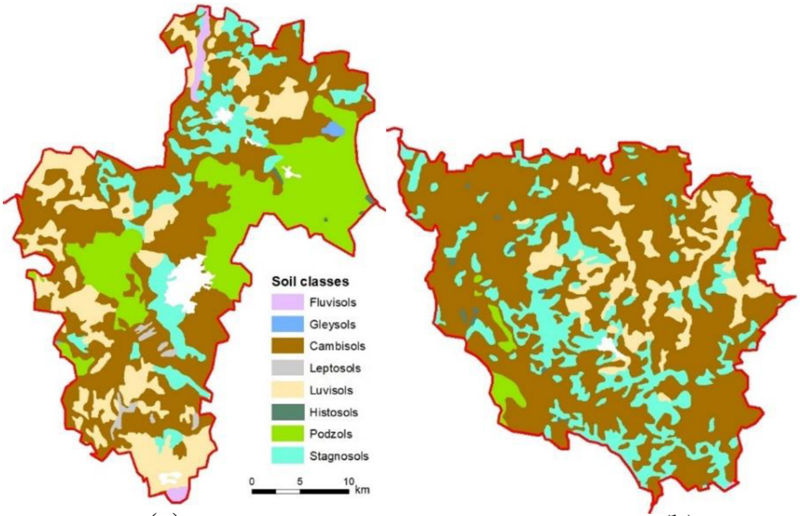
As shown in Figure 3.2, simple random sampling design was used; the altitude of sampling locations ranged between 337 to 436 m for the Liberec district (71 samples) and 383 to 691 m for the Domažlice district (67 samples). Composite samples were created from 3 subsamples from an area of several square meters at each location. These subsamples were collected using a steel soil sample probe. Initially, each sample location was navigated by means of a handheld GPS tracker after clearance of debris (e.g. grasses, twigs, etc.) at each point. The probe was inserted in the mineral topsoil at the depth of 0–30 cm. This depth was selected because the 0–30 cm depth indicates the plough depth and SOC estimation in this depth is an important factor in farm management. Organic horizons (forest floor) were omitted and not sampled to get corresponding layers from all land-use types. It should be noted that the top 100 cm soil depth shows a rooting depth for many field crops (Adhikari et al. 2014).



(a) (b)

Figure 3. 2. Distribution of sampling locations maps in a) Liberec, and b) Domažlice districts with altitudes

The soil was classified according to the Czech taxonomic soil classification system and WRB system (Němeček & Kozák 2002; IUSS 2015). Six major reference soil groups were recognized in the Liberec district including Cambisols, Podzols, Gleysols, Stagnosols, Luvisols, and Fluvisols (Figure 3.3a). Five dominant reference soil groups were also observed in the Domažlice district: Cambisols, Gleysols, Stagnosols, Luvisols, and Fluvisols (Figure 3.3b).



(a) (b)

Figure 3. 3. Dominant reference soil groups in laboratory analyses and data processing

The debris, rocks together with plant roots were manually removed from the collected soil samples. For analyses, the soil samples were air-dried, sieved to 2 mm, and thoroughly mixed. For SOC determination, the samples were further ground to pass the 0.25 mm mesh. Then, soil samples were analysed through the oxidimetric modified Tyurin method (Pospíšil 1964). Clay content of each sample was measured using the hydrometer method (Elfaki et al. 2016). Terrain characteristics were calculated by system for automated geoscientific analyses (SAGA) software, (Ver. 7.2.0) (Conrad et al. 2015), using digital terrain model 4G (DTM 4G) acquired from airborne laser scanning (ALS), also commonly known as light detection and ranging (LiDAR) with an original resolution of 5×5 m. Variables used in this study were altitude and slope. Both of these variables have the potential to contribute to SOC and clay spatial distribution. Land-use categories were obtained from the database CORINE Land Cover 2018 (EEA 2018) at the resolution of 100 m. (Taghizadeh-Mehrjardi et al. 2014; Figure 3.4).

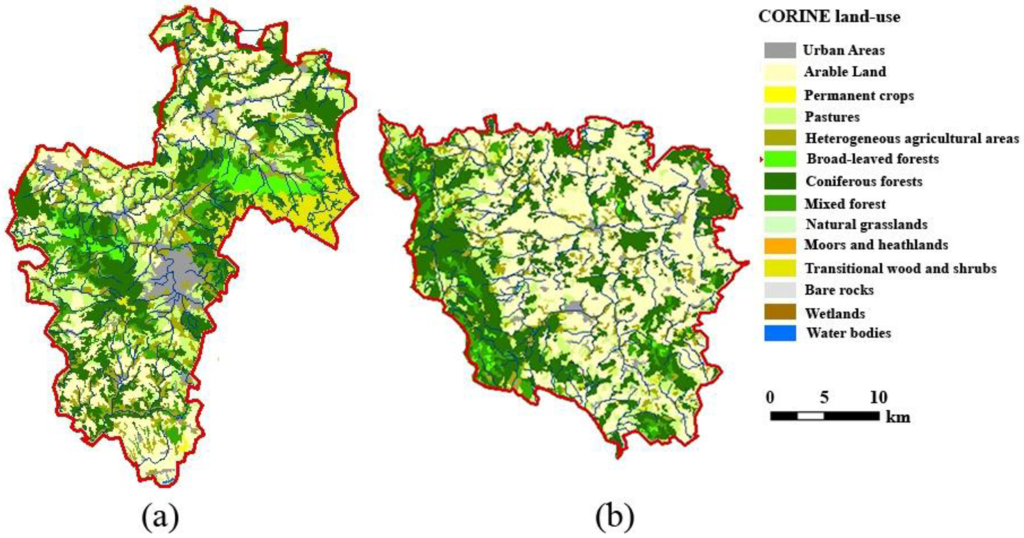


Figure 3. 4. Land-use maps for a) Liberec, and b) Domažlice districts

3.3. Results and Discussion

3.3.1. Basic statistical terrain analysis

Statistical differences in mean values of SOC and slope were identified by one-way ANOVA using R (Ver. 3.5.1) (The R Foundation for Statistical Computing, 2018) and SPSS (Ver. 11) (SPSS Inc). Additionally, the correlation matrix between the selected variables (SOC, clay, slope, and altitude) was determined using the SPSS software package. Linear regression analysis was conducted using SPSS to identify the relationships between slope, SOC, and clay. The topographic parameters were used as independent variables that were altitude and slope. Summary statistics for SOC, clay, and terrain parameters (altitude and slope) for sampling sites in the Liberec and Domažlice districts are presented in Table 3.1.

Table 3. 1. Summary statistics for soil organic carbon (SOC), clay, and terrain parameters (altitude and slope) for sampling sites in the Liberec (71 locations) and Domažlice (67 locations) districts

Variable	Minimum	Maximum	Mean	standard deviation
Liberec district				
SOC (%)	0.42	11.33	2.83	2.50
Clay (%)	2.70	24.32	8.58	3.46
Altitude (m)	239.7	719.1	412.6	98.3
Slope (radian)	0.005	0.534	0.081	0.091
Domažlice district				
SOC (%)	0.00	9.33	2.83	2.39
Clay (%)	2.18	23.71	11.95	4.42
Altitude (m)	356.3	719.7	481.4	77.4
Slope (radian)	0.003	0.158	0.048	0.039

3.3.2. The relationship between slope and SOC content

Tables A4 and A5 present the correlation matrices for Liberec and Domažlice districts. Results show that the distribution of SOC varies with changing the slope (Li et al. 2016). Tables A4 and A5 show that $r = 0.525$ and $P < 0.01$ for Liberec district and $r = 0.444$ and $P < 0.01$ for Domažlice district indicating that there were significant correlations between slope and SOC for both study areas, which is consistent with observations from other studies (Nozari & Borůvka 2020, 2021). A positive correlation between slope and SOC means that there are bigger SOC contents on steeper slopes. Therefore, spots with steeper slopes shown in Figure 3.5 (slope maps produced by SAGA) may contain more SOC.

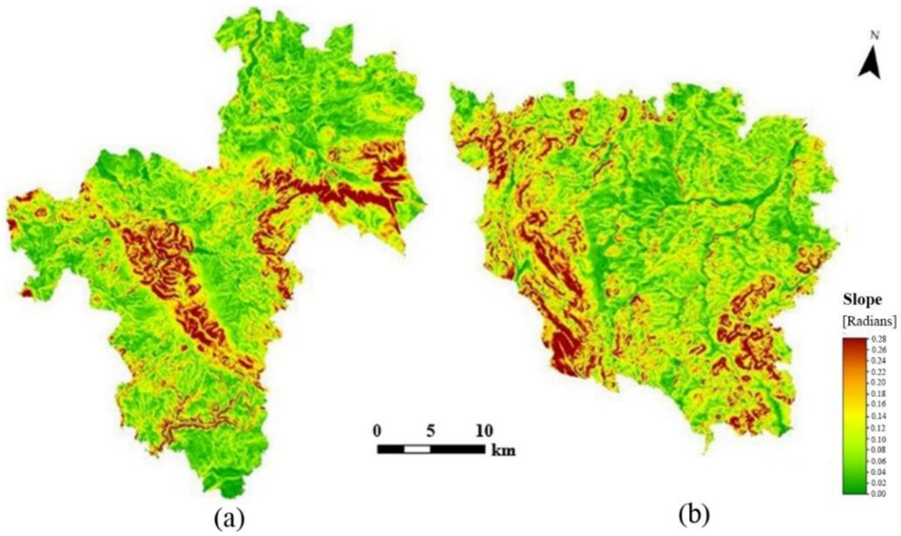


Figure 3. 5. Slope maps for a) Liberec, and b) Domažlice districts

Figure 3.6 also shows a positive linear relationship between slope and SOC for both districts (SOC increases with increasing slope). However, results from the linear model indicated that slope does not explain much the variation of SOC as the dependent variable (small R^2).

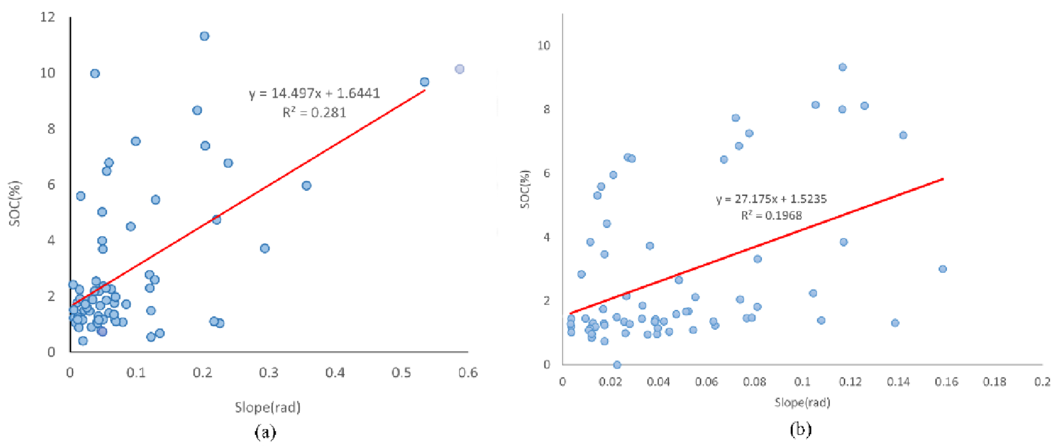


Figure 3. 6. Relationship between slope and soil organic carbon (SOC) (linear regression) in a) Liberec, and b) Domažlice districts

3.3.3. The relationship between slope and clay content

As shown in Figure 3.7, linear models between clay and slope showed a decreasing trend for both districts. This can be attributed to the greater transport of SOC and clay on steeper slopes due to the greater erosion. Also, clay content in the soil was relatively low based on this study dataset, which corresponds to the mostly granitic parent material, particularly in the Liberec district, as well as Podzols and Cambisols as dominant soil classes. This may explain why there was no significant correlation between slope components and clay. Additionally, logarithmic, exponential, power, and polynomial functions were used, however, similar results were observed.

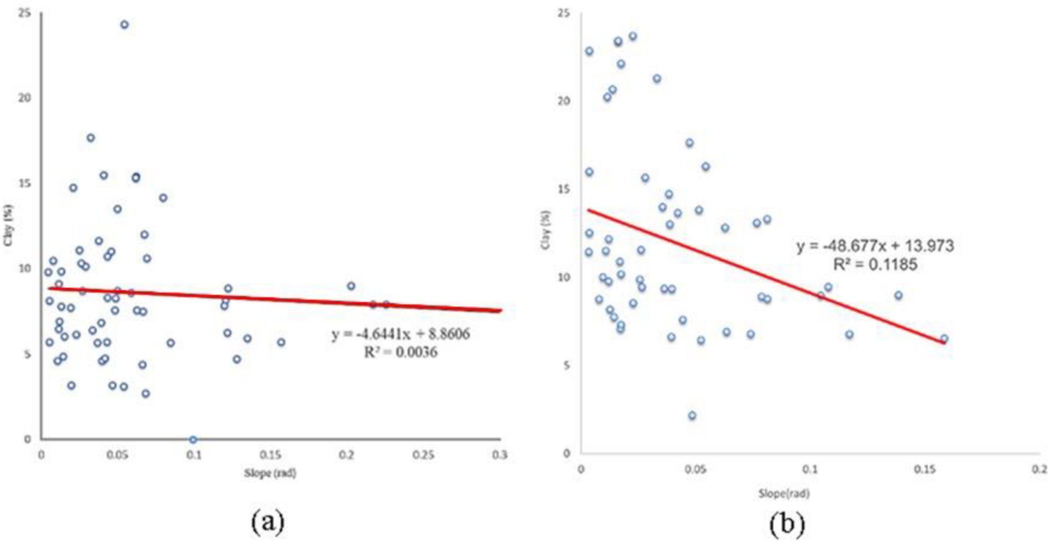


Figure 3. 7. Relationship between slope and clay (linear regression) in a) Liberec, and b) Domažlice districts

3.3.4. The relationship between altitude and SOC content

The correlations between altitude and SOC content were identified. Figure 3.8 shows that SOC increased with increasing altitude. These results showed that the average SOC concentrations increased with altitude, even after considering the effects of land-use and landscape position. This suggests that SOC is responding to climatic variables (the most likely temperature that decreases as altitude increases). This may also be confounded by the recent nature of land-use change (i.e. agricultural lands at higher altitudes are more likely to have been recently converted) and higher levels of soil acidity at higher altitudes, which may decrease decomposition rates. Therefore, the effect of altitude on SOC obtained from this study may be due to the combined effects of increased leaching at higher altitudes (subsurface pH change) as well as soil acidification through reduced decomposition and the build-up of a high organic matter litter layer with organic acids. Similar studies showed that SOM and soil nutrients are significantly correlated with altitude (Wu et al. 2016; Massawe et al. 2017; Gebrehiwot et al. 2018). Additionally, the coefficient of determination (R^2) between altitude and SOC was 0.144 and 0.433 for Liberec and Domažlice districts, respectively, indicating that there was a moderate correlation between altitude and SOC. Borůvka et al. (2022) reported that the importance of environmental variables in the models for SOC stock prediction varies in different regions and altitudes.

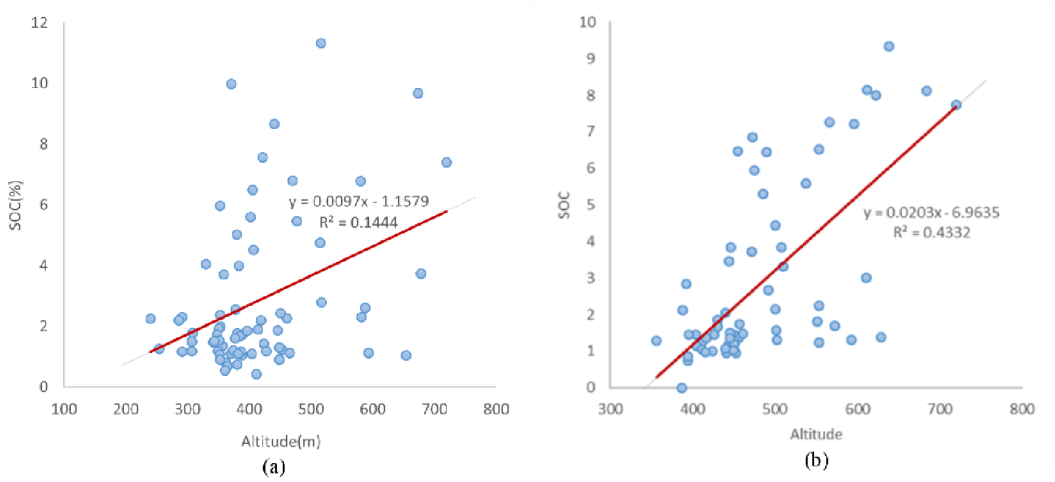


Figure 3. 8. Relationship between altitude and soil organic carbon (SOC) (linear regression) in a) Liberec, and b) Domažlice districts

3.3.5. The relationship between altitude and clay content

Figure 3.9 shows the correlations between altitude and clay content for both Liberec and Domažlice districts. The R^2 value for the relationship between altitude and clay was 0.002 and 0.039 for Liberec and Domažlice districts, respectively, indicating that altitude and clay content were not correlated in both areas as shown in Figure 3.9.

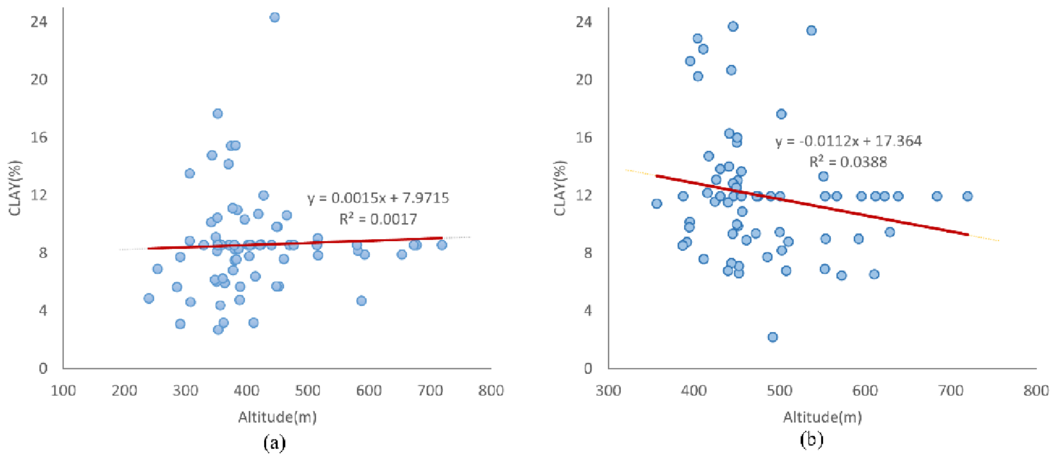


Figure 3. 9. Relationship between altitude and clay (linear regression) in a) Liberec, and b) Domažlice districts

The weak correlation between altitude and clay content can be attributed to different factors. Geological processes such as erosion can occur independently of altitude variations, leading to different clay content in areas at different altitudes solely based on local geological conditions, rather than being directly influenced by altitude. Additionally, although climate and parent material affect clay content, only climatic conditions (such as temperature and precipitation) are directly influenced by altitude. Moreover, there can be localized variations in clay content due to factors such microclimates, drainage patterns, and land-use practices that may not directly correlate with altitude. Finally, soil movement through erosion or runoff can cause spatial variations in clay content, irrespective of altitude, as landscape dynamics re-distribute clay particles (Gebrehiwot et al. 2018).

3.3.6. The relationship between SOC and clay content

As presented in Tables A4 and A5, the correlation between SOC and clay was insignificant in both districts with r = -0.026 for Liberec and r = -0.108 for Domažlice districts. Additionally, clay content in the soil was mostly less than 20% as presented in Table 3.1, which could not have significantly minimized mineralization. On the other hand, soils containing lower content of clay may contain lower SOC content due to the high decomposition rate of organo-mineral fractions (Lee et al. 2009; Pronk et al. 2012). This can be due to the parent material in sampling locations (including granites, loamy glacial sediments, micaceous schist, and phyllites) because parent materials influence the organic matter stock. As an example, soils developed from inherently rich materials, such as basalt, are more fertile and have higher SOC than soils formed from granitic materials which include fewer mineral nutrients (Straaten 2011; Hartati & Sudarmadji 2016). Moreover, very acidic reaction of the soils under study can reduce the decomposition rate and thus lead to higher SOC content. In the present study, soil textural classes for the Liberec district included sedimentary coarse-textured rocks, acid granites, similar rocks (coarse textured), polygenetic loams, and loamy glacial sediments. Similarly, soil textural classes for the Domažlice district included polygenetic loams, loamy glacial sediments, micaceous schist, phyllites (medium textured), and a small proportion of acid granites, and similar rocks (coarse textured). The high content of SOC in sandy soils particularly at higher altitudes under coniferous forests can be caused by reduced mineralization due to strong acidity, as it is typical for Podzols and some Cambisols.

3.3.7. Comparing observed and predicted variables (SOC and clay)

Based on the relationships found in this study, multiple linear models were constructed to predict SOC and clay contents from altitude and slope separately for the Liberec and Domažlice districts. Figures 3.10 and 3.11 show the relationship (linear regression) between observed and predicted values for two variables including SOC and clay contents, respectively. Results indicate that although correlations between observed and predicted values were better in the Domažlice district than Liberec district, the correlation between observed and predicted SOC is much better than the correlation between observed and predicted clay content in both districts. This can be due to the low content and great variability in observed clay contents that decrease the possibility of creating a regression model with a strong correlation.

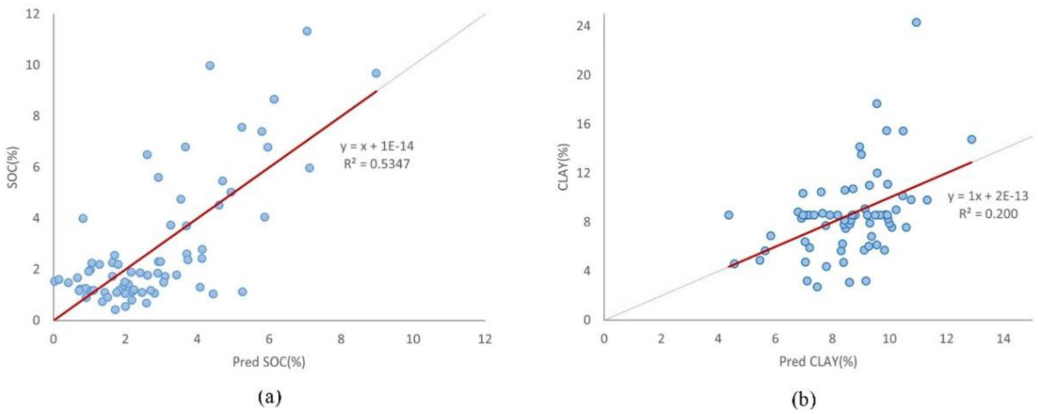


Figure 3. 10. Relationship (linear regression) between a) observed and predicted soil organic carbon (SOC) content, and b) observed and predicted clay in the Liberec District

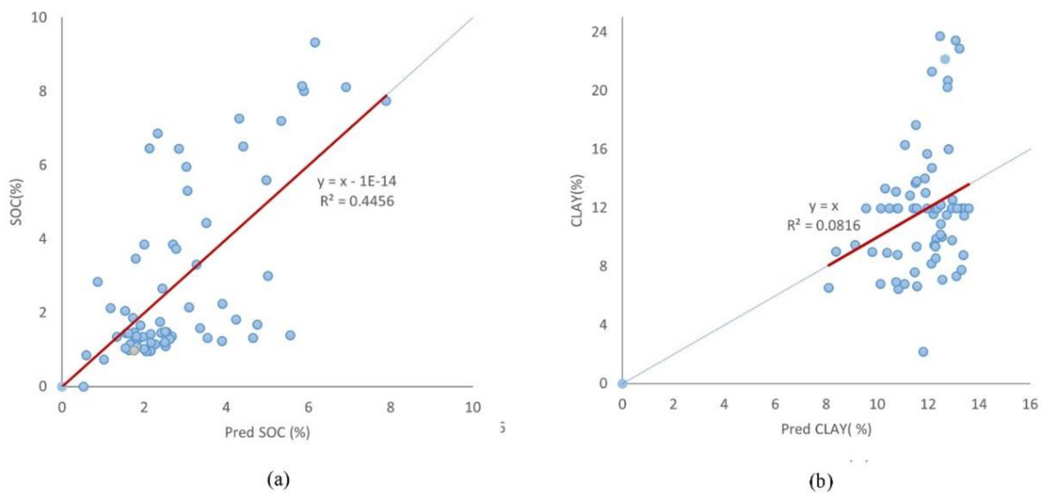


Figure 3. 11. Relationship (linear regression) between a) observed and predicted soil organic carbon (SOC) content, and b) observed and predicted clay in the Domažlice district

3.3.8. Data separation by land-use classes and model analysis

Generally, SOC is affected by land-use and increases with increasing altitude within land-use categories due to the convergent effects of temperature decrease, precipitation changes, acidification, and intactness of native ecosystems. To analyse the correlation between SOC and slope for Liberec and Domažlice districts in more detail, datasets were divided into three subsets based on the land-use including arable land, forest, and complex system of agriculture, parcels, and forests (Figure 3.4). In the following sub-sections, the relationship between slope and SOC, across three different land-use classes, is evaluated. Additionally, a comprehensive comparison between the observed and predicted variables (SOC and clay) is conducted.

3.3.9. The relationship between slope and SOC content after data separation

As shown in Figures 3.12a through 3.14a, SOC and slopes for Liberec district were correlated at arable land with $R^2 = 0.353$ at $P < 0.05$ (35 samples), at forest land with $R^2 = 0.140$ (22 samples), and at complex system with $R^2 = 0.598$ at $P < 0.05$ (14 samples). As shown in Figures 3.12b through 3.14b, SOC and slopes for Domažlice district were correlated very little at arable land with $R^2 = 0.008$ (35 samples), at forest land with $R^2 = 0.083$ (20 samples), and at complex system with $R^2 = 0.386$ (12 samples).

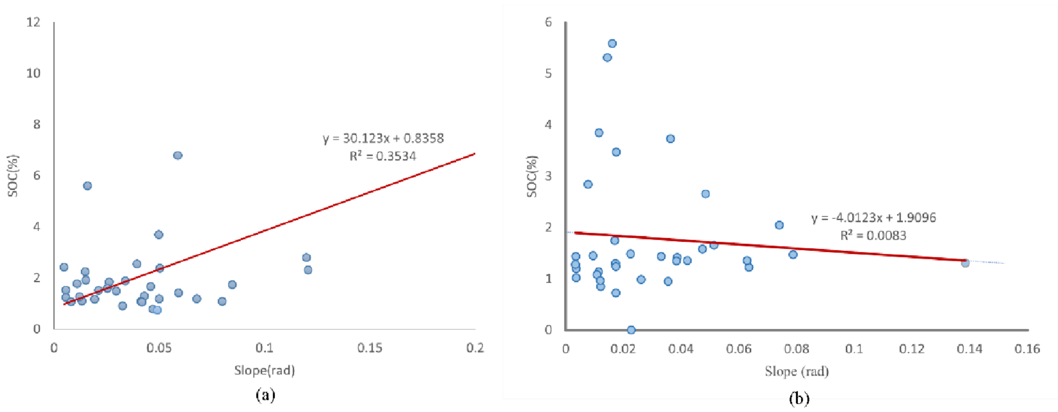


Figure 3. 12. Relationship between slope and soil organic carbon (SOC) (linear regression) in arable lands in a) Liberec, and b) Domažlice districts

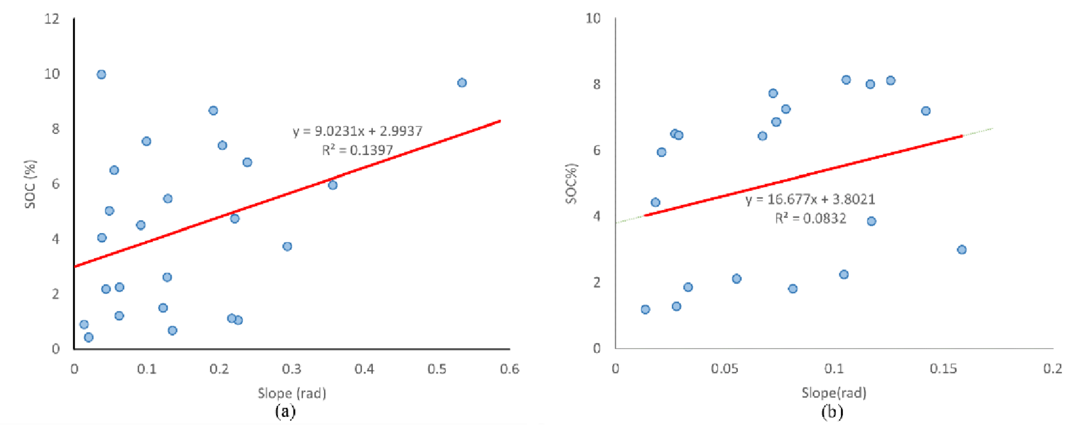


Figure 3. 13. Relationship between slope and soil organic carbon (SOC) (linear regression) in forest areas in a) Liberec, and b) Domažlice districts

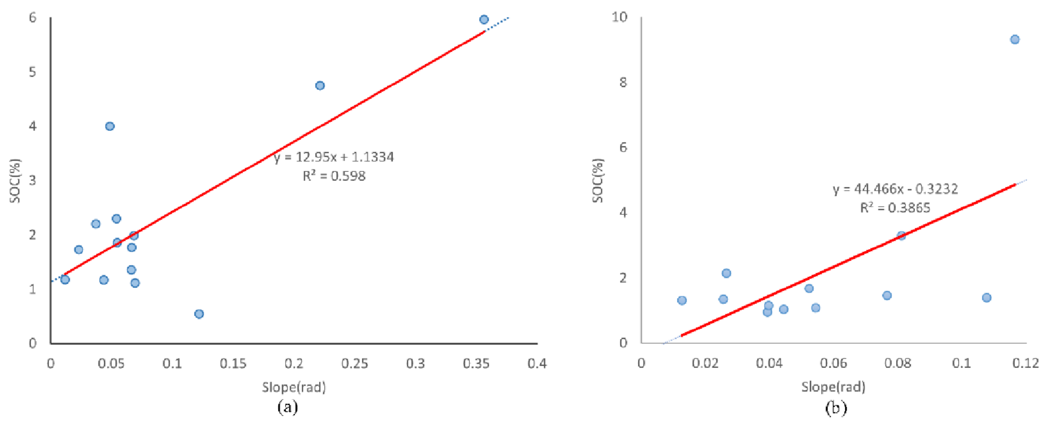


Figure 3. 14. Relationship between slope and soil organic carbon (SOC) (linear regression) in complex systems in a) Liberec, and b) Domažlice districts

Comparing Figure 3.6 with Figures 3.12 through 3.14 shows that the data separation improved the correlations between SOC and slope in some subsets, confirming the effects of land-use on SOC. The correlation for arable areas in the Liberec district increased more than that in the Domažlice district. This is because of the different effects of agricultural systems and management, tillage, slope, soil biology, and erosion on SOC. For instance, agricultural systems with conventional tillage in the Czech Republic have been affected over the years (Šíp et al. 2009).

The results of the surface runoff speed corroborate the significant benefits of soil conservation tillage technology. Also, tillage typically reduces mean SOC content and homogenizes the horizontal and vertical distribution of SOC (Dornbush & Von Haden 2017). Additionally, the spatial distribution of biologically mediated soil ecosystem services is impacted by agricultural practices, changing SOC, because most of the soil biological communities are dependent on SOC substrates. A reduction in SOM decomposition and thus increased SOC concentration is due to an increase in acidity and consequent reduction of biological activities. Soil compaction is another problem due to intensive conventional farming. The original causes of the decrease in SOC are conventional farming without using organic fertilizer or other SOM (Šíp et al. 2009). This shows the necessity for sustainable land management practices, mostly those that reduce erosion and build SOM. As shown in Figure 3.13, the correlation between SOC and slope in forests of the Liberec district was similar to the Domažlice district while both correlation values were low. This can be due to the effects of forest tree species on soil layers. Generally, the soil of coniferous stands, which constituted the largest group in the forests group, contained significantly less stored carbon than the soil of other species, as a large part of SOC is stored in the forest floor that is not considered in this study. Also, carbon stocks can potentially be affected by soil disturbance events depending on forest type or topographic parameters.

It appears that changes in SOC are affected by a range of soil-management practices relating to tillage management, a total of crop residues, fertilizer, organic losses, and different crop rotation programs (Ghimire et al. 2012).

In complex systems, correlations were better than the arable land and forest. Francaviglia et al. (2019) showed that variated arable cropping systems and various management plans in selected European areas had positive effects on SOC.

3.3.10. Multiple linear regression models for SOC and clay content prediction

To achieve more appropriate results, multiple regressions were conducted on different land-use types to assess the relationship between SOC and other environmental variables (altitude and slope) as well as the relationship between clay and other parameters (altitude and slope) in both districts. The results showed the multicollinearity between environmental variables and predicting SOC and clay with these variables using simple regression would not be reliable while using multiple regression could effectively improve the results. The results also illustrated that the multiple regression was significantly changed by dividing the study area into various land-use types because the distribution of SOC and clay varies with land-use. Similarly, Lettens et al. (2005) separated the regions into 289 landscape units and predicted the SOC stocks for each landscape unit in Belgium. They reported that SOC stocks were continuously influenced by some external characteristics, mainly land-use history and usual land management and climate.

Table 3.2 presents the distribution of SOC and clay as dependent variables, respectively, based on the altitude and slope for different land-uses and entire dataset in both districts. It can serve as a basis for multiple regression model for the selection of the predictors. For instance, multiple Pearson correlation coefficients of 11 different variables for arable land in the Liberec district are also presented in Table A3. SOC was mostly positively correlated with altitude and slope, while clay was mostly negatively correlated with altitude and slope (Table 3.2).

Table 3. 2. Regression coefficients of multiple regression models for soil organic carbon (SOC) and clay (dependent variable) prediction based on altitude and slope (independent variables) in different land-uses and for the entire dataset

Source	Liberec District				Domažlice District			
	arable	forest	complex	entire dataset	arable	Forest	complex	entire dataset
SOC								
Altitude	0.421	0.142	-0.177	0.380	0.373	0.594	0.622	0.658
Slope	0.595	0.374	-0.306	0.525	-0.091	0.289	0.622	0.444
Clay								
Altitude	-0.019	-0.154	0.688	0.042	-0.161	-0.293	-0.257	-0.197
Slope	0.056	-0.061	-0.085	-0.025	-0.267	-0.599	0.180	-0.276

Figures 3.15 through 3.20 show multiple regression (linear trend) of the relationship between observed and predicted SOC as well as observed and predicted clay for arable land, forest, and complex system in both districts. Results showed that the correlations obtained from multiple regression (linear trend) for separate land-uses were more significant (greater R^2) compared to the correlations obtained from simple regression (linear trend) without land-use separation. For instance, the relationship between observed and predicted SOC in the Liberec district improved with increasing $R^2 = 0.535$ (Figure 3.10a) to $R^2 = 0.970$ (Figure 3.17a) when considering only complex system. As another example, R^2 value for the relationship between observed and predicted clay increased from 0.200 (Figure 3.10b) to an impressive 0.914 (Figure 3.17b) by exclusively considering complex systems within the Liberec district.

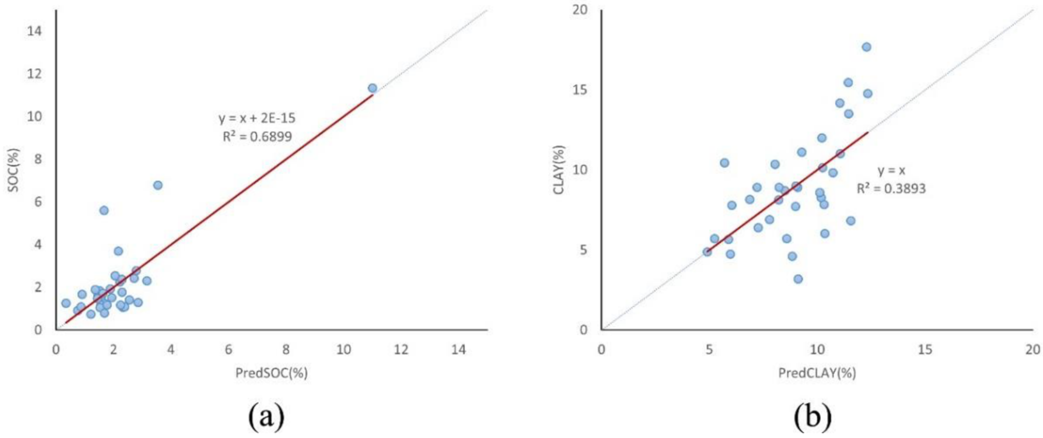


Figure 3. 15. Linear trend of the relationship between observed values and values predicted by multiple regression for a) soil organic carbon (SOC), and b) clay in the arable land (Liberec)

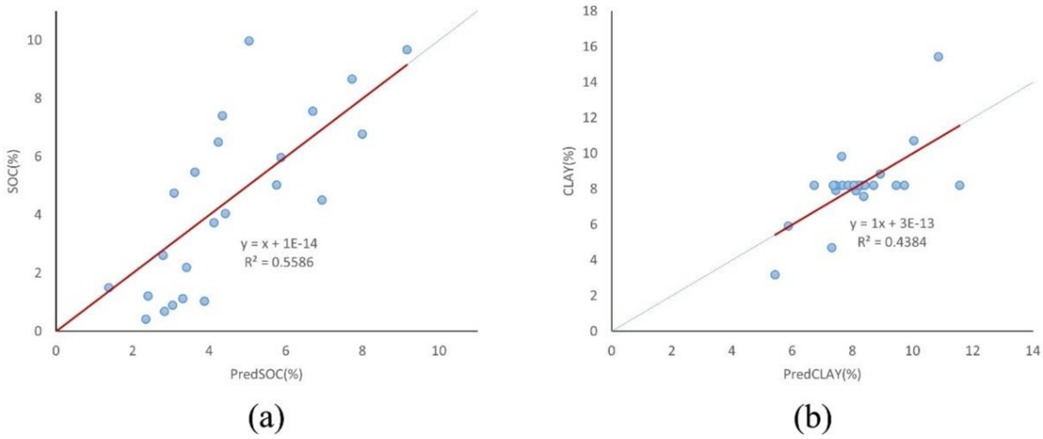


Figure 3. 16. Linear trend of the relationship between observed values and values predicted by multiple regression for a) soil organic carbon (SOC), and b) clay in the forest (Liberec)

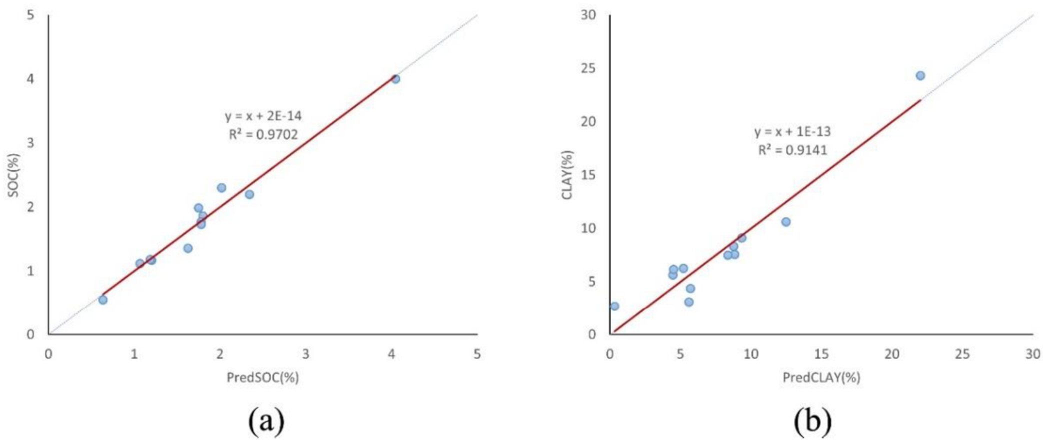


Figure 3. 17. Linear trend of the relationship between observed values and values predicted by multiple regression for a) soil organic carbon (SOC), and b) clay in the complex system (Liberec)

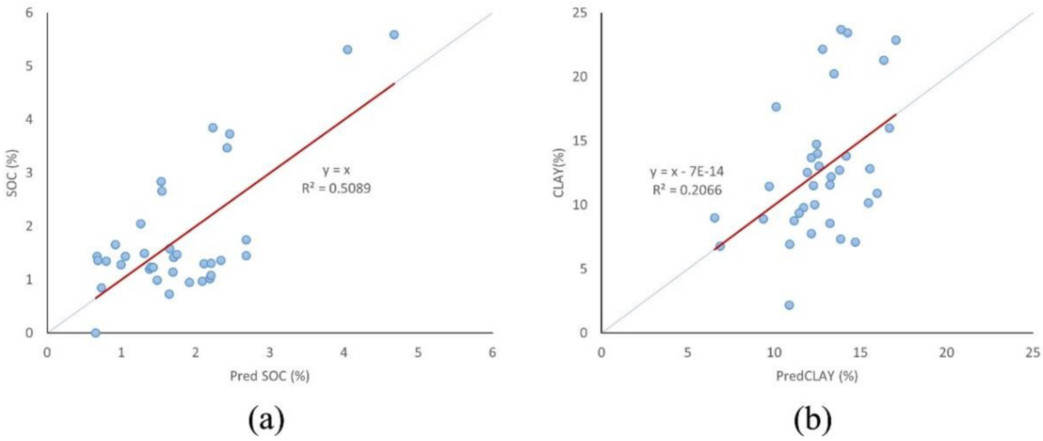


Figure 3. 18. Linear trend of the relationship between observed values and values predicted by multiple regression for a) soil organic carbon (SOC), and b) clay in the arable land (Domažlice)

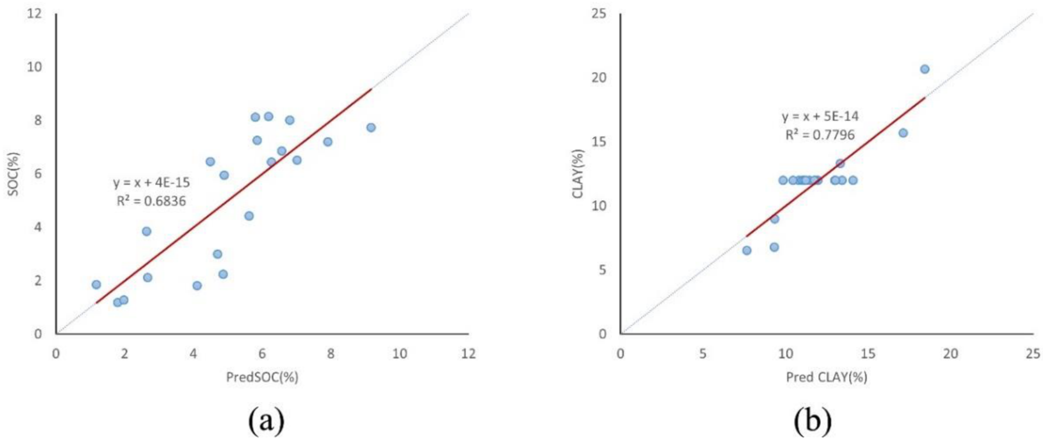


Figure 3. 19. Linear trend of the relationship between observed values and values predicted by multiple regression for a) soil organic carbon (SOC), and b) clay in the forest (Domažlice)

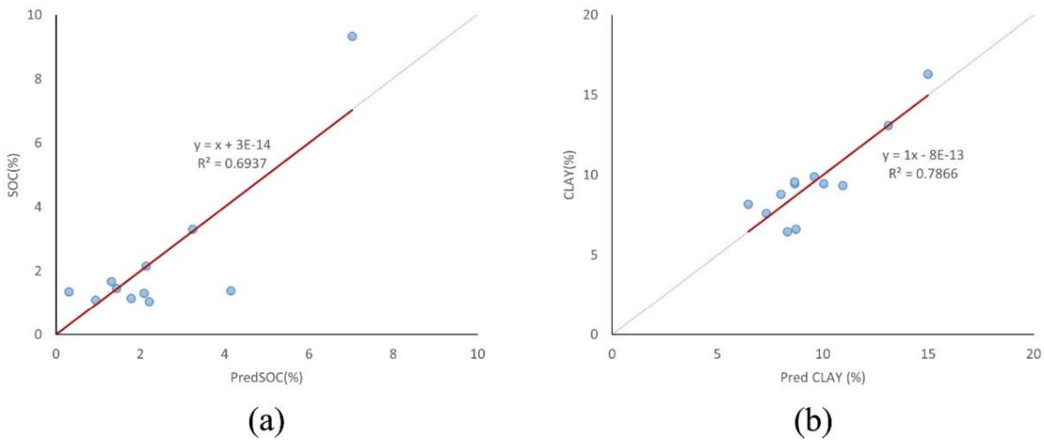


Figure 3. 20. Linear trend of the relationship between observed values and values predicted by multiple regression for a) soil organic carbon (SOC), and b) clay in the complex system (Domažlice)

3.4. Conclusions

From the results of this study, the following conclusions can be drawn:

1. Although SOC content increases when slope or clay increase, the correlation between SOC and clay was not significant in both districts.

2. SOC increases with increasing altitude, most likely due to the combined effects of increased leaching at higher altitudes and soil acidity leading to reduced decomposition. However, there was only a moderate correlation between SOC and altitude.

3. Clay increases with decreasing slope for both districts.

4. Clay and altitude were not correlated well in both areas, most likely due to the effects of erosion and runoff which transported sediments from higher altitudes and accumulated them in the lower parts of the basin.

5. The correlation between observed and predicted SOC is much better than the correlation between observed and predicted clay content in both districts, due to the low content and great variability in observed clay contents.

6. Multiple linear regression models based on topographical variables were constructed for SOC and clay content prediction separately for each district. A better correlation between observed and predicted values (SOC and clay) was observed in the Domažlice district than in the Liberec district. This can be due to the low content and great variability in observed clay contents in Liberec district that decreases the possibility of creating a linear model with a strong correlation.

7. Data separation by land-use types improved the correlations between SOC and slope in some subsets, showing the significant effects of land-use on SOC.

8. Overall, it can be concluded that the variation of land-uses was influential in SOC distribution of the study areas and should be taken into consideration when evaluating the effects of future land-use changes on SOC and clay content on a regional scale. Additionally, a combination of slope and altitude could provide a better understanding of the effect of topography on SOC and clay.

3.5. 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: e105519.

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. Science of the Total Environment, 667: 833–845.

Baltensweiler A., Heuvelink G.B., Hanewinkel M., Walthert L. (2020): Microtopography shapes soil pH in flysch regions across Switzerland. Geoderma, 380: 114663.

Birkeland P.W. (1984): Soils and Geomorphology. New York, Oxford University Press.

Borůvka L., Vašát R., Šrámek V., Hellebrandová K.N., Fadrhonsová V., Sáňka M., Pavlů L., Sáňka O., Vacek O., Němeček K., Nozari S. (2022): Predictors for digital mapping of forest soil organic carbon stocks in different types of landscape. Soil and Water Research, 17: 69–79.

Conrad O., Bechtel B., Bock M., Dietrich H., Fischer E., Gerlitz L., Wehberg J., Wichmann V., Böhner J. (2015): System for Automated Geoscientific Analyses (SAGA) 2.1.4. Geoscientific Model Development, 8: 1991–2007.

Dornbush M.E., Von Haden A.C. (2017): Intensified agroecosystems and their effects on soil biodiversity and soil functions. Soil Health and Intensification of Agroecosystems, 2017: 173–193.

EEA (2018): CORINE Land Cover (CLC) 2018, Version 20. Kopenhagen, European Environment Agency. Available at https://www.eea.europa.eu/data-and-maps/ data/corine-land-cover-accountinglayers (accessed Oct 23, 2020).

Elfaki J.T., Gafer M.A., Sulieman M.M., Ali M.E. (2016): Hydrometer method against pipette method for estimating soil particle size distribution in some soil types selected from Central Sudan. International Journal of Engineering Research and Advanced Technology, 2: 25–41.

Francaviglia R., Álvaro-Fuentes J., Di Bene C., Gai L., Regina K., Turtola E. (2019): Diversified arable cropping systems and management schemes in selected European regions have positive effects on soil organic carbon content. Agriculture, 9: 261.

Gao G.L., Ding G.D., Zhao Y.Y., Wu B., Zhang Y.Q., Qin S.G., Bao Y.F., Yu M.H., Liu Y.D. (2014): Fractal approach to estimating changes in soil properties following the establishment of Caragana Korshinskii shelterbelts in Ningxia, NW China. Ecological Indicators, 43: 236–243.

Gebrehiwot K., Desalegn T., Woldu Z., Demissew S., Teferi E. (2018): Soil organic carbon stock in Abune Yosef afroalpine and sub-afroalpine vegetation, northern Ethiopia. Ecological Processes, 7: 1–9.

Ghimire R., Adhikari K.R., Chen Z.S., Shah S.C., Dahal K.R. (2012): Soil organic carbon sequestration as affected by tillage, crop residue, and nitrogen application in rice-wheat rotation system. Paddy and Water Environment, 10: 95–102.

Gingrich S., Erb K.H., Krausmann F., Gaube V., Haberl H. (2007): Long-term dynamics of terrestrial carbon stocks in Austria: A comprehensive assessment of the time period from 1830 to 2000. Regional Environmental Change, 7: 37–47.

Griffiths R.P., Madritch M.D., Swanson A.K. (2009): The effects of topography on forest soil characteristics in the Oregon Cascade Mountains (USA): Implications for the effects of climate change on soil properties. Forest Ecology and Management, 257: 1–7.

Gruba P., Socha J. (2019): Exploring the effects of dominant forest tree species, soil texture, altitude, and pH H2O on soil carbon stocks using generalized additive models. Forest Ecology and Management, 447: 105–114.

Hartati W., Sudarmadji T. (2016): Relationship between soil texture and soil organic matter content on mined-out lands in Berau, East Kalimantan, Indonesia. Nusantara Bioscience, 8: 83–88.

Hontoria C.R.D.J., Saa A., Rodríguez-Murillo J.C. (1999): Relationships between soil organic carbon and site characteristics in peninsular Spain. Soil Science Society of America Journal, 63: 614–621.

IUSS Working Group WRB (2015): World Reference Base for Soil Resources 2014, Update 2015. International Soil Classification System for Naming Soils and Creating Legends for Soil Maps. World Soil Resources Reports No. 106. FAO, Rome.

Khan H., Shafique M., Khan M.A., Bacha M.A., Shah S.U., Calligaris C. (2019): Landslide susceptibility assessment using Frequency ratio, a case study of northern Pakistan. The Egyptian Journal of Remote Sensing and Space Science, 22: 11–24.

Lee S.B., Lee C.H., Jung K.Y., Do Park K., Lee D., Kim P.J. (2009): Changes of soil organic carbon and its fractions in relation to soil physical properties in a long-term fertilized paddy. Soil and Tillage Research, 104: 227–232.

Lettens S., Van Orshoven J., van Wesemael B., De Vos B., Muys B. (2005): Stocks and fluxes of soil organic carbon for landscape units in Belgium derived from heterogeneous data sets for 1990 and 2000. Geoderma, 127: 11–23.

Li Z., Nie X., Xiaofeng Ch., Lin L., Liying S. (2016): Characteristics of soil and organic carbon loss induced by water erosion on the loess Plateau in China. PLoS ONE, 11: e0154591.

Liu D., She D., Shao G., Chen D. (2015): Rainfall intensity and slope gradient effects on sediment losses and splash from a saline–sodic soil under coastal reclamation. Catena, 128: 54–62.

Martin M.P., Wattenbach M., Smith P., Meersmans J., Jolivet C., Boulonne L., Arrouays D. (2011): Spatial distribution of soil organic carbon stocks in France. Biogeosciences, 8: 1053–1065.

Massawe B.H., Winowiecki L., Meliyo J.L., Mbogoni J.D., Msanya B.M., Kimaro D., Deckers J., Gulinck H., Lyamchai C., Sayula G., Msoka E. (2017): Assessing drivers of soil properties and classification in the West Usambara mountains, Tanzania. Geoderma Regional, 11: 141–154.

Miko L., Hošek M. (2009): The State of Nature and the Landscape in the Czech Republic. 1st Ed. Prague. Agency for Nature Conservation and Landscape Protection of the Czech Republic.

Němeček J., Kozák J. (2002): The Czech taxonomic soil classification system and the harmonization of soil maps. In: Micheli E., Nachtergaele F.O., Jones R.J.A., Montanarella L. (eds.): Soil Classification 2001. European Soil Bureau Research Report No.7. Luxembourg, Office for Official Publications of the European Communities: 47–53.

Ngatia L.W., Moriasi D., Grace III J.M., Fu R., Gardner C.S., Taylor R W. (2021): Land use change affects soil organic carbon: An indicator of soil health. In: Environmental Health. IntechOpen.

Nisha R., Kaushik A., Kaushik C.P. (2007): Effect of indigenous cyanobacterial application on structural stability and productivity of an organically poor semi-arid soil. Geoderma, 138: 49–56.

Nozari S., Borůvka L. (2020): The effect of landscape slope on soil organic carbon in the Liberec district in the Czech Republic. International Journal of Advanced Science and Technology, 29: 3934–3942.

Nozari S., Borůvka L. (2021): The effect of landscape slope on soil organic carbon in Domažlice district in the Czech Republic. In: Proc. Global Symposium on Soil Biodiversity. Keep Soil Alive, Protect Soil Biodiversity. Rome, Apr 19–22, 2021: 317–322.

Olson K., Gennadiyev A., Zhidkin A., Markelov M. (2012): Impacts of land-use change, slope, and erosion on soil organic carbon retention and storage. Soil Science, 177: 269Y278.

Pospíšil F. (1964): Fractionation of humus substances of several soil types in Czechoslovakia. Rostlinná Výroba, 10: 567–579.

Pronk G.J., Heister K., Ding G.C., Smalla K., Kögel-Knabner I. (2012): Development of biogeochemical interfaces in an artificial soil incubation experiment; aggregation and formation of organo-mineral associations. Geoderma, 189: 585–594.

Šíp V., Růžek P., Chrpová J., Vavera R., Kusá H. (2009): The effect of tillage practice, input level, and environment on the grain yield of winter wheat in the Czech Republic. Field Crops Research, 113: 131–137.

Straaten P.V. (2011): The geological basis of farming in Africa. In: Innovations as Key to the Green Revolution in Africa. Dordrecht, Springer: 31–47.

Taghizadeh-Mehrjardi R., Minasny B., Sarmadian F., Malone B.P. (2014): Digital mapping of soil salinity in Ardakan region, central Iran. Geoderma, 213: 15–28.

Thai S., Pavlů L., Tejnecký V., Vokurková P., Nozari S., Borůvka L. (2021): Comparison of soil organic matter composition under different land uses by DRIFT spectroscopy. Plant, Soil and Environment, 67: 255–263.

Wang G.X., Li S.N., Hu H.C., Li Y.S. (2009): Water regime shifts in the active soil layer of the Qinghai–Tibet Plateau permafrost region, under different levels of vegetation. Geoderma, 149: 280–289.

Wei W., Fang J. (2009): Soil respiration and human effects on global grasslands. Global Planetary Change, 67: 20–28.

Wei S., Zhan X., McLaughlin N.B., Chen X., Jia S., Liang A. (2017): Impact of soil water erosion processes on catchment export of soil aggregates and associated SOC. Geoderma, 294: 63–69.

Wiesmeier M., Spörlein P., Geuß U.W.E., Hangen E., Haug S., Reischl A., Schilling B., von Lützow M. Kögel-Knabner I. (2012): Soil organic carbon stocks in southeast Germany (Bavaria) as affected by land use, soil type and sampling depth. Global Change Biology, 18: 2233–2245.

Wu X., Zhao L., Fang H., Zhao Y., Smoak J.M., Pang Q., Ding Y. (2016): Environmental controls on soil organic carbon and nitrogen stocks in the high-altitude arid western Qinghai-Tibetan Plateau permafrost region. Journal of Geophysical Research: Biogeosciences, 121: 176–187.

Xiaojun N., Jianhui Z., Zhengan S. (2013): Dynamics of soil organic carbon and microbial biomass carbon in relation to water erosion and tillage erosion. PLoS ONE, 8: e64059.

Zhao W.Z., Xiao H.L., Liu Z.M, Li J. (2005): Soil degradation and restoration as affected by land use change in the semiarid Bashang area, northern China. Catena, 59: 173–186.

Zhong Z., Chen Z., Xu Y., Ren C., Yang G., Han X., Ren G., Feng Y. (2018): Relationship between soil organic carbon stocks and clay content under different climatic conditions in Central China. Forests, 9: 598.

Zhu B., Xiangping W., Jingyun F., Shilong P., Haihua S., Shuqing Z., Changhui P. (2010): Altitudinal changes in carbon storage of temperate forests on Mt Changbai, Northeast China. Journal of Plant Research, 123: 439–452.

CHAPTER 4

Predictors for Digital Mapping of Forest Soil Organic Carbon Stocks in Different Types of Landscape

Citation:

Borůvka L., Vašát R., Šrámek V., Neudertová Hellebrandová K., Fadrhonsová V., Sáňka M., Pavlů L., Sáňka O., Vacek O., Němeček K., **Nozari S.**, Oppong Sarkodie V.Y. (2022): Predictors for digital mapping of forest soil organic carbon stocks in different types of landscape. Soil & Water Res., 17: 69−79.
https://doi.org/10.17221/4/2022-SWR

Abstract: Forest soils have a high potential to store carbon and thus mitigate climate change. The information on spatial distribution of soil organic carbon (SOC) stocks is thus very important. This study aims to analyse the importance of environmental predictors for forest SOC stock prediction at the regional and national scale in the Czech Republic. A big database of forest soil data for more than 7000 sites was compiled from several surveys. SOC stocks were calculated from SOC content and bulk density for the topsoil mineral layer 0–30 cm. Spatial prediction models were developed separately for individual natural forest areas and for four subsets with different altitude range, using random forest method. The importance of environmental predictors in the models strongly differs between regions and altitudes. At lower altitudes, forest edaphic series and soil classes are strong predictors, while at higher altitudes the predictors related to topography become more important. The importance of soil classes depends on the pedodiversity level and on the difference in SOC stock between the classes. The contribution of forest types as predictors is limited when one (mostly coniferous) type dominates. Better prediction results can be obtained in smaller, but consistent regions, like some natural forest areas.

Keywords: stocks; digital soil mapping; environmental covariates; random forests; spatial distribution; terrain attributes

4.1. Introduction

The total forest ecosystem carbon (C) stock is large and in dynamic equilibrium with its environment (Lal 2005). There is a high potential for C sequestration and forest soils can thus contribute significantly to climate change mitigation. The ratio of C storage between tree biomass and soil depends on climate. At colder climate, lower C amounts are incorporated in tree biomass, but the soil organic carbon (SOC) stocks in soil are increased due to slower decomposition (Wen & He 2016). As the built-up of organic matter is a long-term process, forest continuity is an important factor of the SOC stocks in forest soils (Nitsch et al. 2018), as well as forest age (Jonard et al. 2017). Recovery of SOC stocks after forest soil disturbance can take decades (Dobor et al. 2018).

Factors influencing SOC amount in forest soils include (Lal 2005): climatic factors, topography, soil characteristics, natural disturbance, and anthropogenic factors (forest management, afforestation, and deforestation). Chuman et al. (2021) concluded that elevation (reflecting temperatures and precipitation levels) belongs to the most important factors controlling SOC pools in Podzols and Cambisols, together with legacy acid deposition of S and N compounds. The anthropogenic influence is particularly pronounced in forest floor and mineral topsoil.

The information on SOC stock spatial distribution and the influencing factors is important for the assessment of forest ecosystem functioning, soil ecosystem services, soil fertility, as well as a support for decision making in forest and environmental management. Digital soil mapping (DSM) provides a useful and efficient tool for the description and assessment of soil properties spatial distribution. General framework of DSM as the quantitative prediction of soil properties or classes using soil information and environmental covariates (scorpan model) was formalised by McBratney et al. (2003). Digital mapping of SOC contents or stocks is one of the most frequent applications of DSM (e.g. Lamichhane et al. 2019). Various prediction models are used, and various sets of covariates (predictors) are tested. Miller et al. (2015) tested a pool of 412 potential predictors and found that models with limited predictor pools can substitute other predictors to compensate for the missing variables.

Random forests (RF, Breiman 2001) is one of the most often used prediction methods in DSM (e.g. Calvo de Anta et al. 2020; Yamashita et al. 2022). However, Were et al. (2015) found that RF overestimated SOC stocks compared to models based on support vector regression and artificial neural networks. Martin et al. (2014) found that robust geostatistical modelling of residuals from tree-based models improved the prediction accuracy significantly when a limited number of predictors were included.

Many studies on digital mapping of SOC stocks focus on mineral topsoil 0–30 cm as there is usually the highest amount of SOC stored (e.g. Wiesmeier et al. 2012; Minasny et al. 2013; Yamashita et al. 2022). According to De Vos et al. (2015), the mineral layer 0–30 cm contains approximately 55–65% of the total SOC stock in forest soil profiles.

Various surveys of forest SOC content have been performed and various legacy data are available. However, different sampling designs, protocols and depth, different analytical methods, and data aging make the combination of data from different sources difficult and challenging (Borůvka et al. 2018; Bai & Fernandez 2020).

The aim of this study was to analyse the importance of environmental predictors for forest SOC stock prediction at the regional and national scale in the Czech Republic and to compare relative importance of the predictors in contrasting subsets of the national forest soil database compiled from several large-scale soil surveys.

4.2. Materials and Methods

4.2.1. Study area and soil data

This study is done on the whole forested area of the Czech Republic, belonging to the temperate forest zone. The country has an elevation ranging from 115 to 1602 m a.s.l. Mean annual temperatures are in the range from 1 to 10°C, with mean annual precipitation ranging between 400 and 1400 mm. Forests cover 26,551 km^2, forming 34.2% of the total country area. The Czech Republic is divided into 41 natural forest areas (NFA, http://www.uhul.cz/what-we-do/regional-plans-offorest- development). These spatially compact areas are rather homogeneous territories defined on the basis of geological, climatic, orographic and phytogeographical conditions.

A database of forest soil data from the years 2000 to 2020 was compiled from several resources: (i) National Forest Inventory (NFI) done by the Forest Management Institute (FMI, Forest Management Institute 2007); (ii) Data from permanent typological areas collected also by the FMI; (iii) Forest Soil Monitoring (FSM) done by the Central Institute for Supervising and Testing in Agriculture (Fiala et al. 2013); (iv) Data originating from the international projects ICP Forest and BioSoil (Lorenz & Becher 2012; Šrámek et al. 2013). As the surveys used different methodology and different sampling depths or horizons, the data were recalculated to the topsoil mineral layer 0–30 cm using weighted average. SOC content was mostly determined by oxidimetric method; comparability of other methods used in the surveys was tested. SOC stocks were calculated from the SOC content and bulk density (BD). Where the BD was not available, an estimate of BD was calculated using the model by Honeysett and Ratkowsky (1989):

$$BD = 1/(0.564 + 0.0556 \times OM)\ (g/cm^3)$$

Where; OM (organic matter) = 1.724 × SOC (%).

Rock fragments were not taken into account as this information was not available on all sites and, moreover, the accuracy of rock fragment content is generally low. In total, SOC stock values at the 0–30 cm depth were collected from 7338 forest stands all over the country, though the spatial distribution is not even and there are some gaps (see Figure 4.1D).

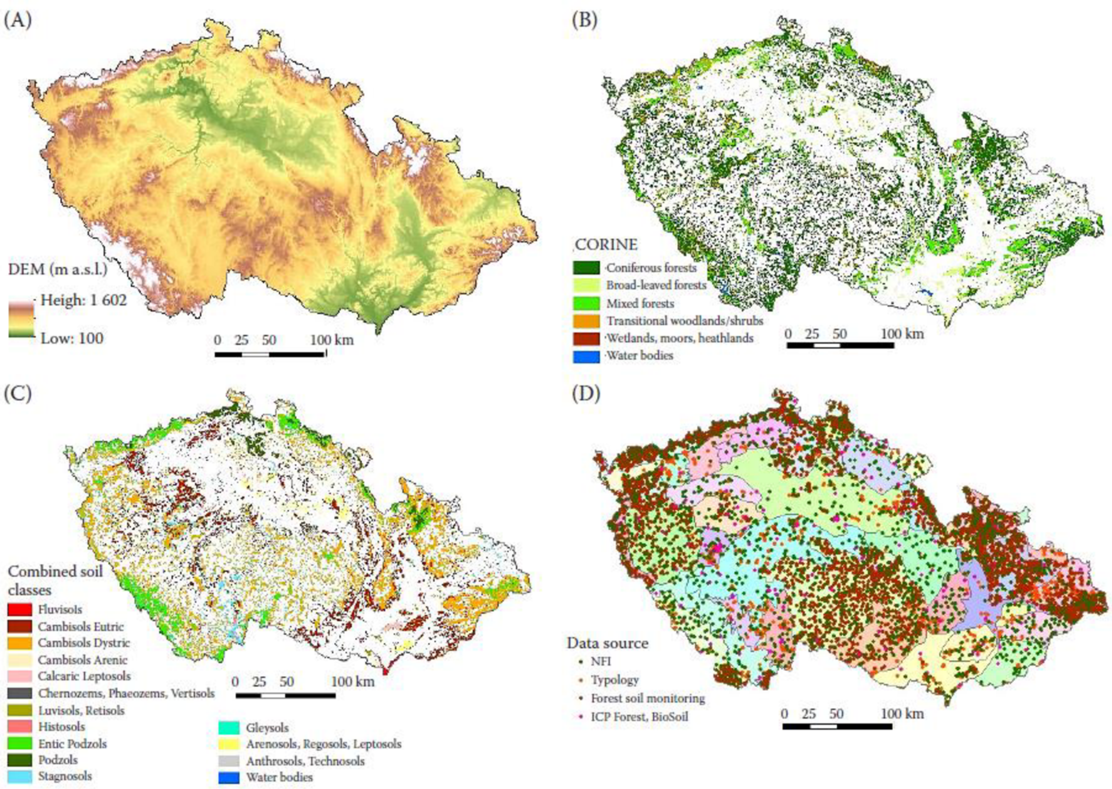


Figure 4. 1. Map of the Czech Republic with digital elevation model (DEM) (A), forest types (B), combined soil classes (C) and sampling points in the natural forest areas (D)

Terrain data were extracted from the digital elevation model (DEM) ArcČR®500 with resolution 200 m (ARCDATA PRAHA, ZÚ, ČSÚ, 2016; Figure 4.1A). Secondary terrain characteristics were calculated using Terrain Analysis Toolbox in SAGA GIS 2.1.4 (Conrad et al. 2015). The following terrain attributes were determined: elevation (m a.s.l.), slope, aspect (cos and sin), planar and profile curvatures, convergence index, catchment area, valley depth, relative slope position (RSP), channel network base level (CNBL), channel network distance (CND), topographic wetness index (TWI), LS factor (LSF), and analytical hillshade.

Soil classes were obtained from the Czech soil information system PUGIS at the resolution 1 : 250,000 (Kozák et al. 1996). The individual classes were grouped into 13 groups (see Table 4.3). While some soil classes were grouped to larger sets as they are less represented in forests (like Chernozems, Phaeozems and Vertisols), or have similar properties (like Luvisols and Retisols), the most abundant Cambisols forming in total more than 50% of the country were divided into 3 subclasses (mostly Eutric, Dystric and Arenic Cambisols; Figure 4.1C). Mean annual precipitation and temperatures were obtained from the database WorldClim.org at resolution 1 km (Fick & Hijmans 2017). Land cover/land-use categories, particularly forest types (deciduous/mixed/coniferous) were obtained from the database CORINE Land Cover 2018 (EEA 2018) at resolution of 100 m (Figure 4.1B). Forest typology (Viewegh et al. 2003) information on stands (forest vegetation zones – FVZ, and edaphic series) were obtained from the map of forest typology at scale 1 : 10,000 (ÚHÚL 2019).

4.2.2. Model selection, calibration and validation

Several model types were tested for SOC stock prediction, namely artificial neural networks, boosted regression trees, random forests (RF), and multivariate adaptive regression splines. Based on the results, and taking into account its common utilization, robustness to model overfitting and intercorrelation of predictors, and its ability to quantify relative predictor importance, the method of random forests (Breiman 2001) was chosen. 70% of data were used for model calibration, 30% for model validation. Index of determination (R^2) and root mean square error (RMSE) of validation were used for model performance evaluation.

4.3. Results and Discussion

4.3.1. General data description and predictor selection

The calculated SOC values in the depth 0–30 cm ranging from 0.07 to 38.59 kg/m^2, with a mean of 10.30 kg/m^2 (Table 4.1) correspond to values compiled by Lal (2005) for temperate forests, as well as those reported for Germany and other Central European countries (Wiesmeier et al. 2012), Slovakia (Priwitzer et al. 2009), Austria (Baumgarten et al. 2021), Spain (Calvo de Anta et al. 2020), or EU (De Vos et al. 2015). Prietzel and Christophel (2014) found slightly lower values in mineral topsoils in German Alps, which may be caused by higher elevations and consequently higher proportion of SOC in forest floor, and by the rock fragments that were not taken into account in our study. Lower SOC stock values were found also in Russian forests (Osipov et al. 2021) or in Hesse, Germany (Heitkamp et al. 2021).

Table 4. 1. Basic statistical parameters of soil organic carbon (SOC) stock dataset (in kg/m^2, layer 0–30 cm)

Parameter	SOC stock
Count	7338
Mean	10.30
Median	10.05
Geometric mean	9.06
Variance	22.04
SD	4.69
CV (%)	45.59
Standard error	0.05
Minimum	0.07
Maximum	38.59
Range	38.52
Lower quartile	6.53
Upper quartile	13.65
Skewness	0.33
Kurtosis	-0.33

SD: standard deviation, and CV: coefficient of variation

Correlation analysis showed that SOC stocks are positively correlated with altitude (r = 0.438; Table 4.2), forest vegetation zones (0.413) and mean annual precipitation (0.347), and negatively correlated with average annual temperature (–0.425). An increase of SOC stocks with increasing altitudes was reported also by Bojko and Kabala (2017), but only to the altitude of 1000 m a.s.l. Above this level, the SOC stocks started to drop again. Decreasing SOC stocks with increasing altitudes above 900 m a.s.l. were found also by Tungalag et al. (2020) in Mongolia. Weak correlation of the other predictors with SOC stock does not necessarily mean that there are no relationships; there can be some, but not linear.

Table 4. 2. Correlation matrix of soil organic carbon (SOC) stocks and covariates for the whole dataset (layer 0–30 cm); the predictors finally used for model development are in bold in the first column

Variable	SOC stock	FVZ	Altitude	Anal. hillshade	Slope	Aspect (sin)	Aspect (cos)	Planar curv.	Prof. curv.	Conv. index	Catch. area	TWI	LSF	CNBL	CND	Valley depth	RSP	Aver. temp.	Ann. prec.
SOC stock	1.000	0.413	0.438	0.241	0.199	−0.021	−0.004	−0.012	−0.007	0.037	−0.001	−0.061	0.221	0.210	0.162	0.171	0.138	−0.425	0.347
FVZ	0.413	1.000	0.900	0.202	0.229	0.047	−0.015	0.012	0.026	0.054	−0.061	−0.235	0.215	0.659	−0.027	−0.035	0.413	−0.824	0.633
Altitude	0.438	0.900	1.000	0.186	0.295	0.017	−0.045	0.040	0.068	0.132	−0.066	−0.299	0.259	0.700	−0.088	−0.108	0.508	−0.889	0.718
Analytical hillshade	0.241	0.202	0.186	1.000	−0.026	0.302	−0.317	−0.067	−0.062	0.045	0.022	0.210	0.067	−0.052	0.288	0.329	0.000	−0.227	0.158
Slope	0.199	0.229	0.295	−0.026	1.000	0.017	0.027	0.070	0.003	−0.009	−0.047	−0.575	0.923	0.178	0.168	0.151	0.147	−0.353	0.310
Aspect (sin)	−0.021	0.047	0.017	0.302	0.017	1.000	0.011	−0.012	−0.001	0.004	0.013	−0.053	0.011	0.000	−0.005	−0.013	0.043	−0.053	0.060
Aspect (cos)	−0.004	−0.015	−0.045	−0.317	0.027	0.011	1.000	0.017	−0.017	−0.007	−0.038	−0.012	0.011	−0.045	−0.028	−0.029	−0.013	0.040	−0.005
Planar curv.	−0.012	0.012	0.040	−0.067	0.070	−0.012	0.017	1.000	0.521	0.339	−0.010	−0.280	−0.063	−0.002	−0.066	−0.138	0.218	−0.035	0.022
Profile curv.	−0.007	0.026	0.068	−0.062	0.003	−0.001	−0.017	0.521	1.000	0.178	−0.012	−0.190	−0.080	0.017	−0.064	−0.187	0.312	−0.063	0.042
Conv. index	0.037	0.054	0.132	0.045	−0.009	0.004	−0.007	0.339	0.178	1.000	−0.051	−0.332	−0.095	−0.019	−0.232	−0.244	0.349	−0.111	0.075
Catch. area	−0.001	−0.061	−0.066	0.022	−0.047	0.013	−0.038	−0.010	−0.012	−0.051	1.000	0.227	0.013	−0.056	0.033	0.034	−0.084	0.070	−0.042
TWI	−0.061	−0.235	−0.299	0.210	−0.575	−0.053	−0.012	−0.280	−0.190	−0.332	0.227	1.000	−0.357	−0.201	0.179	0.226	−0.470	0.308	−0.251
LSF	0.221	0.215	0.259	0.067	0.923	0.011	0.011	−0.063	−0.080	−0.095	0.013	−0.357	1.000	0.154	0.270	0.272	0.031	−0.323	0.282
CNBL	0.210	0.659	0.700	−0.052	0.178	0.000	−0.045	−0.002	0.017	−0.019	−0.056	−0.201	0.154	1.000	−0.025	−0.057	0.221	−0.753	0.682
CND	0.162	−0.027	−0.088	0.288	0.168	−0.005	−0.028	−0.066	−0.064	−0.232	0.033	0.179	0.270	−0.025	1.000	0.943	−0.478	−0.017	0.073
Valley depth	0.171	−0.035	−0.108	0.329	0.151	−0.013	−0.029	−0.138	−0.187	−0.244	0.034	0.226	0.272	−0.057	0.943	1.000	−0.601	0.022	0.036
RSP	0.138	0.413	0.508	0.000	0.147	0.043	−0.013	0.218	0.312	0.349	−0.084	−0.470	0.031	0.221	−0.478	−0.601	1.000	−0.489	0.352
Aver. temp.	−0.425	−0.824	−0.889	−0.227	−0.353	−0.053	0.040	−0.035	−0.063	−0.111	0.070	0.308	−0.323	−0.753	−0.017	0.022	−0.489	1.000	−0.744
Annual prec.	0.347	0.633	0.718	0.158	0.310	0.060	−0.005	0.022	0.042	0.075	−0.042	−0.251	0.282	0.682	0.073	0.036	0.352	−0.744	1.000

FVZ – forest vegetation zones; TWI – topographic wetness index; LSF – LS factor; CNBL – channel network base level; CND – channel network distance; RSP – relative slope position; correlations at $P \leq 0.05$ are in bold

The correlation analysis showed also mutual relationships between the predictors. Thanks to the large dataset, even weak relationships are significant. Though RF model is not too sensitive to interrelations of predictors, we removed from further model calibration the predictors strongly correlated with other predictors to avoid redundant information in the model input. Finally, only seven continuous auxiliary variables were retained: annual precipitation, analytical hillshade, LS factor, catchment area, profile curvature, convergence index, and channel network distance. Three categorical ones were added: combined soil classes, edaphic series indicating trophic conditions and thus indirectly reflecting soil and geological conditions, and forest type. These ten predictors were used in all further models and their relative importance was evaluated.

4.3.2. SOC stocks prediction in natural forest areas

Separate models for SOC stock prediction were developed for individual NFA if the number of sampling points was sufficient, or for groups of two or a few neighbouring NFA that were similar. The NFA can correspond to the soil-landscape systems described by Mulder et al. (2015) who concluded that these systems have homogeneous conditions with respect to the combination of SOC controlling factors. This may explain why the prediction in some of these NFA was more successful than the groups defined by altitude ranges as shown further, or than the whole national model; the highest R^2 was 0.564, the lowest RMSE 2.31 kg/m^2. However, prediction accuracy for some other NFA was rather poor (minimum R^2 0.001, highest RMSE 4.53 kg/m^2). Similar results were reported by Hounkpatin et al. (2021) after comparison of national model with local (regional) models. Though the prediction accuracy generally improves (R^2 increases and RMSE decreases) with increasing size of the dataset (Figure 4.2), there are large datasets with poor models, and, in contrast, small datasets with good prediction accuracy. Moreover, though there are different combinations of important predictors for lower NFA and higher NFA, there is not a consistent trend of a better model performance in any altitude group of NFA. To analyse the different combination of important predictors at different altitudes, and to avoid criticism for different size of the groups, we divided the whole national dataset to four equal groups according to altitudes.

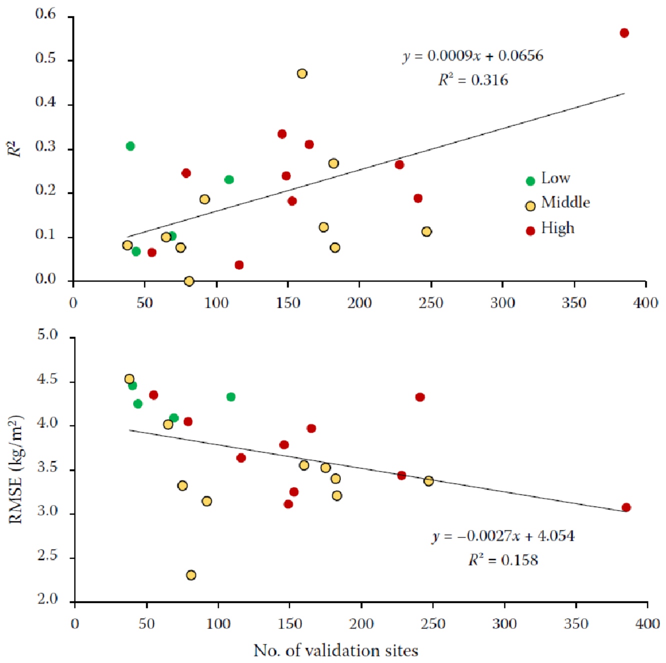


Figure 4. 2. Relationship between model performance measures (R^2 and RMSE of validation) and the number of validation sites for regional models on individual NFA or groups of several neighbouring NFA

In Figure 4.2, the sets are divided to groups according to the prevailing altitude range (lower/middle/higher), RMSE is root mean square error, and NFA is natural forest areas.

4.3.3. SOC stocks prediction in different altitude ranges

Figure 4.3 shows that the relative importance of predictors differs between different altitudes. At the first group with the lowest altitudes, there is the strongest effect of edaphic series, followed by combined soil classes, catchment area, annual precipitation, and analytical hillshade. Edaphic series indicate the trophic state of the stands, which definitely has a strong effect on SOC accumulation. The effect of soil classes is important because there is a strong variation of soil types in group 1, as it is shown by higher level of pedodiversity (Vacek et al. 2020), and also there are significant differences in SOC stock between soil types as confirmed by analysis of variance (ANOVA, Table 4.3). The highest stocks are in Fluvisols, which corresponds to the general features of this class, and in Calcareous Leptosols (mainly Rendzinas), where soil organic matter is stabilized by carbonates. However, Ostrowska et al. (2010) stated that the SOC accumulation in the profile is to a greater extent affected by the site type and stand age than by the soil type. In contrast, rather low importance was found at low altitudes for most relief-related predictors. Tziachris et al. (2019) also reported that terrain-based covariates have the least importance in flatness area. Only analytical hillshade, which is a terrain parameter, but with a strong relationship to the extent of solar radiation reaching the stand, has similar importance in all the altitude groups, as the sunlight undoubtedly influences organic matter production, decomposition and accumulation.

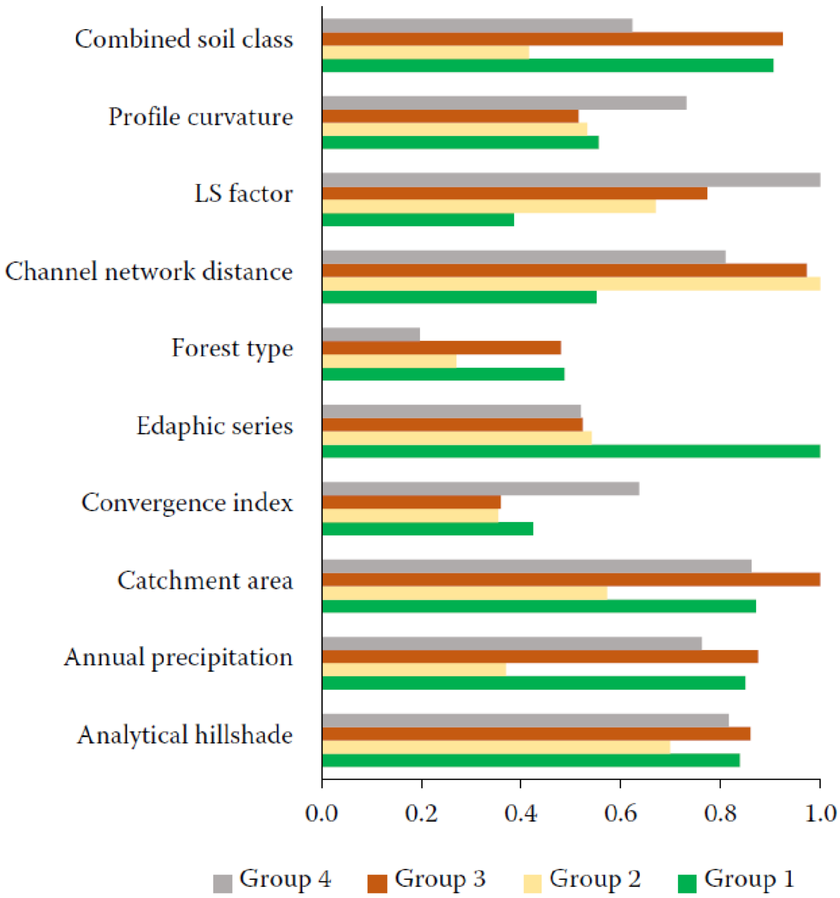


Figure 4. 3. Relative importance of predictors for four altitude groups of equal size according to the altitude

Table 4. 3. Basic characteristics of four altitude classes and mean soil organic carbon (SOC) stocks in the layer 0–30 cm (in kg/m^2) in separate soil class and forest type subsets; number of sampling points in each subset is given in parentheses

	Group 1	Group 2	Group 3	Group 4
Altitude range (m a.s.l.)	145–421	421–550	550–748	748–1479
Mean annual temperature range (°C)	2.7–9.5	2.6–9.5	2.9–8.9	1.3–8.5
Annual precipitation range (mm)	470–1157	494–1233	494–1175	519–1318
Mean C stock	8.48 (1 836)	8.54 (1 836)	10.44 (1 836)	13.73 (1 836)
Combined soil classes				
Chernozems, Phaeozems, Vertisols	9.16 (44)^{bc}	–	–	–
Fluvisols	11.77 (91)^{e}	–	–	–
Cambisols Eutric	8.40 (349)^{abc}	9.59 (300)^{d}	11.87 (284)^{cd}	13.80 (146)^{bc}
Cambisols Dystric	7.87 (557)^{a}	8.42 (1 163)^{bc}	9.78 (1 100)^{b}	12.93 (281)^{a}
Cambisols Arenic	8.33 (312)^{ac}	9.04 (85)^{cd}	10.19 (21)^{abc}	–
Calcaric Leptosols	11.27 (37)^{de}	12.87 (8)^{e}	–	–
Luvisols, Retisols	8.02 (272)^{ac}	7.39 (56)^{ab}	8.89 (1)^{abcde}	15.20 (2)^{abcd}
Histosols	5.66 (3)^{abc}	7.25 (9)^{abcd}	10.78 (13)^{abcd}	12.40 (149)^{a}
Entic Podzols	–	10.80 (11)^{cde}	12.48 (218)^{de}	13.84 (664)^{b}
Podzols	–	–	13.55 (26)^{e}	14.31 (575)^{c}
Stagnosols	9.15 (151)^{b}	7.31 (189)^{a}	9.03 (129)^{a}	18.26 (7)^{d}
Gleysols	9.01 (1)^{abcde}	9.35 (12)^{abcde}	9.88 (44)^{ab}	11.77 (12)^{ab}
Technosols	9.54 (19)^{abcd}	12.42 (3)^{cde}	–	–
F ratio	9.65	6.65	19.22	7.28
P	< 0.001	< 0.001	< 0.001	< 0.001
Shannon index of pedodiversity (relative)	1.846 (0.780)	1.192 (0.518)	1.275 (0.580)	1.485 (0.714)
Forest types				
Coniferous	7.61 (668)^{a}	8.10 (1 172)^{a}	10.04 (1 318)^{a}	13.69 (1 476)^{a}
Mixed	8.20 (303)^{b}	8.99 (269)^{b}	11.04 (264)^{b}	13.91 (230)^{a}
Deciduous	9.27 (859)^{c}	9.57 (391)^{b}	11.85 (254)^{c}	13.93 (130)^{a}
F ratio	29.96	20.61	23.54	0.44
P	< 0.001	< 0.001	< 0.001	0.645
Prediction results (validation subset)				
R^2	0.140	0.207	0.240	0.093
RMSE	3.96	3.62	3.71	3.76

Identical letters in each column indicate homogeneous groups according to ANOVA at $P \leq 0.05$; RMSE – root mean square error

In group 2, the biggest importance was achieved for channel network distance, followed by other terrain characteristics like LS factor, catchment area or profile curvature. The importance of forest type is rather low, as this group is dominated by coniferous forests and the difference between broadleaved and mixed forests is not significant. The low importance of soil classes is probably caused by the dominance of Cambisols in this group (reflected by the lowest pedodiversity), and even if there are three sets of Cambisols distinguished, they do not differ much in SOC stocks.

The higher altitudes, groups 3 and 4, have generally even more heterogenous relief, and therefore the importance of relief-related predictors is rather high. Similarly, Ellili et al. (2019)

found that slope and elevation are the most important covariates for predicting SOC. Soil classes are very important predictor in group 3 as there are Cambisols and Podzols that differ in SOC stocks. Group 4 is dominated by Podzols and therefore the importance of soil classes as predictor is again smaller. The highest SOC stocks at higher altitudes are in Podzols which corresponds to results reported by Bojko and Kabala (2017), and in Stagnosols where water saturation reduces mineralization process. The importance of forest type is still rather small, as the forests are dominated by conifers, and moreover, in group 4 the SOC stock in mineral topsoil under broadleaved and mixed forests does not differ significantly from coniferous forests.

4.3.4. General discussion

The validation results of the models were mostly weak, with quite low R^2 values. Similarly, Yamashita et al. (2022) obtained R^2 value of 0.38 in spatial prediction of SOC stocks in forested areas of Japan. Even lower R^2 values were obtained by Ottoy et al. (2017), Hounkpatin et al. (2021), Nussbaum et al. (2014), Baltensweiler et al. (2021) and Hoffmann et al. (2014). The predictions overestimated low values and underestimated high values, creating thus much narrower range of values.

Similar result was obtained for Swedish forest soils by Hounkpatin et al. (2021). Much better prediction accuracy was obtained by Li et al. (2021) when using remote sensing indices as additional predictors. Another potential source of auxiliary information for SOC prediction can be found in soil spectroscopy (Gholizadeh et al. 2021). Using some covariates in a more detailed resolution can possibly improve the prediction. However, more detailed environmental covariates do not need necessarily lead to more accurate soil maps (Samuel-Rosa et al. 2015). An important part of the uncertainty in the models could have been introduced by combination of data from different surveys using different sampling designs, methods and approaches, by recalculation of the data to unified depth, and by uncertainty in bulk density estimation. Potential sources of errors and uncertainties in the assessment of forest SOC stocks from sample to continental scale are clearly reviewed and summarized by Vanguelova et al. (2016).

The importance of soil classes depends on the heterogeneity of soil cover (described for example by Shannon's index of pedodiversity), and also on the significance of difference between soil classes in SOC stocks. Surprisingly, the SOC stocks in Histosols were among the lowest. However, there are just a few sites with Histosols particularly in the first three altitude groups, so that it cannot be considered significant, either. It indicates rather some inconsistencies or errors in the database, in spite of numerous checks applied.

SOC stock in the depth of 0–30 cm is lower under coniferous (mainly spruce) forests than under broadleaved and mixed forests (Table 4.3); at lower altitudes this difference is significant. However, as the coniferous forests have usually thicker O horizons and larger SOC amounts are retained in the surface organic horizons (Kjønaas et al. 2021), the total SOC stock in the whole profile is generally bigger under coniferous forests than under broadleaved ones (Bojko & Kabala 2017; Nitsch et al. 2018). Nevertheless, Cremer et al. (2016) reported higher SOC stocks under coniferous forests even in the mineral topsoil. The dominance of coniferous forests, particularly at higher altitudes, and very similar SOC stock values in all forest types make forest type a less important predictor. A more detailed description of forest species composition might improve the prediction. The effect of climate on building SOC stocks was shown e.g. by Rial et al. (2017), Černý et al. (2020), or Calvo de Anta et al. (2020).

4.4. Conclusions

The study showed that the importance of environmental predictors in the models for SOC stock prediction can strongly differ between regions and altitudes. At lower altitudes, edaphic series and soil classes are strong predictors, while at higher altitudes the predictors related to topography become more important. The importance of soil classes depends on the pedodiversity level and on the difference in SOC stock between the soil classes distinguished. The contribution of forest types as predictor is limited when one type dominates. Collection and selection of influential covariates is a very important part of digital mapping of soil properties. It was found that better prediction results can be obtained in smaller, but consistent regions, like in some natural forest areas; however, in some NFA the models failed. It was also shown than even very exhaustive datasets used for modelling do not ensure highly accurate prediction. Data harmonization, transformation, standardization and recalculation bring additional uncertainty and error that are projected in developed prediction models and model estimates. Nevertheless, in spite of the uncertainties of the models, the prediction shows well the general trends and factors of SOC stock distribution, at least at the national scale (Figure 4.4).

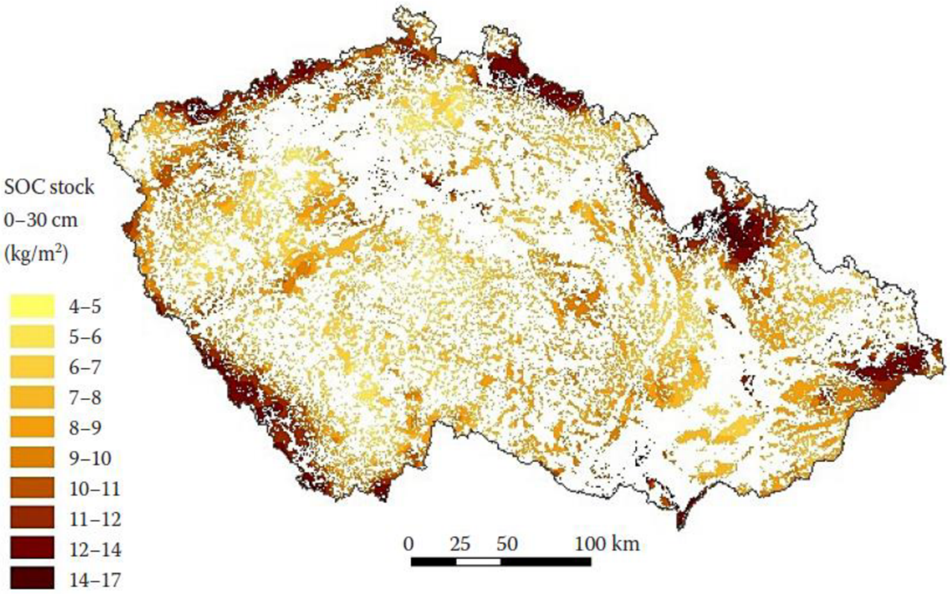


Figure 4. 4. Predicted soil organic carbon (SOC) stock values for the mineral topsoil (0–30 cm) of forest soils using random forest model (R^2 = 0.32, RMSE = 3.91 kg/m^2)

In Figure 4.4, the agricultural and other non-forest soils are masked by white colour and RMSE is root mean square error.

4.5. References

ARCDATA PRAHA, ZÚ, ČSÚ (2016): ArcČR 500 – Digital Geogrphic Database, Version 3.3. Available at www.arcdata. cz (accessed Oct 23, 2020). (in Czech)

Bai X., Fernandez I.J. (2020): Comparing publicly available databases to evaluate soil organic carbon in Maine, USA. Soil Science Society of America Journal, 84: 1722–1736.

Baltensweiler A., Walthert L., Hanewinkel M., Zimmermann S., Nussbaum M. (2021): Machine learning based soil maps for a wide range of soil properties for the forested area of Switzerland. Geoderma Regional, 27: e00437.

Baumgarten A., Haslmayr H.-P., Schwarz M., Huber S., Weiss P., Obersteiner E., Aust G., Englisch M., Horvath D., Leitgeb E., Foldal C., Rodlauer C., Bohner A., Spiegel H., Jandl R. (2021): Organic soil carbon in Austria – Status quo and foreseeable trends. Geoderma, 402: 115214.

Bojko O., Kabala C. (2017): Organic carbon pools in mountain soils – Sources of variability and predicted changes in relation to climate and land use changes. Catena, 149: 209–220.

Borůvka L., Penížek V., Kozák J., Marhoul A.M., Vacek O., Šrámek V., Žížala D., Poláková Š., Smatanová M., Fiala P. (2018): Combining and harmonizing soil data from different sources: Problems and approaches. In: Arrouays D., Savin I., Leenaars J., McBratney A.B. (eds.): GlobalSoil- Map. Digital Soil Mapping from Country to Globe. Leiden, CRC Press/Balkema: 29–34.

Breiman L. (2001): Random forests. Machine Learning, 45: 5–32.

Calvo de Anta R., Luís E., Febrero-Bande M., Galiñanes J., Macías F., Ortíz R., Casás F. (2020): Soil organic carbon in peninsular Spain: Influence of environmental factors and spatial distribution. Geoderma, 370: 114365.

Černý J., Pokorný R., Vejpustková M., Šrámek V., Bednář P. (2020): Air temperature is the main driving factor of radiation use efficiency and carbon storage of mature Norway spruce stands under global climate change. International Journal of Biometeorology, 64: 1599–1611.

Conrad O., Bechtel B., Bock M., Dietrich H., Fischer E., Gerlitz L., Wehberg J., Wichmann V., Böhner J. (2015): System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geoscientific Model Development, 8: 1991–2007.

Cremer M., Kern N.V., Prietzel J. (2016): Soil organic carbon and nitrogen stocks under pure and mixed stands of European beech, Douglas fir and Norway spruce. Forest Ecology and Management, 367: 30–40.

Chuman T., Oulehle F., Zajícová K., Hruška J. (2021): The legacy of acidic deposition controls soil organic carbon pools in temperate forests across the Czech Republic. European Journal of Soil Science, 72: 1780–1801.

De Vos B., Cools N., Ilvesniemi H., Vesterdal L., Vanguelova E., Carnicelli S. (2015): Benchmark values for forest soil carbon stocks in Europe: Results from a large scale forest soil survey. Geoderma, 251–252: 33–46.

Dobor L., Hlásny T., Rammer W., Barka I., Trombik J., Pavlenda P., Šebeň V., Štěpánek P., Seidl R. (2018): Postdisturbance recovery of forest carbon in a temperate forest landscape under climate change. Agricultural and Forest Meteorology, 263: 308–322.

EEA (2018): CORINE Land Cover (CLC) 2018, Version 20, European Environment Agency, Kopenhagen, Denmark. Available at https://www.eea.europa.eu/ data-and-maps/ data/corine-land-cover-accountinglayers (accessed Oct 23, 2020).

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.

Fiala P., Reininger D., Samek T., Němec P., Sušil A. (2013): Forest Nutrition Survey at the Territory of the Czech Republic 1996–2011. Brno, Central Institute for Supervising and Testing in Agriculture. (in Czech)

Fick S.E., Hijmans R.J. (2017): WorldClim 2: New 1 km spatial resolution climate surfaces for global land areas. International Journal of Climatology, 37: 4302–4315.

Forest Management Institute (2007): National Forest Inventory in the Czech Republic 2001–2004. Introduction, Methods, Results. Brandýs and Labem, Forest Management Institute.

Gholizadeh A., Viscarra Rossel R.A., Saberioon M., Borůvka L., Kratina J., Pavlů L. (2021): National-scale spectroscopic assessment of soil organic carbon in forests of the Czech Republic. Geoderma, 385: 114832.

Heitkamp F., Ahrends B., Evers J., Meesenburg H. (2021): Spatial 3D mapping of forest soil carbon stocks in Hesse, Germany. Journal of Plant Nutrition and Soil Science, 184: 635–656.

Hoffmann U., Hoffmann T., Johnson E.A., Kuhn N.J. (2014): Assessment of variability and uncertainty of soil organic carbon in a mountainous boreal forest (Canadian Rocky Mountains, Alberta). Catena, 113: 107–121.

Honeysett J., Ratkowsky D. (1989): The use of ignition loss to estimate bulk density of forest soils. European Journal of Soil Science, 40: 299–308.

Hounkpatin K.O.L., Stendahl J., Lundblad M., Karltun E. (2021): Predicting the spatial distribution of soil organic carbon stock in Swedish forests using a group of covariates and site-specific data. Soil, 7: 377–398.

Jonard M., Nicolas M., Coomes D.A., Caignet I., Saenger A., Ponette Q. (2017): Forest soils in France are sequestering substantial amounts of carbon. Science of the Total Environment, 574: 616–628.

Kjønaas O.J., Bárcena T.G., Hylen G., Nordbakken J.-F., Økland T. (2021): Boreal tree species change as a climate mitigation strategy: Impact on ecosystem C and N stocks and soil nutrient levels. Ecosphere, 12: e03826.

Kozák J., Němeček J., Jetmar M. (1996): The database of soil information system – PUGIS. Rostlinná Výroba, 42: 529–534.

Lal R. (2005): Forest soils and carbon sequestration. Forest Ecology and Management, 220: 242–258.

Lamichhane S., Kumar L., Wilson B. (2019): Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. Geoderma, 352: 395–413.

Li Y., Zhao Z., Wei S., Sun D., Yang Q., Ding X. (2021): Prediction of regional forest soil nutrients based on Gaofen-1 remote sensing data. Forests, 12: 1430.

Lorenz M., Becher G. (eds.) (2012): Forest Condition in Europe, 2012. Technical Report of ICP Forests. Work Report of the Thünen Institute for World Forestry 2012/1. Hamburg, ICP Forests.

Martin M.P., Orton T.G., Lacarce E., Meersmans J., Saby N.P.A., Paroissien J.B., Jolivet C., Boulonne L., Arrouays D. (2014): Evaluation of modelling approaches for predicting the spatial distribution of soil organic carbon stocks at the national scale. Geoderma, 223–225: 97–107.

McBratney A.B., Mendonca Santos M.L., Minasny B. (2003): On digital soil mapping. Geoderma, 117: 3–52.

Miller B.A., Koszinski S., Wehrhan M., Sommer M. (2015): Impact of multi-scale predictor selection for modeling soil properties. Geoderma, 239–240: 97–106.

Minasny B., McBratney A.B., Malone B.P., Wheeler I. (2013): Digital mapping of soil carbon. Advances of Agronomy, 118: 1–47.

Mulder V.L., Lacoste M., Martin M.P., Richer-de-Forges A., Arrouays D. (2015): Understanding large-extent controls of soil organic carbon storage in relation to soil depth and soil-landscape systems. Global Biogeochemical Cycles, 29: 1210–1229.

Nitsch P., Kaupenjohann M., Wulf M. (2018): Forest continuity, soil depth and tree species are important parameters for SOC stocks in an old forest (Templiner Buchheide, northeast Germany). Geoderma, 310: 65–76.

Nussbaum M., Papritz A., Baltensweiler A., Waltheart L. (2014): Estimating soil organic carbon stocks of Swiss forest soils by robust external-drift kriging. Geoscientific Model Development, 7: 1197–1210.

Osipov A.F., Bobkova K.S., Dymov A.A. (2021): Carbon stocks of soils under forest in the Komi Republic of Russia. Geoderma Regional, 27: e00427.

Ostrowska A., Porebska G., Kanafa M. (2010): Carbon accumulation and distribution in profiles of forest soils. Polish Journal of Environmental Studies, 19: 1307–1315.

Ottoy S., De Vos B., Sindayihebura A., Hermy M., Van Orshoven J. (2017): Assessing soil organic carbon stocks under current and potential forest cover using digital soil mapping and spatial generalisation. Ecological Indicators, 77: 139–150.

Prietzel J., Christophel D. (2014): Organic carbon stocks in forest soils of the German Alps. Geoderma, 221–222: 28–39.

Priwitzer T., Pajtík J., Konôpka B., Ištoňa J., Pavlenda P. (2009): Carbon allocation in forest ecosystem – Case studies from young pine stands. Lesnícky Časopis – Forestry Journal, 55: 239–250.

Rial M., Martínez Cortizas M., Rodríguz-Lado L. (2017): Understanding the spatial distribution of factors controlling topsoil organic carbon content in European soils. Science of the Total Environment, 609: 1411–1422.

Samuel-Rosa A., Heuvelink G.B.M., Vesques G.M., Anjos L.H.C. (2015): Do more detailed environmental covariates deliver more accurate soil maps? Geoderma, 243–244: 214–227.

Šrámek V., Jurkovská L., Fadrhonsová V., Hellebrandová- Neudertová K. (2013): Forest soil chemistry in relation to the forest site classification categories used in the Czech Republic – Results of the EU "BioSoil" forest soil monitoring project. Zprávy Lesnického Výzkumu, 58: 314–323.

Tungalag M., Gerelbaatar S., Lobanov A.I. (2020): Organic carbon stocks in the forest soils of Northern Mongolia. Lesnoy Zhurnal, 2020: 169–176.

Tziachris P., Aschonitis V., Chatzistathis T. (2019): Assessment of spatial hybrid methods for predicting soil organic matter using DEM derivatives and soil parameters. Catena, 174: 206–216.

ÚHÚL (2019): Typological map 2019. Ústav pro hospodářskou úpravu lesů. Available at http://geoportal.uhul.cz/ (in Czech)

Vacek O., Vašát R., Borůvka L. (2020): Quantifying the pedodiversity- elevation relations. Geoderma, 373: 114441.

Vanguelova E.I., Bonifacio E., De Vos B., Hoosbeek M.R., Berger T.W., Vesterdal L., Armolaitis K., Celi L., Dinca L., Kjønaas O.J., Pavlenda P., Pumpanen J., Püttsepp Ü., Reidy B., Simončič P., Tobin B., Zhiyanski M. (2016): Sources of errors and uncertainties in the assessment of forest soil carbon stocks at different scales – Review and recommendations. Environmental Monitoring and Assessment, 188: 630.

Viewegh J., Kusbach A., Mikeska M. (2003): Czech forest ecosystem classification. Journal of Forest Science, 49: 85–93.

Wen D., He N. (2016): Forest carbon storage along the north-south transect of eastern China: Spatial patterns, allocation, and influencing factors. Ecological Indicators, 61: 960–967.

Were K., Bui D.T., Dick Ø.B., Singh B.R. (2015): A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. Ecological Indicators, 52: 394–403.

Wiesmeier M., Spörlein P., Geuß U., Hangen E., Haug S., Reischl A., Schilling B., von Lützow M., Kögel-Knabner I. (2012): Soil organic carbon stocks in southeast Germany (Bavaria) as affected by land use, soil type and sampling depth. Global Change Biology, 18: 2233–2245.

Yamashita N., Ishizuka S., Hashimoto S., Ugawa S., Nanko K., Osone Y., Iwahashi J., Sakai Y., Inatomi M., Kawanishi A., Morisada K., Tanaka N., Aizawa S., Imaya A., Takahashi M., Kaneko S., Miura S., Hirai K. (2022): National-scale 3D mapping of soil organic carbon in a Japanese forest considering microtopography and tephra deposition. Geoderma, 406: 115534.

CHAPTER 5

Digital Soil Mapping using Machine Learning-Based Methods to Predict Soil Organic Carbon in Two Different Districts in the Czech Republic

Citation:

Nozari S., Pahlavan-Rad M. R., Brungard C., Heung B., and Borůvka L. (2023): Digital Soil Mapping using Machine Learning-Based Methods to Predict Soil Organic Carbon in Two Different Districts in the Czech Republic. Soil & Water Res., XX: 00–00. (in press)

Abstract: Soil organic carbon (SOC) is an important soil characteristic as well as a way how to mitigate climate change. Information on its content and spatial distribution is thus crucial. Digital soil mapping (DSM) is a suitable way to evaluate spatial distribution of soil properties thanks to its ability to obtain accurate information about soil. This research aims to apply machine learning algorithms using various environmental covariates to generate digital SOC maps for mineral topsoils in the Liberec and Domažlice districts, located in the Czech Republic. The soil class, land cover, and geology maps as well as terrain covariates extracted from the digital elevation model and remote sensing data were used as covariates in modelling. The spatial distribution of SOC was predicted based on its relationships with covariates using random forest (RF), cubist, and quantile random forest (QRF) models. Results of the RF model showed that land cover (vegetation) and elevation were the most important environmental variables in the SOC prediction in both districts. The RF had better efficiency and accuracy than the cubist and QRF to predict SOC in both districts. The greatest R^2 value (0.63) was observed in the Domažlice district using the RF model. However, cubist and QRF showed appropriate performance in both districts, too.

Keywords: cubist; digital soil mapping; quantile random forest; random forest; soil organic carbon

5.1. Introduction

Although human population growth affects soil, soil quality must be maintained to ensure human survival (Pieri 1992; Brevik 2013). Soil organic carbon (SOC) is one of the most important indicators of soil quality and constitutes the largest terrestrial pool of bound carbon (Lal et al. 2021; Victoria et al. 2012). Many studies have been conducted to identify suitable methods to model and monitor SOC due to its substantial influence on atmospheric carbon dioxide (CO_2), which affects climate change (Selvaradjou et al. 2007). However, a high-resolution spatial prediction of SOC is needed to inform sustainable soil management practices and to assess the impacts of land-use.

Numerous studies have been conducted on the prediction of SOC distribution using digital soil mapping (DSM) (Nikou and Tziachris 2022); nevertheless, detailed aspects such as specific features, input data, and models used for spatial prediction in DSM have not been fully compiled for SOC in forest and agricultural soil (Minasny and McBratney 2016). Similarly in the Czech Republic, there are examples of producing high-resolution maps of SOC at the local, regional, or national scale (e.g. Gholizadeh et al. 2018), while there are not adequate studies considering the prediction of SOC in the Liberec and Domažlice districts. Additionally, there is no feasible study elucidating this approach, despite the region's active engagement in agriculture production. In this research, many different aspects of SOC are spatially evaluated concerning specific characteristics, input data, and models for SOC.

Therefore, this study aims to compare three models including random forest (RF), cubist, and quantile random forest (QRF) to assess their prediction accuracy, important variables, and spatial predictions of SOC as well as compare prediction uncertainty maps and suggest the best model that can be used to predict SOC in the Liberec and Domažlice districts in the Czech Republic.

5.1.1. Literature Review

This section reviews information about SOC distribution as well as applications of DSM and machine learning in the prediction of SOC distribution.

5.1.1.1. SOC Distribution

One of the significant parameters that influence SOC distribution and explain the variation in SOC is topography as it is related to the extent of soil erosion, sediment yield, and the rate of incoming solar radiation. In addition, changes in other soil properties (such as changes caused by cultivation) influence the SOC content prediction by affecting aggregate stability, porosity, and bulk density. It has been also found that land-use, land management, vegetation, elevation, slope, rainfall, soil type, and wetness index are the most effective predictors of SOC (Wiesmeier et al. 2019, Mosleh et al. 2016, Badia et al. 2016, and Borůvka et al. 2022). Additionally, Nozari et al. (2020, 2023) showed that there is a clear relationship between SOC and environmental, particularly terrain parameters. Although changes in environmental variables can influence the SOC prediction accuracy, the direct relationship between variables and model accuracy is not straightforward. In other words, a reduction in variables in a specific model may either decrease or increase the accuracy of model, depending on the relationships between variables and model types (Heung et al. 2014).

5.1.1.2. DSM

Although traditional maps are still a major source of information on the distribution of SOC, the development of DSM offers better ways to generate such information in the Czech Republic

(Žížala et al. 2022). DSM provides essential tools to improve the understanding of the distribution of SOC for both forest and agricultural soils. It increases the efficiency of mapping process and provides a more detailed, accurate, and quantitative prediction of soil properties for different areas (Lorenzetti et al. 2015). Additionally, DSM has become a powerful tool for optimal decision-making in environmental and agricultural management by providing relevant soil information (McBratney et al. 2003). DSM integrates information from observed soil attributes with interdependent environmental covariates obtained from terrain analysis, geospatial data sources, and remote sensing images using geographical information systems (GIS) and machine learning (ML) to generate grid-based maps of different soil types and properties and predict the spatial distribution of soil properties using a quantitative framework (McBratney et al. 2003; Mulder et al. 2011).

5.1.1.3. Machine learning

ML is the self-adaptive method where a fitted pattern can be used to set prediction targets for new data. Brungard et al. (2015) reported that covariates selected by soil scientists familiar with the study area did not yield the most accurate models compared to covariates automatically selected by ML algorithms. Additionally, Borůvka et al. (2022) showed that even large datasets used for modelling do not guarantee highly accurate prediction. Although the number of studies using ML algorithms have been increasing, only a few studies have compared different learners and most studies are limited to the evaluation of a few common models such as random forest (Fathololoumi et al. 2020). QRF takes into account both landscape properties and the density, which is closer to the experience of a soil surveyor. QRF is based on the hypothesis that the clustering, optimizing the prediction of the mean, also optimizes the prediction of the other quantiles and the uncertainty. Although this has not been fully proven yet, Meinshausen (2006) showed that QRF clearly outperformed the quantile regression algorithms estimating each quantile separately in five different case studies (Vaysse et al. 2017).

5.2. Materials and Methods

An overall evaluation of the performance of RF, cubist, and QRF models for SOC mapping was conducted using R, version 3.5.1 (R Core Team 2018), to provide a framework for interpretation. The local uncertainty was assessed through a rigorous cross-validation approach. The evaluation consisted of comparing metrics of the performance, visual inspection, and interpretation of anomalies with geographic knowledge to figure out why the accuracy of the model for some locations in the landscape is not acceptable.

5.2.1. Study area and soil sampling

The Liberec district (989 km^2) is located in the northern Czech Republic (Figure 5.1) with elevations ranging from 210 to 1124 m above mean sea level and is covered by 47.2% agricultural land, 42.4% forest area, and less than 6% mix of agriculture and forests (Miko et al. 2009), as illustrated in Figure 5.3c. On the other hand, the Domažlice district (1123 km^2) is located in the western Czech Republic with elevations ranging from 383 to 1042 m and is covered by 53.0% agricultural land, 38.2% forest land, and less than 6% mix of agriculture and forests (Miko et al. 2009), as illustrated in Figure 5.4c.

Figure 5. 1. Location of the Liberec (red) and Domažlice (green) districts in the Czech Republic

The soil was classified according to the Czech Taxonomic Soil Classification System and WRB system (Němeček et al. 2011 and IUSS Working Group WRB 2015). Eight classes of mineral parent material including sedimentary rocks, acid granites and similar rocks, basalts, loess-like sediments, micaceous schists and phyllites, polygenetic loams, gneisses, alluvial (fluvial) and six major reference soil groups including Cambisols, Podzols, Gleysols, Stagnosols, Luvisols, and Fluvisols were identified in the Liberec district. Eight classes of mineral parent material including sedimentary rocks, acid granites and similar rocks, other mafic rocks, loess-like sediments, micaceous schists and phyllites, polygenetic loams, gneisses, alluvial (fluvial) sediments as well as five major reference soil groups including Cambisols, Gleysols, Stagnosols, Luvisols, and Fluvisols were observed in the Domažlice district (Němeček et al. 2011, IUSS Working Group WRB 2015).

In this study, 71 samples for the Liberec and 67 samples for the Domažlice districts were randomly collected in 2004 (Figure 5.2 and Table 5.1).

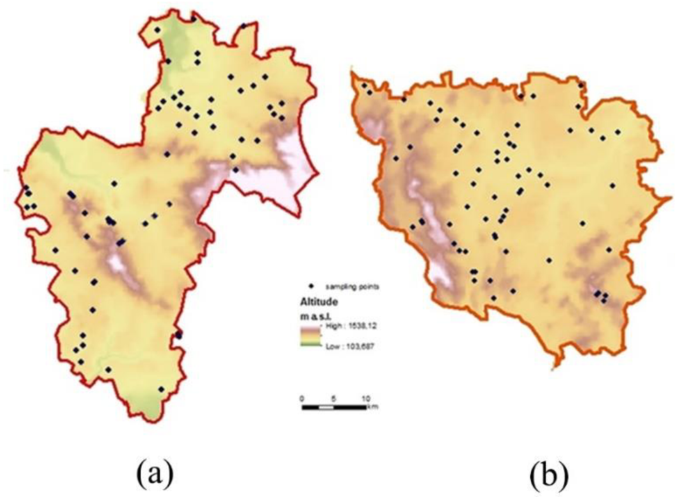


Figure 5. 2. Elevation maps and distribution of sampling locations maps in a) Liberec, and b) Domažlice districts with altitudes

Table 5. 1. Number of samples collected in different land-uses

District	Agricultural land	Forest area	Mix of agricultural and forest lands	Total
Liberec	35	22	14	71
Domažlice	35	20	12	67

The sampling depth was 0-30 cm because it represents the plow depth and SOC estimation in this depth is an important factor in farm management. In each location, soil was sampled to a depth of 30 cm using a steel soil auger after removing plant debris such as grass and twigs. In forest, the forest floor was also removed for consistency of samples across different land covers. The collected soil samples were stored in plastic bags and transferred to the laboratory for analysis. To measure SOC, soil samples were air-dried, grinded and sieved using a sieve with mesh size < 0.25 mm, and the SOC was determined through the oxidimetric modified Tyurin method (Pospíšil 1964).

5.2.2. Legacy data and auxiliary environmental covariates

Environmental covariates are essential in the DSM process and can be obtained from a combination of remotely sensed data, digital elevation model (DEM), or other geospatial sources (Lagacherie et al. 2006). Soil survey and soil mapping have a long tradition in the Czech

Republic. Various large scale point or polygon legacy soil data and maps are available in the country (Kozák et al. 1996, Němeček 2000) and Europe (Panagos et al. 2014). In this research, a DEM with a 100 m spatial resolution was obtained from the U.S. Geological Survey database (U.S. Geological Survey 2021). A suite of 15 topographic variables was computed using SAGA GIS 7.2.0 (Conrad et al., 2015). In addition, Normalized Difference Vegetation Index (NDVI) (Landsat TM image [USGS, 2021]), soil, geological (Kozák et al. 1996), and CORINE Land Cover maps (EEA 2018) were used as the predictors. The CORINE database contains four main categories including forest, arable land, pasture, and industrial areas (Figures 5.3 and 5.4). Table 5.2 presents a summary of the total 16 environmental variables used in this study. Borůvka et al. (2022) reported that the importance of environmental variables in the models for SOC stock prediction varies in different regions and altitudes.

Table 5. 2. Soil environmental covariates mostly derived from DEM (McBratney et al. 2003)

Soil-environmental covariates	Code	Significance related to soil development and properties
Elevation	Elev	Climate, vegetation, energy potential
Slope	S	Surface and subsurface flows, flow speed and erosion rate, precipitation, vegetation, geomorphology, soil water content, land-use capacity
Profile curvature	PC	Profile curvature is the rate of change of slope in a downslope direction. It characterizes changes in flow acceleration that may differentiate erosion and deposition zones in landscapes.
Plan curvature	Plan. Cur	Convergent/divergent flows, soil water content, soil characteristics, flow acceleration, erosion rate/ deposition, geomorphology
Length slope factor	LSF	Surface flow volume
Topographic wetness index	TWI	A measure of the topographic control on soil wetness
Valley depth	Va. Dep	Valley depth specifies soil characteristics, influencing composition and fertility, crucial for effective land management
Relative slope position	RSP	It is a measure of the percentage distance a location is from slope bottom to nearest ridge top, influencing drainage, erosion, and microenvironments
Convergence index	CI	It is calculated based on the aspect that shows the structure of the relief and flow convergence affecting water movement
Vertical distance to channel networks	VDCN	A grid provides information about the channel network, influencing drainage patterns and sediment transport
Channel network base level	CNBL	This grid output contains the interpolated channel network of base level elevations, defining landscape lowering and drainage efficiency
Total catchment area	Cat. Area	Expected runoff volume that determines water inflow and sediment transport

Normalized difference vegetation index of Landsat-4	NDVI	It reflects vegetation health and biomass
Geology map	Geology	Polygon map of soil parent material
Soil map	Soil	Polygon map of soil classes
Land-use map	Land-use	Polygon map of CORINE land cover categories that shows vegetation and human activities impacting soil

Vegetation indices are helpful in modeling SOC because the vegetation is the ultimate source of SOC. NDVI is a common unitless remote sensing index that uses the ratio between visible and near-infrared reflectance of vegetation cover. Additionally, it can estimate the green density of the area (Weier and Herring 2000). NDVI is calculated as:

$$NDVI = \frac{NIR - RED}{NIR + RED} \qquad Equ. 1$$

where NIR is the amount of image reflection in the near-infrared band and RED is the amount of image reflection in the red band. NDVI was calculated from a Landsat TM image with < 10% cloud cover. The image was taken in 1992 under clear weather conditions on the 9th of August for the Domažlice district and the 19th of September for the Liberec district. Two spectral bands were selected from Landsat Legacy TM including Band three, containing red reflectance, and Band four with infrared reflectance.

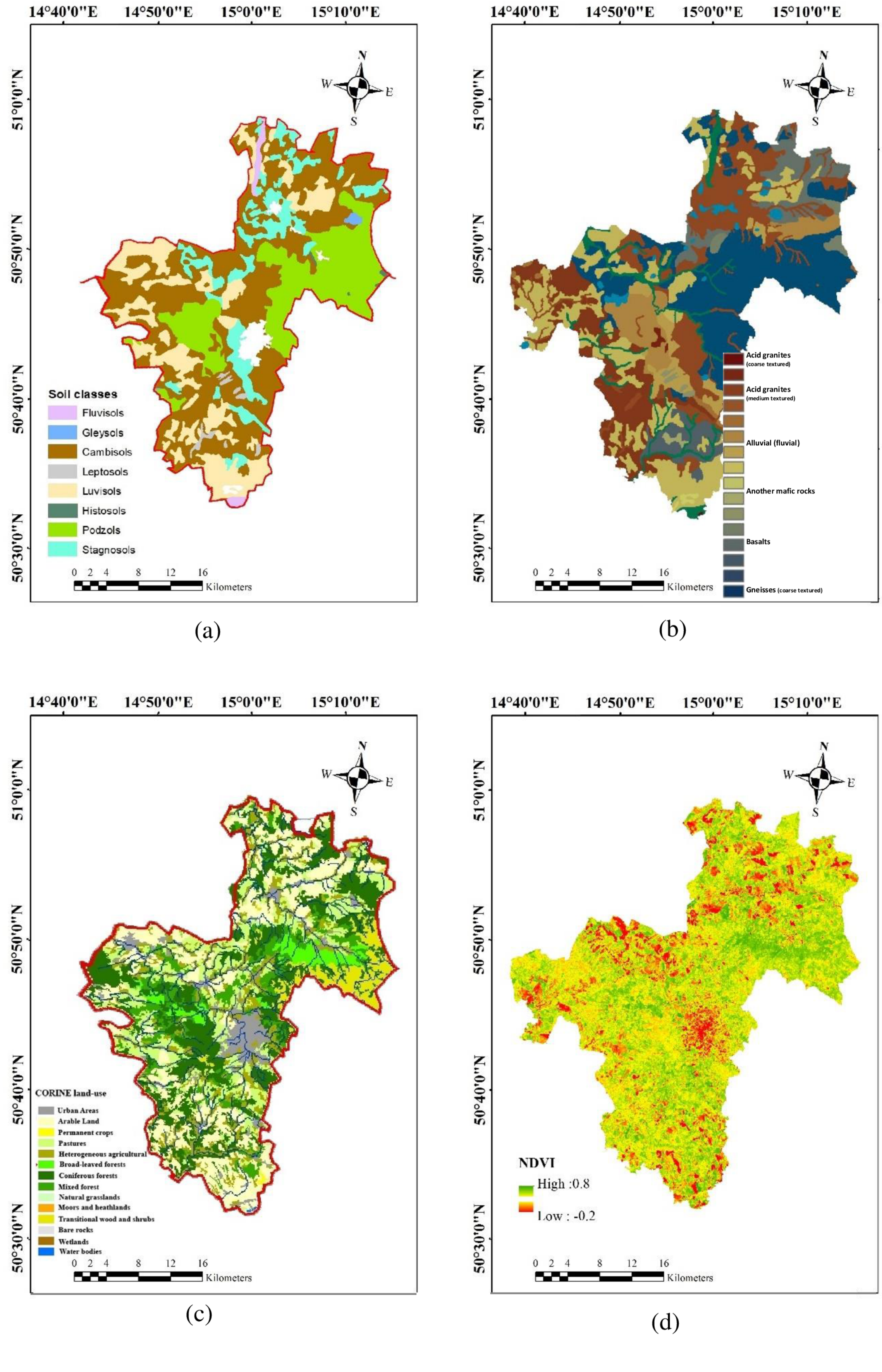


Figure 5. 3. a) Soil map, b) geology map, c) land cover, and d) NDVI for the Liberec district

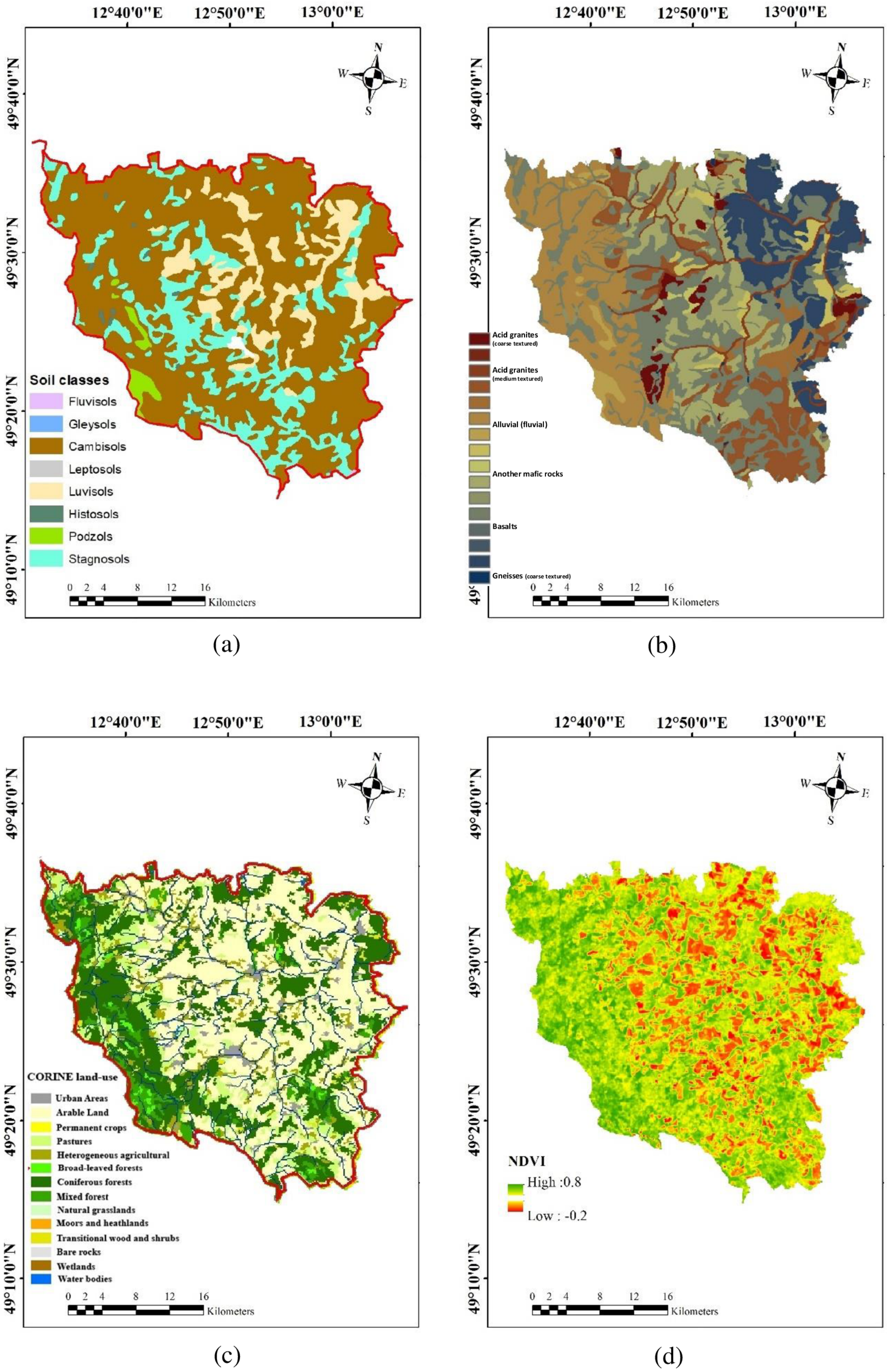


Figure 5. 4. a) Soil map, b) geology map, c) land cover, and d) NDVI for the Domažlice district

5.2.3. Basic statistical analyses

Statistical differences in mean values were computed by one-way analysis of variance (ANOVA) method using SPSS (SPSS version 11) and R, version 3.5.1 (R Core Team 2018). One-way ANOVA was conducted to evaluate the effects of landform types (slope) on soil properties (Duncan's test at the 5% level of significance). The SPSS analysis was also carried out to determine the correlation matrix between variables used in this study. In addition, multiple and linear regression coefficients were calculated to determine the relationships between auxiliary variables and SOC using R and SPSS.

5.2.4. Regression models

The models used in this study include a tree-based methods called RF, cubist, and QRF (Pahlavan-Rad et al. 2020). Many studies have demonstrated that RF has superior performance compared to other models (Brungard et al. 2015, Pahlavan-Rad et al. 2018b, and Zeraatpisheh et al. 2019). Indeed, RF is a modified and extended model of the regression tree model (as a basic idea) and it constructs a forest of low-correlation regression trees (Peters et al. 2007). However, the original implementation of RF was unable to produce spatial estimates of uncertainty. Therefore, QRF was introduced as an alternative to the RF learner, allowing users to leverage the model predictions from each tree of the RF to generate uncertainty estimates. On the other hand, cubist is a modification and extension of the basic classification tree idea (Quinlan 1993).

Although many studies have proved the high accuracy of the RF model, only a few works have shown that the performance of the RF model is not perfectly acceptable (Pouladi et al. 2019). In addition, one of the limitations of the RF model is that the soil properties may be overestimated (Pahlavan- Rad and Akbari Moghaddam 2018a).

5.2.4.1. Random Forest (RF)

RF algorithm, proposed by Breiman (2001), is an ensemble learner which consists of many individual decision trees which are built from a bootstrap sample taken from the population of all samples, ntree. Additionally, the node-splitting rules are generated by randomly selecting a predictor from a subset of predictors based on mtry, which is the main tuning parameter for RF. Mtry and ntree were identified as those returning the lowest out of bag (OOB) error by iterating mtry values from one to the total number of important variables and ntree values were chosen 1000. The results of individual models are aggregated into an ensemble using an averaging function when predicting continuous response variables (i.e., SOC). The ensemble modelling approach is designed to mitigate the impacts of model overfitting. The RF model was implemented using the Caret package (Kuhn 2012 and Brewer et al. 2015) in the R statistical software (R Core Team 2018).

5.2.4.2. Cubist

Cubist is an extension of Quinlan's M5 model tree (Quinlan 1993) and was implemented using the cubist package in R (Kuhn et al. 2013 and R Core Team 2018). Although cubist is comparable to ordinary regression trees, its leaves are in the form of a linear regression equation (Taghizadeh-Mehrjardi et al. 2016). Considering the hybridization of a tree-based model with linear models, cubist can characterize both linear and non-linear relationships. It is also worth mentioning that many researchers have been using the cubist model in different soil prediction and mapping techniques (for example Henderson et al. 2005, Minasny et al. 2008). The cubist method's principal achievement is to use multiple training committees and boosting to make the weights more balanced. The outstanding usage of cubist is to analyse enormous databases that include a great number of records and numeric or nominal fields. Cubist models also compute variable importance to model accuracy as a variable's relative contribution.

5.2.4.3. Quantile Random Forest (QRF)

QRF is an extension of the RF learner, which allows users to leverage the model predictions from each tree of the RF to generate uncertainty estimates. Meinshausen (2006) reported that QRF not only provides information about the conditional mean, but it also provides information about the conditional distribution of the target variable. In addition, only the mean of the observations within the terminal node is used in RF whereas QRF keeps all predicted values for each terminal node. Accordingly, QRF retains the residual distribution at each terminal node, which is used to estimate the prediction interval width. The QRF model was implemented using the Quantreg Forest package (Vaysse et al. 2017) used by RStudio 3.5.0 software.

5.2.5. Uncertainty

Uncertainties in SOC stock assessments are critical in determining the significance of the results. No prediction is free from errors, as every model is a simplified representation of reality. The prediction error can be tracked down to uncertainty introduced in a model either as a result of input uncertainty or during incomplete construction of a model. Therefore, the modelling process is very dependent on training data, not only because of its uncertainties, but also because QRF estimates the cumulative distribution function (CDF) by using an empirical CDF. Therefore, it quantifies the complete error given a certain input vector as it includes a conditional variance estimate for Y by using the information within the leaves. Hence, Meinshausen's (2016) technique can be used for making prediction intervals and not for confidence intervals because the empirical CDF provides no information on the uncertainty of the fit of the RF model itself since the data needs to be a representative sample of the underlying populations (James et al. 2014).

5.2.6. Model evaluation

Spatial models were validated by leave-group-out cross-validation (LGOCV) as well as by independent validation. The latter was performed by randomly splitting the sample set (70% calibration and 30% validation). Each model was fitted using the train data and the test data was used for validation. Differences between observed and predicted values were summarized as the root-mean-squared error of prediction (RMSEP) and the bias of the estimation.

Model evaluation is an essential factor for accurately predicting SOC (Mosleh et al. 2016). K-fold cross-validation is usually used to evaluate model performance. In this research, the training dataset was randomly partitioned into 10 folds (k=10), so 10-fold cross-validation was used. The model was trained using k=9 folds, tested with the one remaining fold, and accuracy metrics were calculated based on the test fold. The process of training and testing was repeated 10 times so each individual fold was selected as the test set once. The model performance was evaluated based on the average accuracy metrics of all folds including mean absolute prediction error (MAE), root mean square error (RMSE), and index of determination (R^2). The metrics were used to evaluate the prediction error rates and model efficiency as well as to investigate the correspondence between predicted and measured data.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \qquad Equ. 2$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \qquad Equ. 3$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \qquad Equ. 4$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \qquad Equ. 5$$

where y_i is the measured value at i-th location, $\hat{y}_i$ is the predicted value of y, $\bar{y}$ is the mean value of y, and N is the number of units (locations).

Although R^2 is a valid statistic to evaluate the prediction accuracy of a model, a high R^2 may not lead to accurate predictions. This is because the model could systematically and considerably over- and/or under-estimate the data at different points along the regression line. As a result, evaluation of the models using other performance statistics appears to be necessary to provide complement information on prediction accuracy. A lower gained value is adequate and evaluated best for the selection of a model using the RMSE and MAE validation criteria evaluation methods. In the current study, the Li et al. (2016) criterion was applied. They proposed a classification criterion for R^2: unacceptable prediction ($R^2 < 0.50$), acceptable prediction ($0.50 \leq R^2 < 0.75$), and good prediction ($R^2 \geq 0.75$).

5.3. Results and Discussion

5.3.1. Summary statistics and correlation analysis

Statistics summary of SOC is presented in Table 5.3. The average SOC was 2.83% and 2.83% in the Liberec and Domažlice districts, respectively (Nozari and Borůvka, 2023). The results of the correlation analysis between environmental covariates and SOC are also presented in Tables A4 and A5. SOC was positively correlated with elevation. Although the correlation between most of the variables and SOC was not high, the RF model can identify nonlinear relationships between variables. Nevertheless, one of the significant parameters to explain SOC content variation is elevation, particularly in areas outside of flat sub-humid climates (Tziachris et al. 2019). The characteristics showing good correlation with SOC indicate potential candidates for strong predictors in SOC models, as analysed in section 5.3.3.

Table 5. 3. Statistics summary of SOC for both districts (%)

District	Number of observations	Minimum	Maximum	Mean	Median	Standard deviation
Liberec	71	0.42	11.33	2.83	1.86	2.50
Domažlice	67	0.00	9.33	2.83	1.49	2.39

5.3.2. Model validation

Table 5.4 presents values for R^2, MAE, and RMSE in the Liberec and Domažlice districts using RF, QRF, and cubist models. R^2 values ranged between 0.40 and 0.68, MAE ranged between 0.98 and 1.49, and RMSE ranged between 1.32 and 2.21. It is generally believed that R^2 values greater than 0.4 indicate the effectiveness of the model in predicting soil properties (Moore et al. 2013 and Prasad et al. 2006). Although R^2, MAE, and RMSE values for all three models used in this study were similar, RF consistently showed greater accuracy metrics (greater R^2 but smaller RMSE and MAE values) compared to cubist and QRF for both districts. Therefore, based on these accuracy metrics RF was considered the most accurate algorithm among the three models used in this research. This finding supports the findings of other studies which also found that RF is suitable for soil spatial modelling due to its high accuracy (Lamichhane et al. 2019 and Ellili et al. 2019). It should be noted that these indicators may not be suitable for prediction accuracy of the local uncertainty.

Table 5. 4. Assessment results for RF, cubist, and QRF models for SOC prediction

Model	Liberec district			Domažlice district		
	R^2	**MAE**	**RMSE**	**R^2**	**MAE**	**RMSE**
RF	0.58	1.40	1.98	0.68	0.98	1.32
Cubist	0.40	1.49	2.21	0.64	1.08	1.57
QRF	0.48	1.28	1.98	0.49	1.18	1.74

R^2 - coefficient of determination; MAE - mean absolute error; and RMSE - root mean square error

5.3.3. Variable importance

Mosleh et al. (2016) reported that the parameters derived from the DEM in low-relief areas are appropriate environmental factors to model soil properties. In this research, the relative importance of the predictor variables in the SOC modelling was evaluated using the Varlmp function in R (R Core Team 2018). It is believed that the climate, temperature, and disaster conditions (e.g. large-scale geological or meteorological events such as flooding, runoff, erosion, drought, and dust storms) are similar in both districts. In the Liberec district, the most effective variables in SOC prediction using RF and cubist models were land cover (vegetation), elevation, valley depth, and slope as illustrated in Figure 5.5. Similarly, Ellili et al. (2019) found that slope and elevation are the most important covariate variables for predicting SOC. These results indicate that the land cover is essential in identifying the SOC distribution. Coniferous forest, broad-leaved forest, and mixed forest containing a great amount of SOC in both districts show the importance of forest management. In addition, slope and valley depth, which are related to the topography and hydrology of the region, have effects on water distribution and runoff transport, affecting the erosion and deposition that changes the spatial variation of SOC in this mountainous area, as well as the soil organic matter (SOM) decomposition and accumulation processes. Generally, the steeper and longer a slope is, the faster water runs off from it, increasing the potential of erosion. Therefore, the influences of slope and valley depth were highlighted as the most significant auxiliary variables in predicting SOC in the Liberec region. In the Domažlice district, the most important variables in SOC prediction using RF and cubist models were land cover and elevation as illustrated in Figure 5.5. It also confirms the significance of vegetation and topographic parameters similar to the Liberec district.

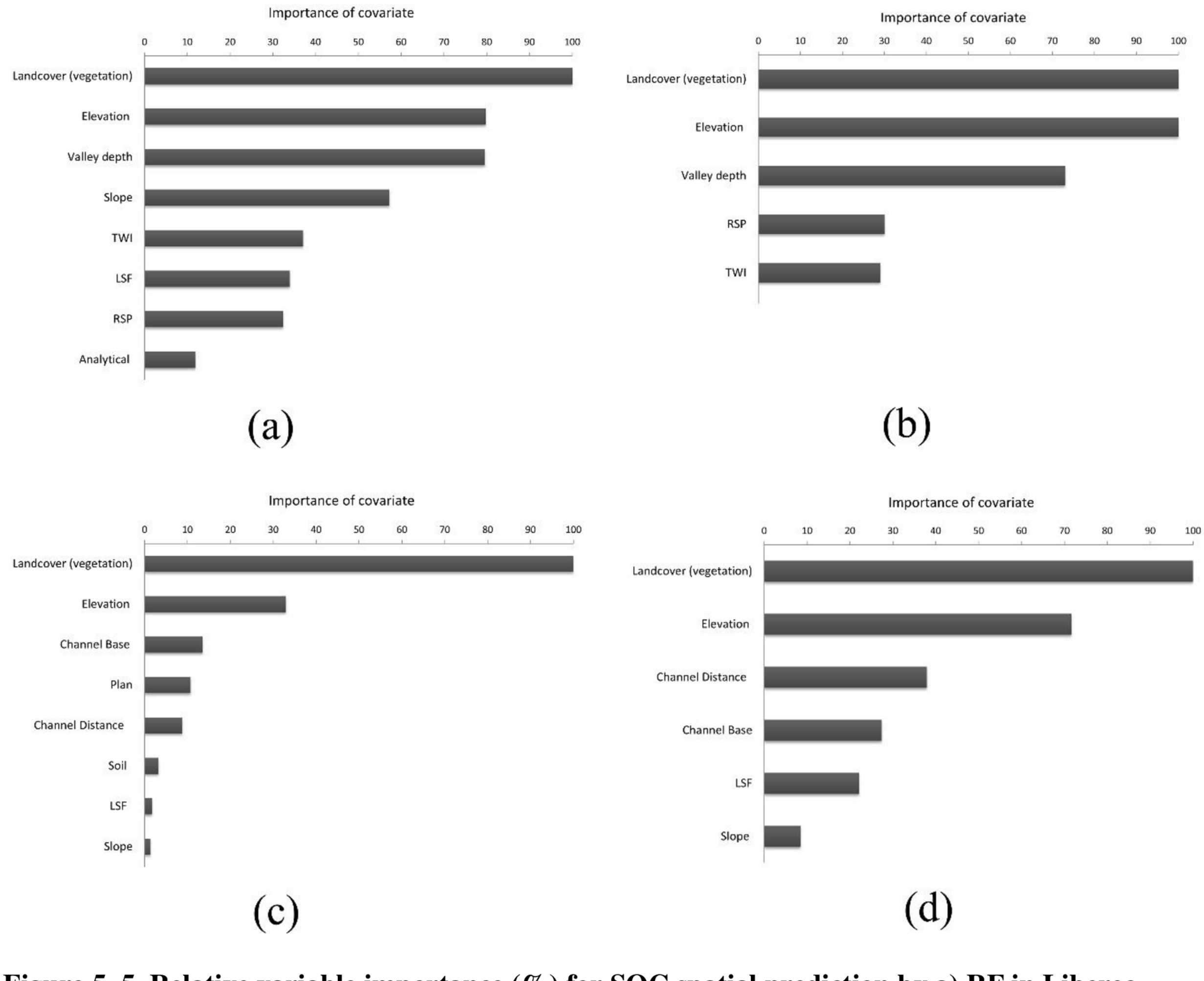
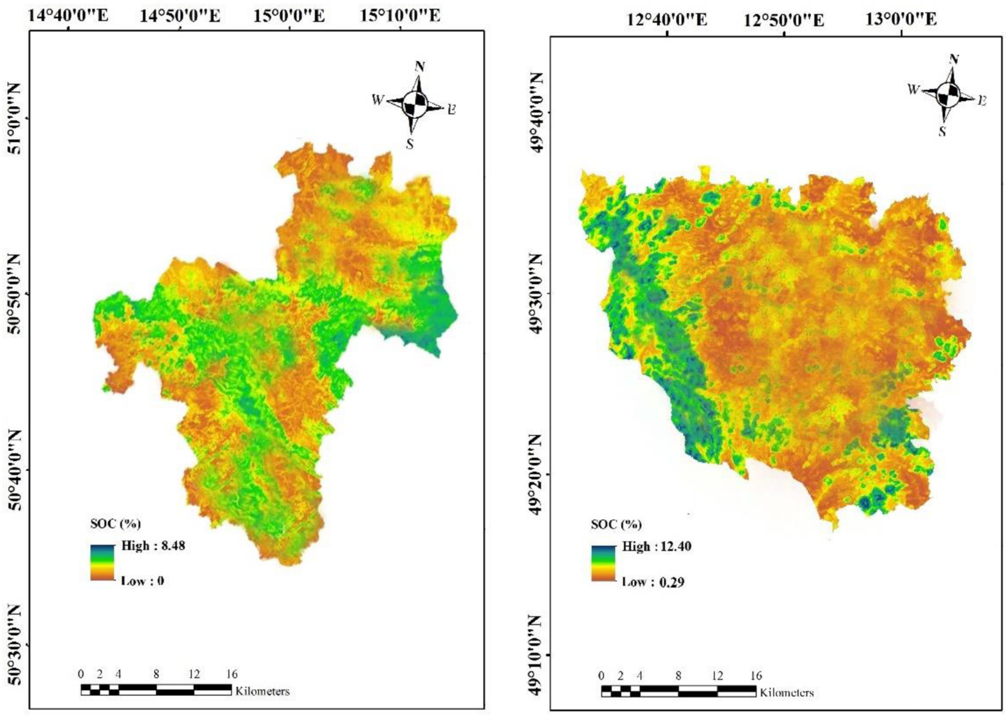


Figure 5. 5. Relative variable importance (%) for SOC spatial prediction by a) RF in Liberec, b) cubist in Liberec, c) RF in Domažlice, and d) cubist in Domažlice.

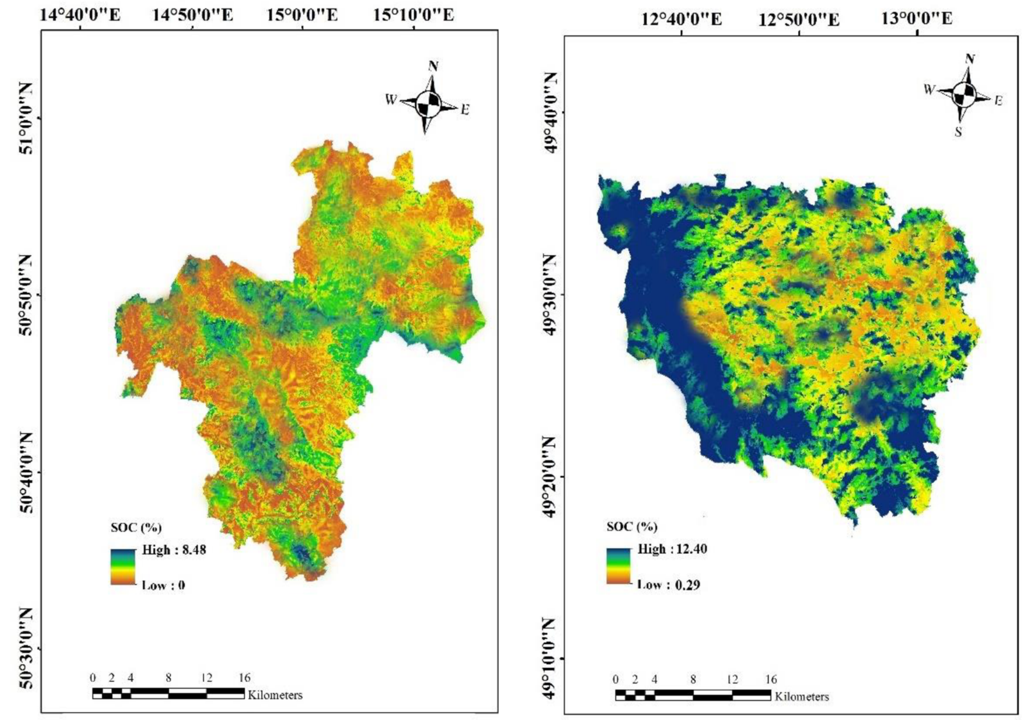
5.3.4. Spatial prediction of SOC

The spatial distribution of SOC content for the topsoil in the Liberec and Domažlice districts using RF, cubist, and QRF models is illustrated in Figures 5.6, 5.7, and 5.8, respectively.



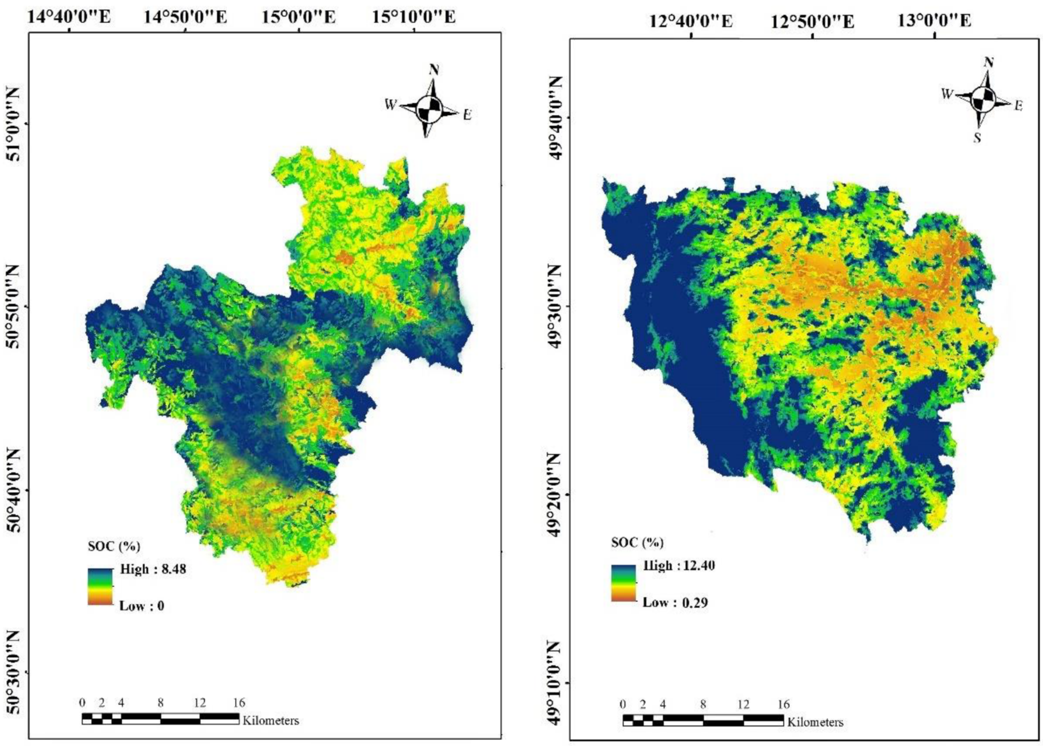
(a) (b)

Figure 5. 6. SOC distribution maps using RF model in the a) Liberec and b) Domažlice districts



(a) (b)

Figure 5. 7. SOC distribution maps using a cubist model in the a) Liberec and b) Domažlice districts



(a) (b)

Figure 5. 8. SOC distribution maps using QRF model in the a) Liberec and b) Domažlice districts

Generally, all three models used in this study were similar in terms of spatial patterns of SOC content. The biggest SOC content was found in high elevations covered by forests, while the lowest SOC content was observed in the areas where croplands have replaced the plantation and indigenous forests, which is consistent with observations by other researchers (Winowiecki et al. 2016 and Tesfahunegn et al. 2011). Therefore, a reduction in SOC stocks could be due to the biomass removal after harvesting, erosive processes, and frequent tillage that breaks up the soil aggregates, alters aeration, and accelerates the microbial decomposition and oxidation of SOM to CO2. It was also found that increasing elevation increased the average SOC concentrations, confirming that SOC responds to climatic variables such as temperature that decreases as elevation increases. Increased SOC may also be due to the recent changes in land-use. For example, agricultural lands at higher elevations are more likely to have been recently changed to another type of land (grassland).

Results also showed greater SOC accumulation extending from northeast to west areas of the Liberec region, and southeast parts of the Domažlice region. Generally, elevation changes affect the soil physicochemical attributes which are the main factors for predicting the SOC content variation. As can be seen in Figures 5.6, 5.7, and 5.8, SOC content has changed sharply based on QRF, while RF shows a more continuous distribution of SOC content for both regions. In addition, QRF maps showed different SOC distribution from RF and cubist maps in the Liberec district. SOC content in the western part of the study areas predicted by QRF maps is much lower than that predicted by cubist. Although RF maps show a rather even distribution of SOC content in agricultural and forestry areas, its results greatly differed from the cubist and QRF maps. RF map indicated that SOC content of the elevation ridge in the eastern parts of the Domažlice district is much lower than that predicted by cubist. Moreover, the RF map predicted a higher amount of SOC in northern parts, southern parts, and around the centre of the Liberec district than cubist and QRF maps.

Topography maps (Figure 5.2) show that the Liberec district is mostly mountainous and sloping, which results in a greater accumulation of SOC. This can be due to the combined effects of soil acidification through reduced decomposition in higher elevations and poor water drainage on lower slopes. Similarly, Zhu et al. (2018) reported that the SOC content is more aggregated and less decomposed in soils with greater slope and poor drainage. In addition, topographic variables such as the depth of the valley have affected water distribution, runoff velocity, and sediment erosion, and so the spatial variation of SOC in both regions has increased. The reduction of SOC is more pronounced in agricultural and residential areas than in areas where human manipulation in nature is limited. Total SOC content in both districts is high due to the high humidity, forest vegetation such as coniferous forest, and a great amount of rainfall resulting in denser vegetation. As a result, this study confirms the importance of terrain-based covariates and vegetation on SOC content variability in sub-humid areas. Also, there is a great variety of land-use and agricultural practices that may generate contrasting organic matter levels.

5.3.5. SOC prediction uncertainty

DSM requires field observations, empirical prediction models, and a variety of environmental covariates to model spatially explicit predictions of soil properties. Therefore, the predictions are always related to uncertainties brought by these three sources. SOC prediction uncertainty maps using RF, cubist, and QRF models for the Liberec and Domažlice districts are illustrated in Figures 5.9 through 5.11. The greatest uncertainty of SOC was in the coniferous forests in both districts (compared to landcover maps in Figures 5.3c and 5.4c). This is likely because there were relatively few SOC samples (15 samples in Liberec and 17 samples in Domažlice districts) and the variability of the covariates increased in these areas. Therefore, it is recommended to take more samples from these areas to ensure lower SOC prediction uncertainty. Interestingly, although each model had similar RMSE, MAE, and R^2, there are differences in uncertainty patterns in each model prediction for both districts. Particularly, differences between RF and QRF, using the same basic algorithm, were surprising. This can be because they differ in how the terminal nodes are dealt with. QRF appears to produce a lower uncertainty at lowlands compared to RF, however, both QRF and cubist show larger values in the mountainous regions where the models are most likely extrapolating. While the QRF model had low RMSE and MSE values similar to RF, the uncertainty distribution is much more uniform for RF.

Similar pattern is repeated in the Domažlice district as well. The higher uncertainty in predicted SOC was observed in mountains and forests compared to the cropland and pastures which is visible in western areas.

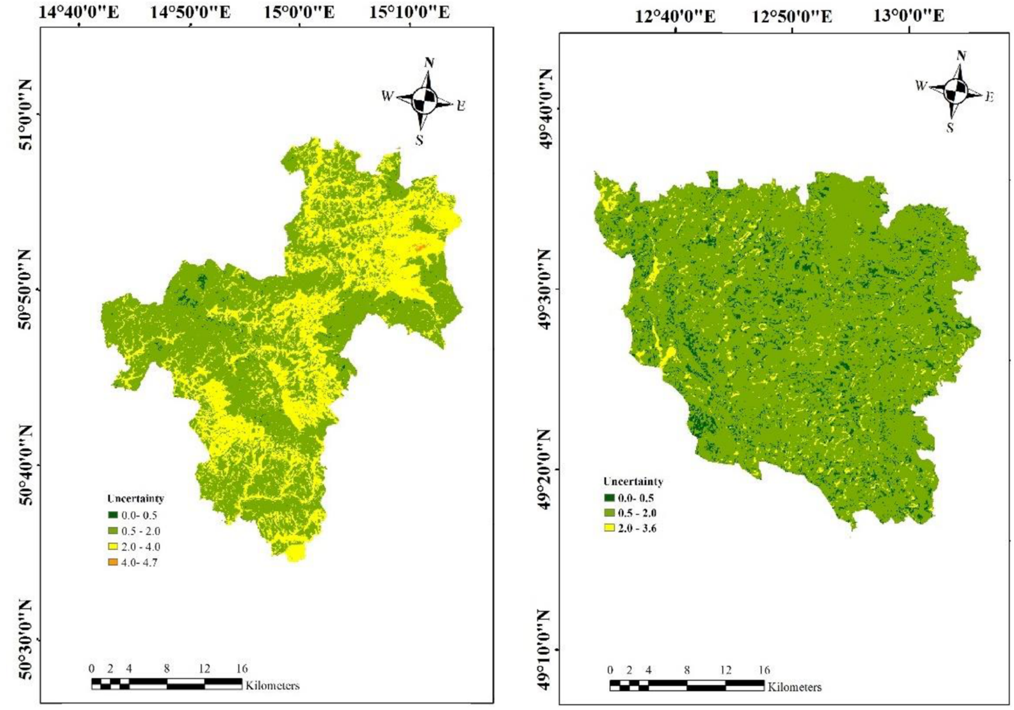


Figure 5. 9. SOC uncertainty maps using RF model in the a) Liberec and b) Domažlice districts

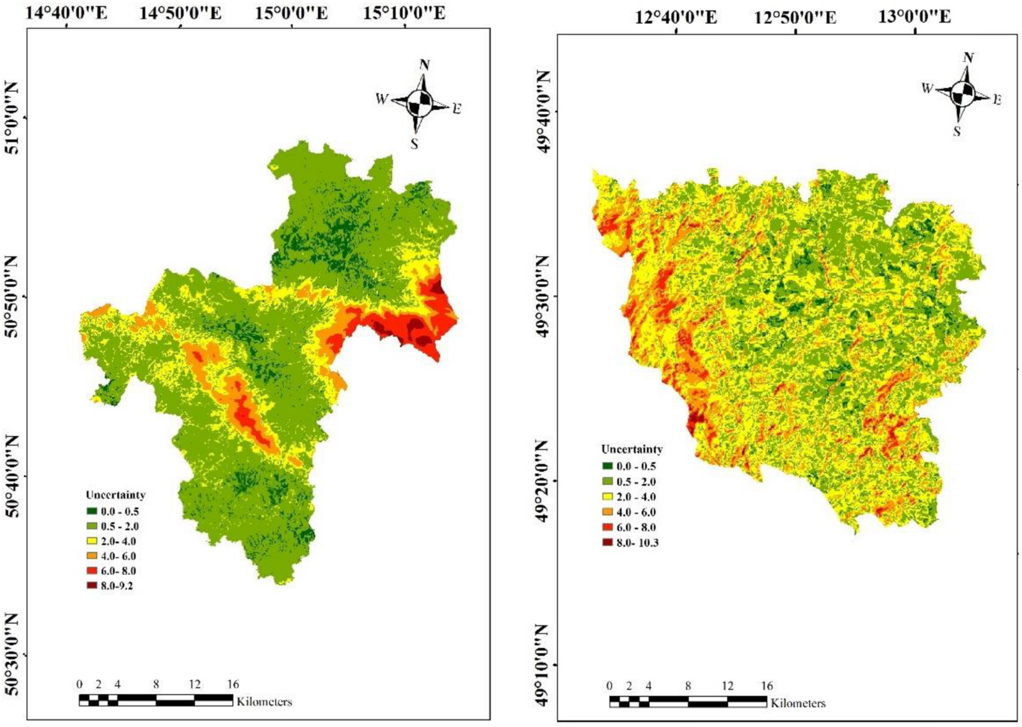


Figure 5. 10. SOC uncertainty maps using cubist model in the a) Liberec and b) Domažlice districts

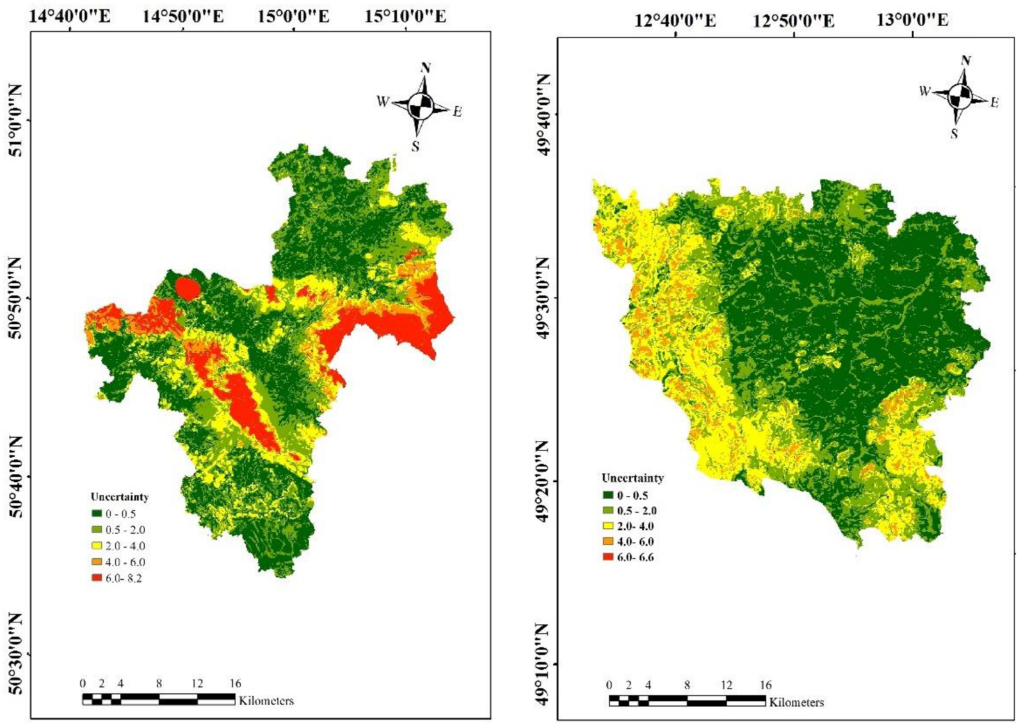


Figure 5. 11. SOC uncertainty maps using QRF model in the a) Liberec and b) Domažlice districts

5.4. Conclusions

This study assessed the spatial distribution of SOC in the Liberec and Domažlice districts in the Czech Republic. From the results of this study, the following conclusions can be drawn:

1. Although the studied models including RF, QRF, and cubist did not have substantially excellent performance (not achieving high R^2 values), RF model consistently showed the best performance among all three models in both districts.

2. Based on the RF model results, land cover (vegetation) and elevation were the most important environmental covariates for the prediction of SOC in both districts.

3. The highest SOC content was predicted in the highest elevation in the forest-dominated areas (northeastern to western parts of the Liberec region and southeast parts of the Domažlice region) while the lowest SOC was found in the lowest elevations in the cropland-dominated areas.

4. The greatest uncertainty of SOC was observed in the coniferous forests in both districts, most likely because there were relatively few SOC samples and the variability of the covariates increased in these areas.

5. Overall, RF can use many terrain covariates which have a strong spatial association with SOC and is considered the most accurate predictive model for both districts because it showed better performance (greater R^2 but smaller RMSE and MAE values, more uniform uncertainty distribution without very high uncertainty values) compared to cubist and QRF for both districts.

6. Finally, to improve the prediction accuracy of SOC distribution, more observations and stratified random sampling using known variables such as habitat type, elevation, or soil type are recommended to be performed in both districts which will enhance the performance of all models.

5.5. References

Badia, D., Ruiz, A., Giona, A., Marti, C., Casanova, J., Ibbara, P., Zufiaurre, R. (2016): The influence of elevation on soil properties and forest litter in the Siliceous Moncayo Massif, SW Europe. Journal of Mountain Science, 13: 2155 - 2169.

Borůvka, L., Vašát, R., Šrámek, V., Neudertová Hellebrandová, K., Fadrhonsová, V., Sáňka, M., Pavlů, L., Sáňka, O., Vacek, O., Němeček, K., Nozari, S., Oppong Sarkodie, V.Y. (2022): Predictors for digital mapping of forest soil organic carbon stocks in different types of landscape. Soil and Water Research, 17(2): 69-79.

Brevik, E.C. (2013): Soil health and productivity. In Verheye, W.H. (ed.): Soils, Plant Growth and Crop Production. Oxford, England: Encyclopedia of Life Support Systems (EOLSS)..

Brungard, C., Boettinger, J.L., Duniway, M.C., Wills, S.A., Edwards, T.C. 2015. Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma, 239: 68-83.

Breiman, L. (2001): Random forests. Machine Learning, 45(1): 5-32.

Brewer, S., Liaw, A., Wiener, M., Liaw, M.A. (2015): Package random Forest. Breiman and Cutler's Random Forests for Classification and Regression. R package version 4.6-14. Availbale on: URL: https://cran.r-project.org/web/packages/randomForest/index.html

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J. (2015): System for Automated Geoscientific Analyses (SAGA) v. 2.1.4, Geoscientific Model Development, 8: 1991–2007.

EEA (2018): CORINE Land Cover (CLC) 2018, Version 20, European Environment Agency, Kopenhagen, Denmark. Accessed from https://www.eea.europa.eu/data-and-maps/data/corine-land-cover-accounting-layers on 2020-10-23.

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.

Fathololoumi, S., Vaezi, A.R., Alavipanah, S.K., Ghorbani, A., Saurette, D., Biswas, A. (2020): Improved digital soil mapping with multitemporal remotely sensed satellite data fusion: A case study in Iran. Science of the Total Environment. 721:137703.

Gholizadeh, A., Žížala, D., Saberioon, M., Borůvka, L. 2018. Soil organic carbon and texture retrieving and mapping using proximal, airborne, and Sentinel-2 spectral imaging. Remote Sensing of Environment, 218, 89-103.

Henderson, BL., Bui, EN., Moran, CJ., Simon., DAP. (2005): Australia-wide predictions of soil properties using decision trees. Geoderma, 124 (3-4): 383-398.

Heung, B., Bulmer, C.E., Schmidt, M.G. (2014): Predictive soil parent material mapping at a regional scale: a random forest approach. Geoderma, 214: 141-154.

IUSS Working Group WRB. (2015): World Reference Base for Soil Resources 2014 (Update 2015). International Soil Classification System for Naming Soils and Creating Legends for Soil Maps, World Soil Resources Reports, No. 106, Rome: FAO.

James, G., Witten, D., Hastie, T., Tibshirani, R. (2014): An Introduction to Statistical Learning: With Applications in R. Springer Publishing Company, Inc.

Kozák, J., Němeček, J., Jetmar, M. (1996): The database of soil information system- PUGIS, Plant Production, 42(12): 529-534.

Kuhn, M., Weston, S., Keefer, C., Coulter, N., Quinlan, R. (2013): Cubist: Rule-and instance-based regression modeling. R package version 0.0, 13.

Kuhn, M. (2012): The caret package. R Found. Statistical Computing. Vienna, Austria. URL https//cran. r-project. org/packa-ge= caret.

Lal, R., Bouma, J., Brevik, E., Dawson, L., Field, D.J., Glaser, B., Hatano, R., Hartemink, A.E., Kosaki, T., Lascelles, B., Monger, C., Muggler, C., Martial Ndzana, G., Norra, S., Pan, X., Paradelo, R., Reyes-Sánchez, L.B., Sandén, T., Singh, B.R., Spiegel, H., Yanai, J., Zhang, J. (2021): Soils and sustainable development goals of the United Nations: An International Union of Soil Sciences perspective. Geoderma Regional, 25: e00398.

Lagacherie, P., McBratney, A.B., Voltz, M. (eds.) (2006): Digital soil mapping - An introductory perspective (2006): Developments in Soil Science, Elsevier Science, Amsterdam, Hardbound, 31: 1-600.

Lamichhane, S., Kumar, L., Wilson, B. (2019): Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. Geoderma, 352: 395-413.

Li, L., Lu, J., Wang, S., Ma, Y., Wei, Q., Li, X., Cong, R., Ren, T. (2016): Methods for estimating leaf nitrogen concentration of winter oilseed rape (Brassica napus L.) using in situ leaf spectroscopy. Industrial Crops and Products, 91: 194–204.

Lorenzetti, R., Barbetti, R., Fantappiè, M., L'Abate, G., Costantini, E.A.C. (2015): Comparing data mining and deterministic pedology to assess the frequency of WRB reference soil groups in the legend of small-scale maps. Geoderma, 237-238.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B. (2003): On digital soil mapping. Geoderma, 117(1-2): 3-52.

Meinshausen, N. (2006): Quantile regression forests. Journal of Machine Learning Research, 7 (6): 983-999.

Meinshausen, N., Hauser, A., Mooij, J., Peters, J., Versteeg, P., Buhlmann, P. (2016): Methods for causal inference from gene perturbation experiments and validation. Proc. National Academy of Sciences. USA, 113: 7361-7368.

Miko, L,. Hošek, M. (2009): The state of nature and the landscape in the Czech Republic. 1st Ed. Prague, Agency for Nature Conservation and Landscape Protection of the Czech Republic: 102.

Minasny, B., McBratney, A.B., Salvador-Blanes, S. (2008): Quantitative models for pedogenesis - A review. Geoderma, 144(1): 140-157.

Minasny, B., McBratney, A.B. (2016): Digital soil mapping: A brief history and some lessons. Geoderma, 264: 301-311.

Moore, D. S., Notz, W.I, Flinger, M.A. (2013): The basic practice of statistics. 6th Ed. New York, NY, W. H. Freeman and Company: 138.

Mosleh, Z.,Salehi, M.H., Jafari, A., Borujeni, I.E., Mehnatkesh, A. 2016. The effectiveness of digital soil mapping to predict soil properties over low-relief areas. Environmental Monitoring and Assessment, 188(3): 1-13.

Mulder, V.L., de Bruin, S., Schaepman, M.E., Mayr, T.R. (2011): The use of remote sensing in soil and terrain mapping - A review. Geoderma, 162 (1-2): 1-19.

Němeček, J. (2000): The status of soil mapping in the Czech Republic. The European Soil Information System. World Soil Resources Reports, 91: 61-65.

Němeček, T., Dubois, D,. Huguenin-Elie, O,. Gaillard, G. (2011): Life cycle assessment of Swiss farming systems: Integrated and organic farming. Agricultural Systems, 104: 217-232.

Nikou, M., Tziachris, P. (2022): Prediction and Uncertainty Capabilities of Quantile Regression Forests in Estimating Spatial Distribution of Soil Organic Matter. ISPRS Int. J. Geo-Inf., 11: 130.

Nozari, S., Borůvka, L. (2020): The effect of landscape slope on soil organic carbon in the Liberec district in the Czech Republic. International Journal of Advanced Science and Technology, 29: 3934-3942.

Nozari, S., Borůvka, L. (2023): The effects of slope and altitude on soil organic carbon and clay content in different land-uses: A case study in the Czech Republic. Soil and Water Research 18(3): 204-218.

Pahlavan-Rad, M.R., Akbari Moghaddam, A.R. (2018a): Spatial variability of soil texture fractions and pH in a flood plain (a case study from eastern Iran). Catena, 160: 275-281.

Pahlavan-Rad, M.R., Dahmardeh, K.H., Brungard, C. (2018b): Predicting soil organic carbon concentrations in a low relief landscape, eastern Iran. Geoderma Regional, 15: e00195.

Pahlavan-Rad, M.R., Dahmardeh, K.H., Hadizadeh, M., Keykha, G., Mohammadnia, N., Gangali, M., Keikha, M., Davatgare, N., Brungard, C. (2020): Prediction of soil water infiltration using multiple linear regression and random forest in a dry flod plain, eastern Iran. Catena, 194:104715.

Panagos, P., Meusburger, K., Ballabio, C., Borrelli, P., Alewell, C. (2014): Soil erodibility in Europe: A high-resolution dataset based on LUCAS. The Science of the Total Environment, 479: 189-200.

Peters, J., De Baets, B., Verhoest, N.E., Samson, R., Degroeve, S., De Becker, P. Huybrechts, W. (2007): Random Forests as a tool for ecohydrological distribution modeling. Ecological Modeling, 207(2-4): 304-318.

Pieri, C.J. (1992): The Effect of Population Growth on the Soil. In Fertility of Soils. Springer Series in Physical Environment, 10. Springer, Berlin, Heidelberg, 84-99.

Pospíšil, F. (1964): Fractionation of humus substances of several soil types in Czechoslovakia. Rostlinná Výroba, 10: 567-579.

Prasad, A.M., Iverson, L.R., Liaw, A. (2006): Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9 (2): 181-199.

Pouladi, N., Møller, A.B. Tabatabai, S., Greve, M.H. (2019): Mapping soil organic matter contents at field level with cubist, random forest and kriging, Geoderma, 342: 85-92.

Quinlan, J. (1993): C4.5: Programs for Empirical Learning Morgan Kaufmann. San Francisco, CA. 38(48): 49.

R Core Team. (2018): R, A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Selvaradjou, S.K., Montanarella, L., Carre, F., Jones, A., Panagos, P., Ragunath, K., Kumaraperumal, R., Natarajan, S. (2007): An Innovative Approach for Updating Soil Information based on Digital Soil Mapping Techniques. Official Publication of European Communities. EUR 22545 EN, 1-34.

SPSS Statistics for Windows, version 11.0, SPSS Inc., Chicago, Ill., USA.

Taghizadeh-Mehrjardi, R., Nabiollahi, K., Kerry, R. (2016): Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. Geoderma, 266: 98-110.

Tesfahunegn, G.B., Tamene, L., Vlek, P.L.G. (2011): Catchment scale spatial variability of soil properties and implications on site-specific soil management in northern Ethiopia. Soil and Tillage Research, 117: 124-139.

Tziachris, P., Aschonitis, V., Chatzistathis, T. (2019): Assessment of spatial hybrid methods for predicting soil organic matter using DEM derivatives and soil parameters. Catena, 174: 206-216.

U.S. Geological Survey (USGS). 2021. retrieved from https://earthexplorer.usgs.gov/.

Vaysse, K., Lagacherie, P., 2017. Using quantile regression forest to estimate uncertainty of digital soil mapping products. Geoderma, 291: 55-64.

Victoria, R., Banwart, S., Black, H., Ingram, J., Joosten, H., Milne, E., Noellemeyer, E. (2012): The benefits of soil organic carbon. Emerging Issues in Our Global Environment. UNEP Division of Early Warning and Assessment United Nations Environment. UNION, Nairobi 00100, Kenya, 80.

Weier, J., Herring, D. (2000): Measuring vegetation (NDVI & EVI). NASA. Available from: https://earthobservatory.nasa.gov/Features/MeasuringVegetation/. Accessed in January 2020.

Wiesmeier, M., Urbanski, L., Hobley, E.; Lang, B., von Lützow, M., Marin-Spiotta, E., van Wesemael, B., Rabot, E., Ließ, M., Garcia-Franco, N. (2019): Soil organic carbon storage as a key function of soils - a review of drivers and indicators at various scales. Geoderma, 333: 149-162.

Winowiecki, L., Vågen, TG., Massawe, B. Jelinski, N., Yamchai, Ch., Sayula, G., Msoka, E. (2016): Landscape-scale variability of soil health indicators: effects of cultivation on soil organic carbon in the Usambara Mountains of Tanzania. Nutrient Cycling in the Agroecosystems, 105: 263-274.

Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., Finke, P. (2019): Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. Geoderma, 338: 445-452.

Zhu, H., Hud, W., Jing, Y., Cao, Y. (2018): Soil organic carbon prediction based on scale-specific relationships with environmental factors by discrete wavelet transform, Geoderma, 330: 9 -18.

Žížala, D., Minařík, R., Skála, J., Beitlerová, H., Juřicová, A., Rojas, J.R., Penížek, V., Zádorová, T. (2022): High-resolution agriculture soil property maps from digital soil mapping methods, Czech Republic. Catena, 212:106024.

CHAPTER 6

Comparison of Soil Organic Matter Composition under Different Land Uses by DRIFT Spectroscopy

Citation:

Thai S., Pavlů L., Tejnecký V., Vokurková P., **Nozari S.,** Borůvka L. (2021): Comparison of soil organic matter composition under different land uses by DRIFT spectroscopy. Plant Soil Environ., 67: 255–263. https://doi.org/10.17221/11/2021-PSE

Abstract: The study aimed to estimate and characterise soil organic matter under different land-uses (cropland, grassland, and forest) and soil depths. The soil organic matter composition of the soil was assessed by diffuse reflectance infrared spectroscopy (DRIFT). Humic and fulvic acids (HAs, FAs) were extracted from soils and their compositions were evaluated by DRIFT. Low molecular mass organic acids content was also measured. Our result revealed that the largest differences of the spectra in the composition of organic matter were observed in the upper parts of the soil profile. The forest soil spectra had more intense aliphatic bands, carboxylic, and CH bands than spectra of grassland and cropland soils. The difference of HAs spectra was at 3010 to 2800/cm where the most intensive aliphatic bands were in forest soil HAs, followed by grassland and cropland soil HAs. The grassland topsoil FAs spectrum differs most from the other land-uses. It has lower peaks around 1660–1600/cm and 1200/cm than cropland and forest. The concentration of low molecular mass organic acid (LMMOA) was the highest in the forest soil and the most abundant acid was citrate.

Keywords: terrestrial ecosystem; Luvisols; humus; organic compounds; functional groups

6.1. Introduction

Soil organic matter (SOM) plays an important role in biological, chemical, and physical soil improvement and productivity (Strosser 2010). The living biomass including microorganisms breaks down the plant residues or detritus and animal waste into humus or organic matter by using carbon as an energy source and nitrogen as a source of protein production (Allison et al. 2007). The decomposition of plant residues releases the organic chemical compounds and helps to cling together with the mineral soil particles that improve the chemical soil properties by soil sorption complex creating and physical soil properties by establishing of soil structure (Davidson and Ackerman 1993).

However, SOM in the ecosystem has been stored in different layers with different concentrations as a result of different stages of decomposition (Ribeiro et al. 2001). The quantity and the quality of SOM depend on several factors such as duration of decomposition, residues, roots, amount of fine materials, type of decomposers (microorganisms), chemical composition, and temperature (Lal 2018). On the other hand, the fluctuation of the organic matter concentration in the soil is related intensively to slopes, elevation, topography, soil types, and land-uses and management (Slepetiene and Slepetys 2005, Jakšík 2015). The humic substances such as humic acids (HAs) and fulvic acids (FAs) are also the component used to identify the quality of SOM. They play an important role in the terrestrial ecosystem (Trevisan et al. 2010), and they are known as a mixture of substances in the form of supramolecular structures (Piccolo 2001). Humic substances make up about 20% of the total of SOM and result from the decomposition and humification process of the SOM (Pavlů and Mühlhanselová 2017). They are often understood as relatively stable components of SOM, which are involved in the fixation and sequestration of carbon in the soil (Lal 2005). The low molecular mass organic acid (LMMOA), which makes up about 10% of dissolved organic carbon (DOC), also characterises the compositions of soil organic matter. They are carboxylic acids of low molecular weight (Ash et al. 2016) and could be aromatic or aliphatic (Hubová et al. 2017). LMMOA are understood as relatively variable and unstable components of SOM (Strobel 2001).

Diffuse reflectance infrared fourier transform spectroscopy (DRIFT), as one of the types of infrared spectroscopy conventionally used for solid powder samples, is commonly used to analyse peat soil, composts, and the transformation of organic matter during composting within various stages (Haberhauer and Gerzabek 1999, Zaccheo et al. 2002, Pavlů and Mühlhanselová 2017). The DRIFT spectra have been recognised as one of the spectroscopic techniques used to distinguish the fluctuation in the abundance of organic functional groups during decomposition and to identify the changes of SOM in the soil profile under different vegetation covers (Veum et al. 2014).

The study hypothesised that different land-uses are connected with different incoming fresh organic materials and these differences can be seen throughout the whole soil in the composition of organic matter. Therefore, the study aimed to describe and compare the SOM compositions and their transformation under different depths and vegetation covers. The comparison of the separated organic compounds such as humic and fulvic acid and low molecular mass organic acid was observed by the combination of the advanced analytical methods.

6.2. Materials and Methods

6.2.1. Site selection and soil sampling

The research was conducted on the outskirts of Prague, Suchdol (Czech Republic). The area is situated in altitude range 250–300 m a.s.l. and has a mean annual precipitation of about 470 mm and a mean average temperature of 11°C. The mixture of loess and sandy river sediments of the Quaternary age creates the bedrock of the research area. Haplic Luvisols are the prevailing soil type in all land-uses. Cropland site (with a dominantly grown wheat (Triticum aestivum L.) crop interspersed with rape (Brassica napus L.) and maize (Zea mays L.)) marginally include the areas of greyic Phaeozems and carbonates were detected in several soil samples mainly in deeper layers. Soil texture belongs to clay loam category. Same soil description applies to grassland site (poorly maintained grassland with Dactylis polygama Horv., Poa annua L., Calamagrostis epigejos Roth). Broadleaf forest site with the dominant abundance of oak (Quercus petraea (Matt.) Liebl.) followed with beech (Fagus sylvatica L.) and hornbeam (Carpinus betulus L.) marginally include the areas of Regosols and Cambisols. The presence of carbonate was not detected in all soil samples from the forest. Soil texture belongs to sandy clay loam category.

Ninety soil samples were collected from each land-uses and categorised for the three different depths (0–10 cm (with exclusion of litter layer in forest), 10–20 cm and 20–30 cm). In all cases, the samples from the first two layers captured humic (A) horizon. Samples from the deepest layer captured either still horizon A (in the case of Phaeozems or its gradual transformation to eluvial horizon in case of Luvisols, to cambic horizon in Cambisols or to mineral substrate in Regosols.

The taken samples were air-dried and sieved with a 2 mm sieve. Furthermore, 2-mm-sieved soil samples were milled (Fritsch Analysette 3 Spartan Pulvensette miller, Idar-Oberstein, Germany) into very fine particles to use for infrared spectroscopy. Fulvic and humic acid were extracted from selected 18 soil samples. The fresh topsoil (15 soil samples) was taken separately and frozen for analysing the DOC and LMMOA.

6.2.2. Soil analysis

The exchangeable (pH_{KCl}) was determined potentiometrically by the pH-electrode SenTix 21 (Inolab pH level 21, WTW, Prague, Czech Republic). Soil organic carbon (SOC) was measured by using rapid dichromate oxidation techniques (Sparks 1996). The quality of humus was determined by the absorbance ratio of sodium pyrophosphate ($Na_4P_2O_7$) soil extract at 400 nm and 600 nm (E4/E6, respectively) (Sparks 1996). The content of LMMOA was measured using ion chromatography (IC) with suppressed conductivity (Hubová et al. 2017). Dissolved organic carbon content was measured by the wet dichromate oxidation method according to Tejnecký et al. (2014).

The extraction of humic substances was carried out by the international humic substance society (IHSS) fraction method, which is modified by Piccolo et al. (2000). A mixture of NaOH and $Na_4P_2O_7$ was used to extract the humic substances. The extract was acidified to pH 1.0 using HCl for precipitation of humic acids and their separation from fulvic acids. The HAs fractions were purified by redissolution with NaOH and reprecipitation with HCl. The purification from co-extracted clay was completed with the solution of HCl and of HF. The suspension was neutralised, centrifuged, and dialysed to release chlorine and then the HAs were freeze-dried. The FAs solutions were purified using the hydrophobic resin in the column. The FAs were released from the sorption of resin using NaOH solution. Finally, the FAs were neutralised, dialysed, and freeze-dried.

DRIFT spectra of pure freeze-dried humic acids, fulvic acid, and dried fine soil samples were recorded by the infrared spectrometer (Nicolet iS10, Waltham, USA). The spectra with a range of 2.5 to 25 µm (4000 to 400/cm) were used. The gold mirror was used as a background reference. The 64 scans with resolution 4/cm and Kubelka-Munk units were applied. OMNIC 9.2.41 software (Thermo Fisher Scientific Inc., Waltham, USA) was applied for spectra analysis.

6.2.3. Data analysis method

The software IBM SPSS (version 26, New York, USA) was used for data analysing and One-way ANOVA was applied for determining the statistical differences among quantitative soil characteristics with different land-uses, and depths at significance level description $P < 0.05$. Tukey test and letters a, b, c were used to describe the significant differences, where a is the highest value, followed by b and c.

6.3. Results and Discussion

6.3.1. Basic soil characteristics

The analysed data (Tables 6.1 and 6.2) indicated that there are no significant differences for the pH_{KCl} among all three depths in all land-uses. However, the soil in cropland is neutral, in grassland is moderately acidic while in forest is strongly acid. Similarly, the method used for the indicative evaluation of SOM quality (E4/E6) in different depths in all the land-uses had no significant differences. The cropland and grassland have very good humus quality while the forest does not. The higher E4/E6 ratio in forest could indicate lower degree of humification processes (Kunlanit et al. 2019). Soil organic carbon content is significantly different among the three depths of all land-uses.

Table 6. 1. Describing soil characteristic among the different depths (0–10, 10–20, and 20–30 cm)

	Depth (cm)	pH_{KCl}	Humus quality index	SOC (%)
	0–10	6.79 ± 0.37	3.27±0.37	1.37 ± 0.17[a]
Cropland	10–20	6.76 ± 0.33	3.25±0.34	1.41 ± 0.20[a]
	20–30	6.85 ± 0.43	3.24±0.43	1.11 ± 0.13[b]
P-value		0.852	0.977	0.001
	0–10	5.93 ± 0.19	3.50±0.36	2.11 ± 0.31[a]
Grassland	10–20	5.73 ± 0.79	3.30±0.17	1.58 ± 0.25[b]
	20–30	5.91 ± 0.70	3.28±0.20	1.51 ± 0.38[b]
P-value		0.788	0.149	0.000
	0–10	3.59 ± 0.21	4.41±0.43	5.64 ± 2.54[a]
Forest	10–20	3.52 ± 0.13	4.56±0.42	1.78 ± 0.49[b]
	20–30	3.65 ± 0.10	5.02±1.17	1.06 ± 0.34[b]
P-value		0.206	0.195	0.000

Data (means ± standard deviation; n = 10)

Table 6. 2. The differences description of basic soil characteristics among the three land-uses (cropland, grassland, and forest)

	pH_{KCl}			E4/E6			SOC		
Depth (cm)	**0–10**	**10–20**	**20–30**	**0–10**	**10–20**	**20–30**	**0–10**	**10–20**	**20–30**
Cropland	a	a	a	b	b	b	b	ns	b
Grassland	b	b	b	b	b	b	b	ns	b
Forest	c	c	c	a	a	a	a	ns	a
P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000

Letters indicate significant difference; ns: none significance; E4/E6: the humus quality index

Table 6.2 shows in the depth 0–10 cm, the SOC has the highest content in the forest, followed by the grassland and cropland. For the 20–30 cm depth, the grassland has the highest SOC compared to SOC in the cropland and forest. Various studies found the same result, that grassland had higher SOC than cropland and forest in deeper soil layers (Muktar et al. 2018).

6.3.2. DRIFT spectra

6.3.2.1. Spectra of soils

Position and identification of soil spectra bands are presented in Table 6.3. The spectra of the cropland soil are very similar in all sampled depths (Figure 6.1). This corresponds well to tillage and soil stirring. In the soil spectra of the different depths under the grassland, there are also no differences in bands position, intensities, or shapes. The forest soil spectra of 0–10 cm layer differ from the two deeper layers. The higher content of aliphatic components is evident from band absorbance in the range between 3000–2800/cm.

Table 6. 3. The assignment of the major bands in infrared spectra of the soil (Tinti et al. 2015, Matamala et al. 2017)

Wavenumber (1/cm)	Assignment of sorption bands
3600–3700	Si–O–H vibration of clays
3440–3320	O–H and N–H stretching, H-bonded OH
3010–2800	Aliphatic CH stretching
2000–1790	Si–O vibration of quartz mineral
1775–1711	C=O stretching in carboxylic group
1691–1642	C=O stretching of amides (amide I), H-bonded conjugated ketones, carboxyls and quinones, lignin, C=N stretching
1642–1569	Amide II of primary amides, aromatic C=C, C=O (quinones), carboxylates
1544–1488	Aromatic C=C stretching, aromatic skeletal vibration, aromatic (lignin), amide II
1479–1444	CH and NH of amide II, aliphatic CH deformation, carbonates
1444–1408	C–H deformation and C–O stretching of phenolic groups
1403–1354	C–O of phenolic OH, COO^- and O–H, CH_3 bending,
1342–1307	C–N (aromatic amines)
1293–1256	C–O of aryl ethers, C–O of phenols, C–O–C ether bond, bentonite
1256–1198	C–O stretching and OH deformation of COOH, C–O of aryl ethers and phenols, silicate
1185–1070	C–OH of aliphatic alcohols, O–Si–O stretching of quartz, sulfates
1056–945	C–O stretching, polysaccharides, Si–OH of alumino-silicate lattice (kaolinite, illite)

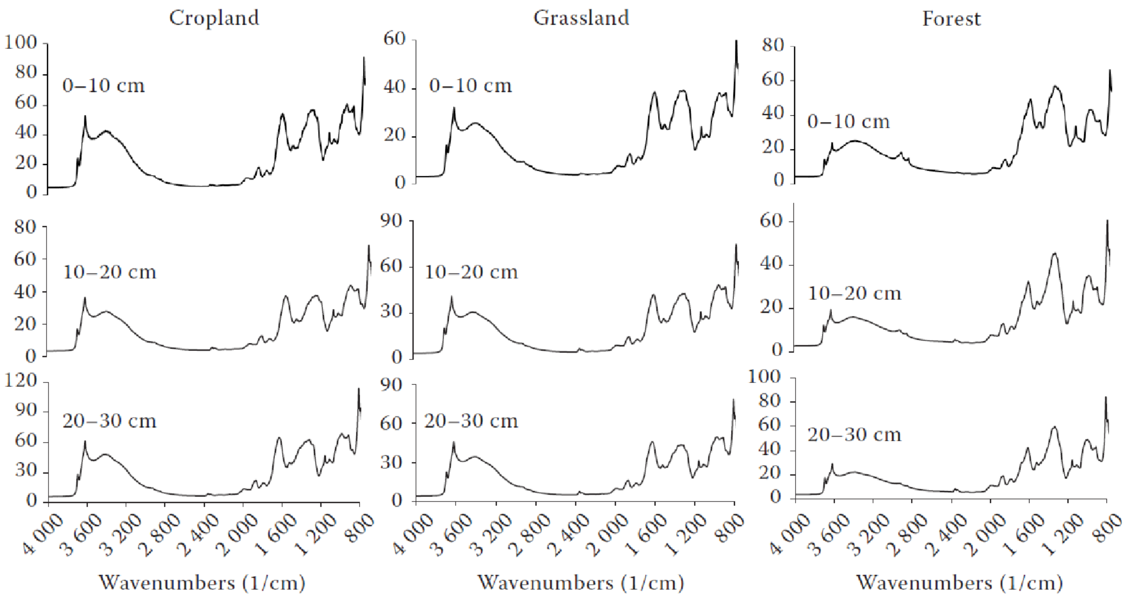


Figure 6. 1. The difference of average soil spectra under different depths (0–10, 10–20 and 20–30 cm) and land-uses (cropland, grassland, and forest)

The most obvious differences among land-uses are visible in soil spectra of the surface layer. The spectrum of forest soil differs from others. The bands of aliphatic groups are well identifiable between 3000 and 2800/cm. The extension of the band with a maximum from around 1660/cm to the region of the carboxyl group (1720/cm) is apparent and the band shape differs between 1500 and 1200/cm (polyphenolic substances and functional groups with nitrogen and phosphorus). The band around 920/cm documents a lower content of secondary alumosilicates (apparent in the whole profile), which corresponds to more sandy soil texture in forest. Hence, a large proportion of aliphatic, carboxylic, aromatic, and CH groups under forest correspond with higher organic carbon content in this soil (Gerzabek et al. 2006).

In the deeper layers of forest soil, the shoulder of carboxyl groups in the band with a maximum around 1640/cm is still visible. The dominant peak of the forest soil spectrum is the band around 1300/cm. In the depth 20–30 cm, the band at 1040–945/cm (indicating C-O stretching, Si-OH of alumino-silicate lattice, and carbohydrate region of polysaccharides) is lower under forest than grassland and cropland. The polysaccharides content decrease through the depth of the forest soils was documented by Sugiura et al. (2017) and probable a higher appearance of inorganic materials (Haberhauer et al. 1998).

6.3.2.2. *Spectra of humic acids*

The main bands of HAs and FAs spectra are described in Table 6.4. The dominant peak of these spectra is a peak around 1740/cm, which represents the carboxylic groups on aromatic rings (Figure 6.2). The vibration band of the carboxylic group is typically placed near 1720/cm in the case of substitution on aliphatic chains. In case of substitution on aromatic rings is placed just near 1740/cm (Reddy et al. 2018).

Table 6. 4. The major bands of humic substances (humic and fulvic acids) in infrared spectra (Stevenson 1995, Tatzber et al. 2007, Pavlů and Mühlhanselová 2017)

Wavenumber (1/cm)	Assignment of sorption bands
3400–3300	O–H stretching, N–H stretching
2950–2800	Aliphatic C–H stretching
1725–1710	C=O stretching of COOH and ketones
1660–1630	C=O stretching of amide I, quinone, H-bonded conjugated ketones
1620–1600	Aromatic C=C stretching
1590–1517	N-H bending and C=N stretching (amide II)
1470–1380	Aliphatic C–H bending
1400–1390	OH deformation of CH_2 and CH_3
1280–1200	C–O stretching and OH deformation of COOH, C–O stretching of Aryl esters
1170–950	C–O stretching of polysaccharides

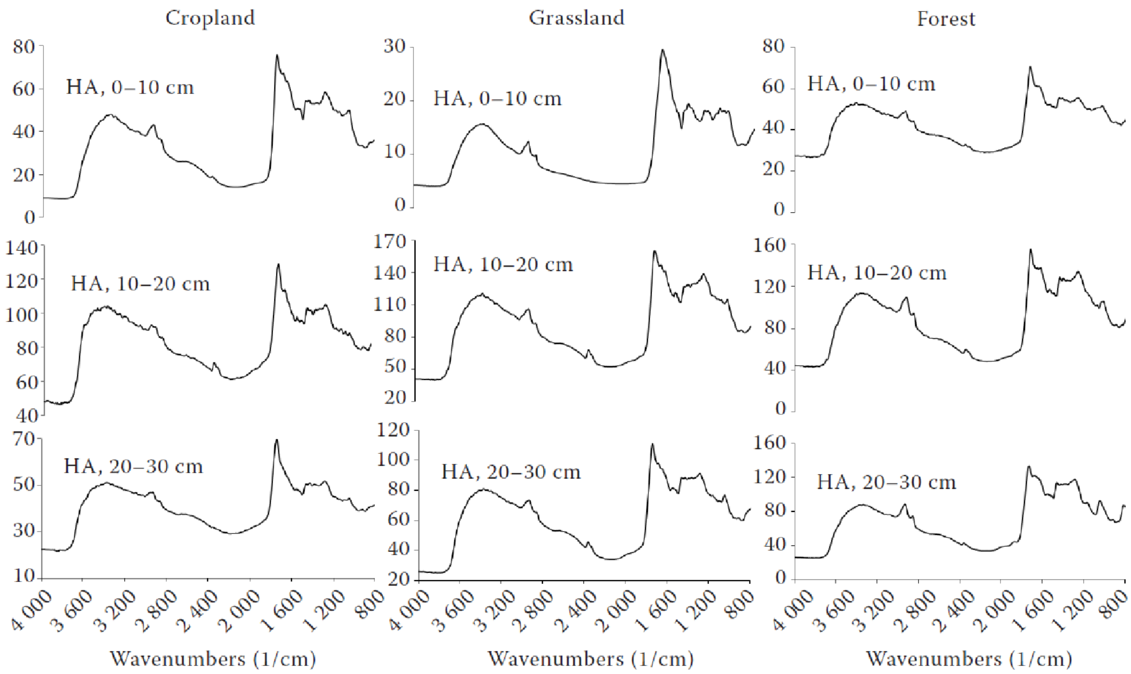


Figure 6. 2. The difference of average humic acids spectra under different depths (0–10, 10–20 and 20–30 cm) and land covers (cropland, grassland, and forest). HA – humic acid

The surface layer of the cropland soil differs from deeper layers, more pronounced peak can be seen around 3000–2800/cm and 1000/cm, which represents higher contents of aliphatic components and polysaccharides chains of HAs. It could point to their lower maturity and stability (Pavlů and Mühlhanselová 2017). The intensity of the C=O group of ketones and amide group (shoulder in range 1690–1630/cm) decreases with soil depth.

The spectrum of the surface layer of the grassland varies by the spectra from other depths and also from other land-use. The band around 1660/cm is dominant, while the carboxyl band is hidden in the spectrum, and the band around 1280/cm is relatively less pronounced in comparison to other spectra. The aliphatic-bending at 1460/cm is shifted to 1425/cm in grassland HAs spectra of the surface soil layer, while in forest and cropland HAs spectra are clearly visible at both positions. It might be the formation of H-bonds between hydroxyl and carboxyl H atoms of HA (Senesi et al. 2001).

HAs spectra of deeper layers of forest soil, differ from others in pronounced aliphatic bands (3000 - 2800/cm). In addition, it could be connected with more sandy substrate in forest as described by Di et al. (2016). The forest HAs spectra have relatively (compared with neighboring band around 1720/cm) the highest peak around 1660/cm in comparison with other land-use and soil depth below 10 cm. The band of aliphatic C-H (1470–1460/cm) is also higher under forest than the grassland and cropland.

6.3.2.3. Spectra of fulvic acids

Generally, the spectra of FAs have a lower amount of peaks in the fingerprint area (Figure 6.3). More details are visible in the spectra of lower parts of the soil profile, where bands of polysaccharide chains and deformation vibrations of OH groups in carboxyl appear.

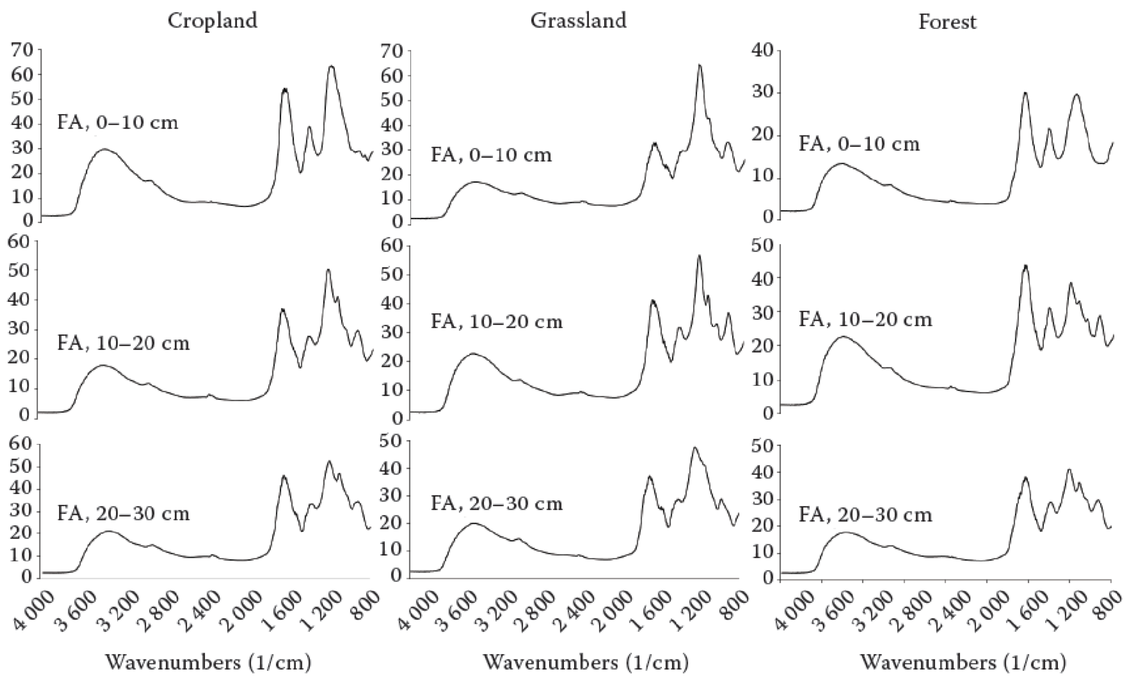


Figure 6. 3. The difference of average fulvic acids spectra under different depths (0–10, 10–20 and 20–30 cm) and land-uses (cropland, grassland, and forest). FA – fulvic acid

The FAs spectra from different soil depths under cropland are quite different. The band at 1670–1600/cm, which mainly characterises carboxyl, ketones, and aromatics, is clearly visible in all three depths. However, the shoulder at 1570–1560/cm (COO- symmetric stretching, N-H deformation, and amides group II) is more pronounced in the deeper layers. The band at 1420–1400/cm (phenols and alcohols) is sharp in the uppermost layer. The spectra of FAs from surface layer of grassland have a bigger amount of peaks in the fingerprint area. The relative intensity of the band 1680–1630/cm and 1100–1200/cm increase with depth. The grassland FAs spectrum differs most from the other land-uses in the depth of 0–10 cm. It has significantly lower peaks around 1660–1600/cm and 1200/cm than cropland and forest. The band at 1560–

1510/cm is more intense under grassland than cropland and forest. Gerzabek et al. (2006) found that aromatic and NH groups were greater in grassland than arable land.

The FAs spectra of the surface forest soil layer have only three wide peaks in the fingerprint area. In the FAs spectra of a deeper layer are visible their splitting on several peaks. The forest FAs spectrum under the depth 10–20 cm has the opposite intensities ratio of bands (1660 ≥ 1200) to the other two land-uses (1660 ≤ 1200). It means that the forest has a higher presence of quinone, ketones, and aromatic C=O than C-O and OH deformation of COOH. It is in accordance with the work of Leinweber et al. (2001).

6.3.2.4. Dissolved organic carbon and low molecular mass of organic acids

On the base of previous results, the most differences among land-use are focused on top parts of the soil profile where LMMOA was mostly found (Hubová et al. 2017). The description of this part of the profile is therefore extended to DOC and LMMOA evaluation (Figure 6.4, Table 6.5).

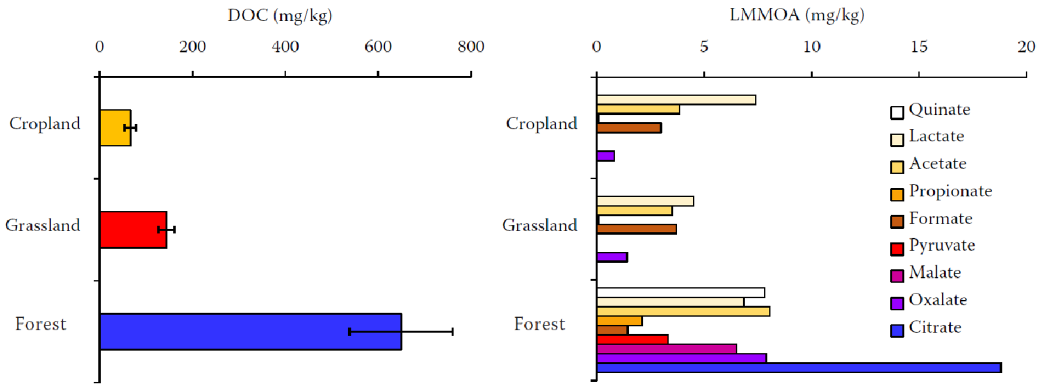


Figure 6. 4. The mean concentrations of dissolved organic carbon (DOC) (error bars show standard deviations) and low molecular mass organic acid (LMMOA) under different land-uses (n = 5)

Table 6. 5. The description of low molecular mass organic acid (LMMOA) and dissolved organic carbon (DOC) concentration under different land-uses in the upper layer (0–10 cm); means ± standard deviation; n = 5

		Cropland	Grassland	Forest
LMMOA (mg/kg)	Quinate	bdl	bdl	7.81 ± 8.13
	Lactate	7.40 ± 6.26	4.52 ± 1.35	6.85 ± 1.57
	Acetate	3.84 ± 2.34	3.51 ± 3.57	8.04 ± 11.89
	Propionate	0.09 ± 0.05	0.08 ± 0.08	2.11 ± 1.67
	Formate	2.99 ± 2.65	3.71 ± 3.89	1.44 ± 0.88
	Pyruvate	bdl	bdl	3.31 ± 2.95
	Malate	bdl	bdl	6.50 ± 5.77
	Oxalate	0.83 ± 0.52	1.41 ± 0.69	7.89 ± 6.63
	Citrate	bdl	bdl	18.8 ± 25.08
DOC (mg/kg)		66 ± 27.36	144 ± 38.55	649 ± 247.92

bdl: below determination limit

The concentration of DOC is relatively high under forest, followed by grassland and cropland. Lower DOC in cropland may result from ploughing, drainage, intensive surface runoff, which cause DOC losses (Manninen et al. 2018). Forest was found to have the highest concentration of LMMOA (citrate, acetate, quinate, oxalate, malate, pyruvate, propionate, formate) followed by grassland and cropland. Citrate concentration is higher under forest while lactate concentration is higher under grassland and cropland. Hubová et al. (2017) showed that more acidic soil contains a higher concentration of citrate. The big value of standard deviation in LMMOA is natural for this slightly stable and highly variable component of soil organic matter. On the other hand, it has a correlation between the amount of LMMOA and DOC with P-value 0.01 ($r = 0.755^{**}$) under all land-uses. The high concentration and amount of LMMOA are based on plant root exudation, residues, and litters decomposition reviewed by (Adeleke et al. 2017, Hubova et al. 2017), and the highest content in forest is as a result of litter decomposition (Berg and McClaugherty 2020) and lower pH in forest area (Rukshana et al. 2014).

6.4. Conclusions

It can be summarized that land-uses influence the amount and qualitative parameters of soil organic matter. Infrared spectroscopy is a useful tool for composition of the SOM evaluation. The most obvious differences in SOM composition according to land-use are evident in surface layer of soil. Forest soil spectra had more intense aliphatic bands (3010– 2800/cm) than the grassland and cropland in the upper layer.

Similarly, the HAs spectra of forest soil have more intense aliphatic bands than the grassland and cropland HAs. More acid characters of organic matter in forest soil are also documented by soil spectra in the intensity of carboxylic bands. The grassland FAs spectrum differs most from the other land-uses in the depth of 0–10 cm. It has significantly lower peaks around 1660–1600/cm and 1200/cm than cropland and forest. In the cropland soils, aromaticity of HAs increase with depth. The concentration of LMMOA was higher under the forest, followed by grassland and cropland. The most abundant acid in LMMOA mixture was citrate in the forest while lactate was in the grassland and cropland.

6.5. References

Adeleke R., Nwangburuka C., Oboirien B. (2017): Origins, roles and fate of organic acids in soils: a review. South African Journal of Botany, 108: 393–406.

Allison V.J., Condron L.M., Peltzer D.A., Richardson S.J., Turner B.L. (2007): Changes in enzyme activities and soil microbial community composition along carbon and nutrient gradients at the Franz Josef chronosequence, New Zealand. Soil Biology and Biochemistry, 39: 1770–1781.

Ash C., Tejnecký V., Borůvka L., Drábek O. (2016): Different low-molecular-mass organic acids specifically control leaching of arsenic and lead from contaminated soil. Journal of Contaminant Hydrology, 187: 18–30.

Berg B., McClaugherty C. (2020): Plant Litter: Decomposition, Humus Formation, Carbon Sequestration. Berlin, Heidelberg, Springer-Verlag. ISBN 978-3-642-38821-7

Davidson E.A., Ackerman I.L. (1993): Changes in soil carbon inventories following cultivation of previously untilled soils. Biogeochemistry, 20: 161–193.

Di X.Y., Dong H., An X.J., Tang H.M., Xiao B.H. (2016): The effects of soil sand contents on characteristics of humic acids along soil profiles. Acta Geochimica, 35: 251–261.

Gerzabek M.H., Antil R.S., Kögel-Knabner I., Knicker H., Kirchmann H., Haberhauer G. (2006): How are soil use and management reflected by soil organic matter characteristics: a spectroscopic approach. European Journal of Soil Science, 57: 485–494.

Haberhauer G., Rafferty B., Strebl F., Gerzabek M.H. (1998): Comparison of the composition of forest soil litter derived from three different sites at various decompositional stages using FTIR spectroscopy. Geoderma, 83: 331–342.

Haberhauer G., Gerzabek M.H. (1999): DRIFT and transmission FT-IR spectroscopy of forest soils: an approach to determine decomposition processes of forest litter. Vibrational Spectroscopy, 19: 413–417.

Hubová P., Tejnecky V., Ash C., Boruvka L., Drabek O. (2017): Low-molecular-mass organic acids in the forest soil environment. Mini-Reviews in Organic Chemistry, 14: 75–84.

Jakšík O., Kodešová R., Kubiš A., Stehlíková I., Drábek O., Kapička A. (2015): Soil aggregate stability within morphologically diverse areas. Catena, 127: 287–299.

Kunlanit B., Butnan S., Vityakon P. (2019): Land-use changes influencing C sequestration and quality in topsoil and subsoil. Agronomy, 9: 1–16.

Lal R. (2005): Forest soils and carbon sequestration. Forest Ecology and Management, 220: 242–258.

Lal R. (2018): Digging deeper: a holistic perspective of factors affecting soil organic carbon sequestration in agroecosystems. Global Change Biology, 24: 3285–3301.

Leinweber P., Schulten H.-R., Kalbitz K., Meißner R., Jancke H. (2001): Fulvic acid composition in degraded fenlands. Journal of Plant Nutrition and Soil Science, 164: 371–379.

Manninen N., Soinne H., Lemola R., Hoikkala L., Turtola E. (2018): Effects of agricultural land use on dissolved organic carbon and nitrogen in surface runoff and subsurface drainage. Science of The Total Environment, 618: 1519–1528.

Matamala R., Calderón F.J., Jastrow J.D., Fan Z.S., Hofmann S.M., Michaelson G.J., Mishra U., Ping C.L. (2017): Influence of site and soil properties on the DRIFT spectra of northern cold-region soils. Geoderma, 305: 80–91.

Muktar M., Bobe B., Kibebew K., Yared M. (2018): Soil organic carbon stock under different land use types in Kersa Sub Watershed, Eastern Ethiopia. African Journal of Agricultural Research, 13: 1248–1256.

Pavlů L., Mühlhanselová M. (2017): Differences in humic acids structure of various soil types studied by DRIFT spectroscopy. Soil and Water Research, 13: 29–35.

Piccolo A. (2001): The supramolecular structure of humic substances. Soil Science, 166: 810–832.

Piccolo A., Celano G., Conte P. (2000): Methods of isolation and characterization of humic substances to study their interactions with pesticides. In: Cornejo J., Jamet P. (eds): Pesticide/Soil Interactions. Paris, Institut National de la Recherche Agronomique (INRA): 103–116.

Reddy S.B., Nagaraja M.S., Kadalli G.G., Champa B.V. (2018): Fourier transform infrared (FTIR) spectroscopy of soil humic and fulvic acids extracted from paddy land use system. International Journal of Current Microbiology and Applied Sciences, 7: 834–837.

Ribeiro J.S., Ok S.S., Garrigues S., de la Guardia M. (2001): FTIR tentative characterization of humic acids extracted from organic materials. Spectroscopy Letters, 34: 179–190.

Rukshana F., Butterly C.R., Xu J.M., Baldock J.A., Tang C.X. (2014): Organic anion-to-acid ratio influences pH change of soils differing in initial pH. Journal of Soils and Sediments, 14: 407–414.

Senesi N., Loffredo E., D'Orazio V., Brunetti G., Miano T.M., La Cava P. (2001): Adsorption of pesticides by humic acids from organic amendments and soils. Soil Science Society of America Journal, 677S: 129–153.

Slepetiene A., Slepetys J. (2005): Status of humus in soil under various long-term tillage systems. Geoderma, 127: 207–215.

Sparks D.L. (1996): Methods of Soil Analysis. Part 3 – Chemical Methods. Madison, SSSA Book Series: 5, Soil Science Society of America, Inc. and American Society of Agronomy Inc.

Strobel B.W. (2001): Influence of vegetation on low-molecular-weight carboxylic acids in soil solution – a review. Geoderma, 99: 169–198.

Strosser E. (2010): Methods for determination of labile soil organic matter: an overview. Journal of Agrobiology, 27: 49–60.

Stevenson F.J. (1995): Humus Chemistry: Genesis, Composition, Reactions. 2th Edition. New York, John Wiley and Sons, Inc. ISBN: 978-0-471-59474-1

Sugiura Y., Watanabe M., Nonoyama Y., Sakagami N., Guo Y., Murayama S. (2017): Saccharides of ectomycorrhizal fungal sclerotia as sources of forest soil polysaccharides. Soil Science and Plant Nutrition, 63: 426–433.

Tatzber M., Stemmer M., Spiegel H., Katzlberger C., Haberhauer G., Mentler A., Gerzabek M.H. (2007): FTIR-spectroscopic characterization of humic acids and humin fractions obtained by advanced NaOH, $Na_4P_2O_7$ and Na_2CO_3 extraction procedures. Journal of Plant Nutrition and Soil Science, 170: 522–529.

Tejnecký V., Drábek O., Nikodem A., Vokurková P., Němeček K., Borůvka L. (2014): Fast determination of water extractable organic carbon from forest soil. Zprávy Lesnického Výzkumu, 59: 155–159.

Tinti A., Tugnoli V., Bonora S., Francioso O. (2015): Recent applications of vibrational mid-infrared (IR) spectroscopy for studying soil components: a review. Journal of Central European Agriculture, 16: 1–22.

Trevisan S., Francioso O., Quaggiotti S., Nardi S. (2010): Humic substances biological activity at the plant-soil interface: from environmental aspects to molecular factors. Plant Signaling and Behavior, 5: 635–643.

Veum K.S., Goyne K.W., Kremer R.J., Miles R.J., Sudduth K.A. (2014): Biological indicators of soil quality and soil organic matter characteristics in an agricultural management continuum. Biogeochemistry, 117: 81–99.

Zaccheo P., Cabassi G., Ricca G., Crippa L. (2002): Decomposition of organic residues in soil: experimental technique and spectroscopic approach. Organic Geochemistry, 33: 327–345.

CHAPTER 7

General Discussion

This doctoral thesis explored the intersection of digital soil mapping (DSM) with geographical information systems (GIS) and spatial statistics. The research delved into soil organic carbon (SOC) distribution in the Czech Republic using digitized data from legacy soil survey maps. The study focused on Liberec and Domažlice districts. The general goal of this research was to comprehend the factors influencing SOC distribution across diverse land-uses and topographic variables, contributing to the advancement of DSM knowledge. Furthermore, it seeked to refine local maps for the Liberec and Domažlice regions and enhance the understanding of SOC spatial patterns.

To achieve this goal, a spatial prediction framework for SOC using DSM was presented and a comparative analysis of machine learning (ML) models was conducted. Key study objectives involved investigating the impact of ML model choice on SOC prediction accuracy and emphasizing the importance of training data quality. Additionally, the research highlighted the comparable suitability of covariates such as the Normalized Difference Vegetation Index (NDVI), land cover and digital elevation models (DEMs). This study also facilitated improvements in ML model selection and training data progression in DSM, contributing to the enhancement of local maps in areas with similar climates.

This study initially examined the correlation between slope, elevation, and clay with SOC content across various land-uses. Also, subsequent evaluations focused on the influence of region and elevation on predictors within SOC prediction models. Then, as the primary objective of this study, DSM and ML methods, including Random Forest (RF), Quantile Random Forest (QRF), and Cubist, were used to generate high-resolution SOC maps at a 100 m spatial resolution. This facilitated the prediction of SOC distribution in the specified districts of the Czech Republic. Finally, the study employed infrared spectroscopy to assess the impact of land-use on soil organic matter (SOM), aiming to comprehend the influence of different land-uses on SOM composition in terms of quality and quantity.

The research outcomes revealed an insignificant correlation between SOC and clay content in the Liberec and Domažlice districts, evidenced by correlation coefficients of r = -0.026 and r = -0.108, respectively. This aligns with findings by Zhong et al. (2018), emphasizing the complex and scale-dependent nature of SOC-clay relationships. The observed lack of a significant correlation may imply the pivotal role of regional variations and soil types in shaping these relationships. Notably, the predominantly low clay content (<20%) in the soil suggests that mineralization processes might not be significantly minimized. Conversely, low SOC levels in soils with low clay content could be attributed to the high decomposition rate of organo-mineral fractions, as highlighted by Lee et al. (2009) and Pronk et al. (2012).

Elevation emerged as a crucial factor influencing SOC content, consistent with the observations of Zhu et al. (2019). SOC consistently increased with rising elevation, particularly pronounced in the Domažlice district. Parent materials were identified as influencers, with soils derived from nutrient-rich materials exhibiting higher fertility and SOC content compared to granitic soils with fewer mineral nutrients. Sandy soils at higher elevations, especially under coniferous forests, demonstrated elevated SOC levels, attributed to reduced mineralization linked to strong acidity, a characteristic of Podzols and some Cambisols. The observed soil acidity likely decreased decomposition rates, contributing to higher SOC content.

The study also highlighted the impact of data separation based on land-use in improving correlations between SOC and slope, which is consistent with other studies, similarly emphasizing the significant influence of land-use on SOC content (Žigová et al., 2017). The present study revealed a more pronounced correlation increase in arable areas in the Liberec district compared to Domažlice. This variation could be attributed to factors such as agricultural practices, tillage, slope, soil biology, and erosion. The study also emphasized the benefits of soil conservation tillage technology in shaping SOC distribution, aligning with previous research by Šíp et al. (2009) and supporting the significance of sustainable land management practices, particularly in erosion reduction and soil organic matter building. Within complex systems, SOC-slope correlations were stronger than in individual arable land and forest categories. This observation is consistent with previous research demonstrating the positive effects of varied arable cropping systems and diverse management plans on SOC in European areas (Francaviglia et al., 2020).

The findings also revealed the presence of multicollinearity among environmental variables, suggesting that relying on simple regression to predict SOC content would be unreliable.

Multiple regression proved effective in improving predictive accuracy. This observation is consistent with observations from Lettens et al. (2005), who, similarly, divided regions into 289 landscape units in Belgium. They demonstrated that SOC stocks were continually influenced by external factors, particularly the history of land-use, regular land management practices, and climate conditions.

The analysis of the correlation between environmental covariates and SOC revealed that most variables exhibited a modest correlation with SOC. However, it is evident that variables demonstrating stronger correlations with SOC are deemed more suitable as robust predictors in SOC models. A thorough investigation into the impact of predictors on SOC predictions highlighted the influence of regional disparities and altitude-dependent variations. Strong predictors at lower altitudes included edaphic series and soil classes, while topography-related predictors gained prominence at higher elevations. The relevance of soil classes as predictors depended on pedodiversity levels, with forest types contributing less if one forest type dominates. Optimal predictions in smaller, consistent regions underscored the importance of influential covariates while acknowledging model limitations and uncertainties. Additionally, the VarImp function in R was also utilized to assess the relative importance of predictor variables in the Liberec and Domažlice districts, resulting in the significance of elevation, land cover, valley depth, and slope. These results underscored the importance of terrain-based covariates and vegetation in explaining SOC variability.

This study also aimed to address the common challenges in SOC prediction models, including low R^2 values and predictive range restrictions observed in various studies (Riggers et al., 2019). In another word, the models usually tend to overestimate low values and underestimate high values, restricting the predictive range. Similar challenges were encountered in studies by Yamashita et al. (2022), Ottoy et al. (2017), and others. Generally, model uncertainties are attributed to data variability, survey methodologies, and uncertainties in bulk density estimation (Minasny et al., 2013; Draper, 1995). This study recommended considering uncertainties arising from field observations, empirical prediction models, and environmental covariates in DSM, emphasizing the importance of comprehensive methodologies and recognizing potential errors or inconsistencies in the database.

This research also identified the RF model as the most accurate algorithm, consistently outperforming Cubist and QRF models in both districts. Utilizing diverse terrain covariates, the RF model demonstrated higher R^2 values, smaller Root Mean Square Error (RMSE), and Mean

Absolute Error (MAE) values. This result aligns with the findings of other studies that endorse RF for its high accuracy in soil spatial modeling (Lamichhane et al., 2019; Ellili et al., 2019).

The study on DSM and SOC predictions using machine learning models, particularly RF, Cubist, and QRF, revealed consistent spatial patterns of SOC content among different models, reporting higher SOC content in elevated forested areas and lower content in areas with replaced plantations and croplands. The positive correlation between increasing elevation and SOC concentrations can be attributed to climatic variables, influencing soil acidification and poor water drainage. The study underscored the importance of terrain-based covariates and vegetation in explaining SOC variability in sub-humid areas.

The uncertainty in SOC predictions, particularly in coniferous forests, was acknowledged, and the need for additional samples in these areas was recommended. The study underscored the importance of considering uncertainties arising from field observations, empirical prediction models, and environmental covariates in DSM, in line with the work of Vanguelova et al. (2016) and Samuel-Rosa et al. (2015).

The investigation into DSM and SOC predictions, using machine learning models such as RF, Cubist, and QRF, provided valuable insights. While Mosleh et al. (2016) highlighted the use of parameters from DEMs for modeling soil properties in low-relief areas, the present study took a different approach. Instead of relying solely on DEMs, the importance of predictor variables in SOC modeling was assessed. In simpler terms, the study challenged the typical use of DEMs and emphasized the roles of alternative covariates such as the NDVI and land cover. In the Liberec district, important variables for predicting SOC using RF and Cubist models were land cover, elevation, valley depth, and slope, aligning with findings of Ellili et al. (2019) that identified slope and elevation as crucial variables. This again highlights the significance of land cover in predicting SOC. Topographic factors such as slope and valley depth were also important, impacting water distribution, runoff, erosion, and deposition, influencing the spatial variation of SOC in mountainous areas. Steeper and longer slopes generally led to faster water runoff and increased erosion potential. In the Domažlice district, land cover and elevation were key variables for predicting SOC using RF and Cubist models. This again underscores the consistent importance of land cover and topography across different regions.

The infrared spectroscopy study revealed insignificant differences in pH_{KCl} across various depths in different land-uses. However, different land-uses showed different pHs with cropland exhibited neutrality, grassland showed moderate acidity, and forest displayed strong acidity. Evaluation of SOM quality through the E4/E6 method indicated no significant differences across depths, except for the forest, suggesting a lower degree of humification processes. Additionally, SOC content variations among depths were observed in all land-uses. Spectral analysis of soil bands highlighted similarities in cropland and grassland, with distinct features in forest soil, particularly in aliphatic components. This aligns with previous research (Kunlanit et al., 2019; Hubová et al., 2017; Gerzabek et al., 2006), reinforcing the consistency and broader implications of observed variations in soil characteristics and spectral features across different land-uses and depths. The pivotal role of land-use in shaping SOM composition underscores the importance of integrating diverse land-uses in predicting SOC.

In conclusion, the research significantly contributes to the understanding of SOC distribution in the Czech Republic. The integrative approach, covering correlations with slope and elevation, the impact of land-use on SOC model predictors, DSM and ML predictions, and the insights from the infrared spectroscopy study, underscored the complexity and interconnected nature of factors influencing SOC in the Czech Republic, emphasizing the necessity for comprehensive methodologies considering diverse land-uses, topographic variables, and influential covariates to ensure precise predictions of SOC. The integration of DSM and machine learning techniques, particularly RF, Cubist, and QRF models, proved effective in capturing general trends in SOC distribution across diverse landscapes. The study provides valuable insights for predicting SOC, acknowledging and addressing challenges associated with model uncertainties and variations in environmental covariates.

CHAPTER 8

Conclusions

This chapter consolidates the research conclusions, presenting distinctive insights into soil science within the Czech Republic. The combined findings offer a holistic understanding of the factors shaping soil organic carbon (SOC) distribution across diverse land-uses and topographic variables. Successfully achieving the primary objective of predicting SOC through digital soil maps (DSMs) using three machine learning models in the Liberec and Domažlice districts, this study delves into the complex intersection of DSM, geographical information systems (GIS), and spatial statistics. The research contributes significantly to DSM knowledge, unraveling the complex relationship between land-uses, topography, and environmental predictors, with a particular emphasis on the substantial impact of land-use on SOC distribution.

The analysis of the effects of slope, elevation, and clay on SOC content in diverse land-uses revealed nuanced relationships. While the correlation between clay and SOC lacked significant correlation, elevation and slope showed a moderately strong positive influence on SOC, suggesting increased leaching and reduced decomposition at higher elevations. Multiple linear regression models, incorporating topographical variables, exhibited stronger correlations in the Domažlice district, underlining the influence of region-specific characteristics. The study also underscored the substantial impact of land-use variations on SOC, emphasizing the necessity of incorporating these variations into predictive models. Furthermore, the assessment of the environmental predictor's importance showed variations based on region and elevation, with edaphic series and soil classes emerging as strong predictors at lower elevations and topography-related predictors gaining prominence at higher elevations.

Despite uncertainties in predictive modeling, the study effectively captured general trends and factors influencing SOC content distribution, with the random forest (RF) model consistently outperforming Cubist and Quantile Random Forest (QRF) models. The critical role of land-use and elevation in SOC prediction was highlighted, particularly in forest-dominated areas. The study also recommended additional observations and stratified random sampling to enhance model accuracy and minimize uncertainty, particularly in coniferous forests. Additionally, the exploration of land-use effects on soil organic matter (SOM) using infrared spectroscopy revealed the significant influence of land-use on SOM quantity and quality, emphasizing the pivotal role of land-use in shaping SOM composition. Overall, these findings contributed valuable insights for predicting SOC, addressing challenges, and recognizing the multifaceted nature of soil dynamics influenced by topography, land-use, and environmental variables.

CHAPTER 9

Additional Information

9.1. Project Outputs and Published Journal Papers

The results of this research were published in five journal articles. Maps of spatial prediction and spatial databases will be another output of the projects. The projects will also enable updating and harmonizing the current soil databases and maps in the Czech Republic.

List of the publications:

1. **Nozari S.**, Pahlavan-Rad M. R., Brungard C., Heung B., and Borůvka L. (2023): Digital Soil Mapping using Machine Learning-Based Methods to Predict Soil Organic Carbon in Two Different Districts in the Czech Republic. Soil & Water Res., XX: 00–00. (in press)
   (Nozari S. conurbation: 80% - complete manuscript preparation from data processing, computation, writing, and editing.)

2. **Nozari S.**, Borůvka L. (2023): The effects of slope and altitude on soil organic carbon and clay content in different land-uses: A case study in the Czech Republic. Soil & Water Res., 18: 204–218. https://doi.org/10.17221/105/2022-SWR
   (Nozari S. conurbation: 90% - complete manuscript preparation from data processing, computation, writing, and editing.)

3. Borůvka L., Vašát R., Šrámek V., Neudertová Hellebrandová K., Fadrhonsová V., Sáňka M., Pavlů L., Sáňka O., Vacek O., Němeček K., **Nozari S.**, Oppong Sarkodie V.Y. (2022): Predictors for digital mapping of forest soil organic carbon stocks in different types of landscape. Soil & Water Res., 17: 69−79. https://doi.org/10.17221/4/2022-SWR
   (Nozari S. conurbation: 5%)

4. Thai S., Pavlů L., Tejnecký V., Vokurková P., **Nozari S.**, Borůvka L. (2021): Comparison of soil organic matter composition under different land-uses by DRIFT spectroscopy. Plant Soil Environ., 67: 255–263. https://doi.org/10.17221/11/2021-PSE

(Nozari S. conurbation: 5%)

5. **Nozari S.**, Borůvka L. (2020). The effect of landscape slope on soil organic carbon in Liberec district in the Czech Republic. International Journal of Advanced Science and Technology, 29(9s): 3934 - 3942. http://sersc.org/journals/index.php/IJAST/article/view/16653

   (Nozari S. conurbation: 90% - complete manuscript preparation from data processing, computation, writing, and editing.)

9.2. Financial Support

A project of the Czech Science Foundation No. 17-27726S on the "Spatial prediction of soil properties and classes based on the position in the landscape and other environmental covariates" (2017-2019) and a project of the Ministry of Agriculture QK192016 on the “Development and verification of spatial models of forest soil properties in the Czech Republic“ are acknowledged for support of this research. Support from the European Structural and Investment Funds, Operational Programme Research Development and Education Project No. CZ.02.1.01/0.0/0.0/16_019/0000845 “Centre of the investigation of synthesis and transformation of nutritional substances in the food chain in interaction with potentially risk substances of anthropogenic origin: comprehensive assessment of the soil contamination risks for the quality of agricultural products” (Centre of Excellence NutRisk) is also acknowledged.

CHAPTER 10

References

Abraham, A. 2005. Rule-based Expert Systems. Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn. John Wiley & Sons, Ltd. ISBN: 0-470-02143-8. 910-919.

Aksoy, E., Özsoy, G., Dirim, S.M. 2009. Soil mapping approach in GIS using Landsat satellite imagery and DEM data. African Journal of Agricultural Research, 4, 1295-1302.

Bell, J.C., Cunningham, R.L., Havens, M.W. 1992. Calibration and validation of a soil-landscape model for predicting soil drainage class. Soil Science Society of America Journal, 56, 1860- 1866.

Bell, J.C., Cunningham, R.L., Havens, M.W. 1994a. Soil drainage class probability mapping using a soil-landscape model. Soil Science Society of America Journal, 58, 464-470.

Bell, J.C., Thompson, J.A., Butler, C.A., and McSweeney, K. 1994b. Modeling soil genesis from a landscape perspective. In Transactions of the Fifteenth World Congress of Soil Science, Acapulco, Mexico, 6A. Madison, WI: International Society of Soil Science, 179-190.

Berkal, I., Walter, C., Michot, D., Djili, K. 2014. Seasonal monitoring of soil salinity by electromagnetic conductivity in irrigated sandy soils from a Saharan oasis. Soil Research 52, 769-780.

Bolstad, P. V., Stowe, T. 1994. An evaluation of DEM accuracy: elevation, slope, and aspect. Photogrammetric Engineering and Remote Sensing, 60, 1327-32.

Borrelli, P., Ballabio, C., Panagos, P., Montanarella, L. 2014. Wind erosion susceptibility of European soils. Geoderma 232-234, 471-478.

Borůvka, L. 2010. Effect of covering with natural topsoil as a reclamation measure on mining dumpsites. In Proceedings of the 19th World Congress of Soil Science: Soil solutions for a changing world, Brisbane, Australia. Working Group 3.3 Soils in urban and industrial areas, 17-20.

Borůvka, L., Kozak, J. 2001. Geostatistical investigation of a reclaimed dumpsite soil with emphasis on aluminum. Soil and Tillage Research, 59, 115-12.

Brabyn, L. 1997. Classification of macro landforms using GIS. ITC Journal, 14, 26-40.

Breiman, L. 2001. Random forests. Machine Learning, 45(1), 5-32.

Brewer, S., Liaw, A., Wiener, M., Liaw, M.A. 2015. Package Random Forest. Breiman and Cutler's Random Forests for Classification and Regression. R package version 4.6-14. Available on: URL: https://cran.r-project.org/web/packages/randomForest/index.html

Brough, P.A. 1986. Principle of Geographical Information Systems for Land Resources Assessment.194. Oxford University Press.

Brungard, C., Boettinger, J.L., Duniway, M.C., Wills, S.A., Edwards, T.C. 2015. Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma, 239, 68-83.

Bui, E.N., Moran, C.J. 2001. Disaggregation of polygons of surficial geology and soil maps using spatial modeling and legacy data. Geoderma, 103 (1), 79-94.

Bui, E.N., Moran, C.J. 2003. A strategy to fill gaps in soil survey over large spatial extents: an example from the Murray-Darling basin of Australia. Geoderma, 111, 21-44.

Burrough, P.A., Bouma, J., Yates, S.R., 1994. The state of the art in pedometrics. Geoderma 62, 311-326.

Carré, F., McBratney, A.B., Mayr, T., Montanarella, L. 2007. Digital soil assessments: Beyond DSM. Geoderma, 142, 69-79.

Dewitte, O., Jones, A., Spaargaren, O., Breuning-Madsen, H., Brossard, M., Dampha, A., Deckers, J., Gallali, T., Hallett, S., Jones, R., Kilasara, M., le Roux, P., Michéli, E., Montanarella, L., Thiombiano, L., van Ranst, E., Yemefack, M., Zougmore, R. 2013. Harmonisation of the soil map of Africa at the continental scale. Geoderma, 211-212, 138-153.

Díaz-Uriarte, R., Alvarez de Andrés, S. 2006. Gene selection and classification of microarray data using Random Forest. BMC Bioinformatics 7, 3-15.

Dietrich, W.E., Reiss, R., Hsu, M.L., Montgomery, D.R. 1995. A process-based model for colluvial soil depth and shallow land sliding using digital elevation model data. Hydrological Processes, 9, 383-400.

Dobos, E., Carré, F., Hengl, T., Reuter, H.I., Tóth, G. 2006. Digital Soil Mapping as a support to production of functional maps. EUR 22123 EN, 68. Office for Official Publications of the European Communities, Luxemburg.

Dobos, E., Micheli, E., Baumgardner M.F., Biehl, L., Helt, T. 2000. Use of combined digital elevation model and satellite radiometric data for regional soil mapping. Geoderma, 97, 367-391.

Dorigo, W.A., Zurita-Milla, R., de Wit, A.J.W., Brazile, J., Singh, R., Schaepman, M.E. 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. International Journal of Applied Earth Observation 9, 165-193.

Draper, D. 1995. Assessment and propagation of model uncertainty. Journal of the Royal Statistical Society Series B: Statistical Methodology, 57(1), 45-70.

Eleanor, M., Banwart S. A., Noellemeyer, E. 2013. Soil carbon, multiple benefits. Environmental Development, 13, 33-38.

Elfaki, J.T., Gafer, M.A., Sulieman, M.M., Ali, M.E. 2016. Hydrometer method against the pipette method for estimating soil particle size distribution in some soil types selected from Central Sudan. International Journal of Engineering Research and Advanced Technology, 2(2), 25-41.

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.

European Commission - Joint Research Centre (JRC). 2019. Estimated soil erosion by water, by erosion level, land cover, and NUTS 3 regions. Eurostat, Available on: http://data.europa.eu/88u/dataset/l3afxzhrogvuio1xzwjlw

Faeth, P., Crosson, P. 1994. Building the case for sustainable agriculture. Environment, 36, 16-20.

FAO and Global Soil Partnership. 2016. Global Soil Partnership Pillar 4 Implementation Plan: Towards a Global Soil Information System. Pillar 4 Working Group, 38.

Francaviglia, R., Álvaro-Fuentes, J., Di Bene, C., Gai, L., Regina, K. and Turtola, E. 2020. Diversification and management practices in selected European regions. A data analysis of arable crops production. Agronomy, 10(2), 297.

Gallant, JC., Wilson, JP. 2000. Primary topographic attributes. In Terrain Analysis: Principles and Applications, Wilson JP, Gallant JC (eds). John Wiley, New York, 51 – 85.

Gerzabek, M.H., Antil, R.S., Kögel-Knabner, I., Knicker, H., Kirchmann, H., Haberhauer, G. 2006. How are soil use and management reflected by soil organic matter characteristics: a spectroscopic approach. European Journal of Soil Science, 57, 485-494.

Gessler, PE., Moore, ID., McKenzie, NJ., Ryan, PJ. 1995. Soil-landscape modeling and spatial prediction of soil attributes. International Journal of Geographical Information Systems, 9, 421-43.

Goovaerts, P. 1999. Geostatistics in soil science: state-of-the-art and perspectives. Geoderma, 89, 1-46.

Green, K. 1992. Spatial imagery and GIS. Journal of Forestry, 90(11), 32-36.

Grimm, R., Behrens, T., Märker, M., Elsenbeer, H. 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island-digital soil mapping using Random Forests analysis. Geoderma, 146(1), 102-113.

Grinand, C., Arrouays, D., Laroche, B., Martin, M.P. 2008. Extrapolating regional soil landscapes from an existing soil map: sampling intensity, validation procedures, and integration of spatial context. Geoderma, 143, 180-190.

Guevara, M., Olmedo, G.F., Stell, E., Yigini, Y., Aguilar Duarte, Y., Arellano Hernández, C., Arévalo, G.E., Arroyo-Cruz, C.E., Bolivar, A., Bunning, S., Bustamante Cañas, N., Cruz-Gaistardo, C.O., Davila, F., Dell Acqua, M., Encina, A., Figueredo Tacona, H., Fontes, F., Hernández Herrera, J.A., Ibelles Navarro, A.R., Loayza, V., Manueles, A.M., Mendoza Jara, F., Olivera, C., Osorio Hermosilla, R., Pereira, G., Prieto, P., Alexis Ramos, I., Rey Brina, J.C., Rivera, R., Rodríguez-Rodríguez, J., Roopnarine, R., Rosales Ibarra, A., Rosales Riveiro, K.A., Schulz, G.A., Spence, A., Vasques, G.M., Vargas, R.R., Vargas, R. 2018. No Silver Bullet for Digital Soil Mapping: Country-specific Soil Organic Carbon Estimates across Latin America. Copernicus GmbH.

Guo, P-T., Li, M-F., Luo, W., Tang, Q-F., Liu, Z-W., Lin, Z-M. 2015. Digital mapping of soil organic matter for rubber plantation at regional scale: an application of Random Forest plus residual kriging approach. Geoderma 237-238, 49-59.

Häring, T., Dietz, E., Osenstetter, S., Koschitzki, T., Schröder, B. 2012. Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. Geoderma, 185, 37-47.

Harms, B., Brough, D., Philip, S., Bartley, R., Clifford, D., Thomas, M., Willis, R., Gregory, L. 2015. Digital soil assessment for regional agricultural land evaluation. Global Food Security 54, 25-36.

Hartemink, A E., McBratney, A.B. 2008. A Soil Science Renaissance. Geoderma, 148, 123-29.

Henderson BL, Bui EN, Moran CJ, Simon DAP. 2005. Australia-wide predictions of soil properties using decision trees. Geoderma, 124, 383-398.

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(8), e105992.

Hengl, T., Heuvelink, G. B., Kempen, B., Leenaars, J.G., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., Mendes de Jesus, J., Tamene, L., Tondoh, J. E. 2015. Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. PloS One, 10(6), e0125814.

Heung, B., Bulmer, C.E., Schmidt, M.G. 2014. Predictive soil parent material mapping at a regional scale: a random forest approach. Geoderma, 214, 141-154.

Hewitt, A.E. 1993. Predictive modelling in soil survey. Soils and Fertilizers, 56(3), 305-314.

Hole, F.D. 1978. An approach to landscape analysis with emphasis on soils. Geoderma 21, 1-23.

Hole, F.D., Campbell, J.B. 1985. Soil Landscape Analysis. Rowman and Allanheld, Totowa, NJ.

Hubová P., Tejnecky V., Ash C., Boruvka L., Drabek O. 2017. Lowmolecular-mass organic acids in the forest soil environment. Mini-Reviews in Organic Chemistry, 14, 75-84.

Hudson, B.D. 1992. The soil survey as paradigm-based science. Soil Science Society of America Journal, 56, 836-841.

IUSS Working Group WRB. 2015. World Reference Base for Soil Resources 2014, Update 2015. International Soil Classification System for Naming Soils and Creating Legends for Soil Maps. World Soil Resources Reports No. 106, Rome: FAO.

Izaurralde, R. C., J. R. Williams, W. B. McGill, N. J. Rosenberg, and M. C. Quiroga Jakas. 2006. Simulating soil C dynamics with EPIC: Model description and testing against long-term data. Ecological Modelling, 192(3-4): 362-384.

Jenny, H. 1941. Factors of Soil Formation: A System of Quantitative Pedology. McGraw-Hill, NY.

Kidd, D., Webb, M., Malone, B., Minasny, B., McBratney, A.B. 2015. Digital soil assessment of agricultural suitability, versatility and capital in Tasmania, Australia. Geoderma Regional 6, 7-21.

Kozák, J., Němeček, J. Jetmar, M. 1996. The database of soil information system – PUGIS. Rostlinná Výroba, 42, 529-534.

Krawczynski, R., Milne, E., Banwart, S. 2015. Soil carbon, multiple benefits. Journal of Environmental Research and Development, 13, 33-38.

Krige, D.G. 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. J. Chem. Metall. Min. Soc. S. Africa, 52 (6), 119-139.

Kuhn, M., Weston, S., Keefer, C., Coulter, N., Quinlan, R. 2013. Cubist: Rule-and instance-based regression modeling. R package version 0.0, 13, CRAN: Wien, Austria.

Kunlanit, B., Butnan, S. and Vityakon, P. 2019. Land–use changes influencing c sequestration and quality in topsoil and subsoil. Agronomy, 9(9), 520.

Lacoste, M., Lemercier, B., Walter, C. 2011. Regional mapping of soil parent material by machine learning based on point data. Geomorphology 133, 90-99.

Lagacherie. P., McBratney, A.B., Voltz, M. 2006. Digital soil mapping: an introductory perspective, 31. Elsevier, 31, 1-600.

Lal, R. 2005. Forest soils and carbon sequestration. Forest Ecology and Management, 220, 242-25.

Lamichhane, S., Kumar, L. and Wilson, B. 2019. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. Geoderma, 352, 395-413.

Lark, R.M., Beckett, P.H.T. 1998. A geostatistical descriptor of the spatial distribution of soil classes, and its use in predicting the purity of possible soil map units. Geoderma 83, 243-267.

Lee, S.B., Lee, C.H., Jung, K.Y., Do Park, K., Lee, D. and Kim, P.J. 2009. Changes of soil organic carbon and its fractions in relation to soil physical properties in a long-term fertilized paddy. Soil and tillage research, 104(2), 227-232.

Lemercier, B., Lacoste, M., Loum, M., and Walter, C. 2012. Extrapolation at regional sale of local soil knowledge using boosted classification trees: a two-step approach. Geoderma 171-172, 75-84.

Lettens, S., Van Orshoven, J., van Wesemael, B., De Vos, B. and Muys, B. 2005. Stocks and fluxes of soil organic carbon for landscape units in Belgium derived from heterogeneous data sets for 1990 and 2000. Geoderma, 127(1-2), 11-23.

Lim, K., Woods, M., Pitt, D., Nesbitt, D., Etheridge, D. 2012. LiDAR sampling density for forest resource inventories in Ontario, Canada. Remote Sensing 4, 830-848.

Ließ, M., Glaser, B., Huwe, B. 2012. Uncertainty in the spatial prediction of soil texture: comparison of regression tree and Random Forest models. Geoderma, 170, 70-79.

Liu, T.-L., Juang, K.-W., Lee, D.-Y. 2015. Interpolating soil properties using kriging combined with categorical information of soil maps. Soil Science Society of America Journal, 70(4), 1200-1209.

Lorenzetti, R., Roberto, B., Fantappiè, M., Costantini, E. 2014. Comparing different approaches data mining, geostatistic, and deterministic pedology to assess the frequency of WRB reference soil groups in the Italian Soil Regions. Geophysical Research Abstracts, 15, EGU2013-11171.

Malone, B.P., Minasny, B., Odgers, N.P., McBratney, A.B. 2014. Using model averaging to combine soil property rasters from legacy soil maps and from point data. Geoderma, 232, 34-44.

Marchetti, A., Piccini, C., Francaviglia, R., Santucci, S., Chiuchiarelli, I. 2010. Estimating soil organic matter content by regression kriging. In Digital Soil Mapping: Bridging Research, Environmental Application, and Operations. Progress in Soil Science 2.

Martin, M.P., Orton, T.G., Lacarce, E., Meersmans, J., Saby, N.P.A., Paroissien, J.B., Jolivet, 1 C., Boulonne, L., Arrouays, D. 2014. Evaluation of modeling approaches for 2 predicting the spatial distribution of soil organic carbon stocks at the national scale. Geoderma, 223, 97-107.

Martin, M.P., Wattenbach, M., Smith, P., Meersmans, J., Jolivet, C., Boulonne, L., Arrouays, 40 D. 2011. Spatial distribution of soil organic carbon stocks in France. Biogeosciences, 8(5), 1053-1065.

Matheron, G. 1963. Principles of geostatistics. Economic Geology, 58, 1246-1266.

McBratney, A.B. Mendonça-Santos, M. L., Minasny, B. 2003. On digital soil mapping. Geoderma, 117, 3-52.

McBratney, A.B., Odeh, I.O.A., Bishop, T.F.A., Dunbar, M.S., Sha tar, T.M. 2000. An overview of pedometric techniques for use in soil survey. Geoderma, 97, 293-327.

McKenzie, N.J., Gessler, P.E., Ryan, P.J., O'Connel, D. 2000. The role of terrain analysis in soil mapping. In: Wilson, J.P., Gallant, J.C. (Eds.), Terrain Analysis Principles and Applications. Wiley, New York, 245-265.

McKenzie, N.J. Ryan, P.J. 1999. Spatial prediction of soil properties using environmental correlation. Geoderma, 89, 67-94.

McSweeney, K., Gessler, P. E., Hammer, J., Bell, J. and Peterson, G. W. 1994. Towards a new framework for modeling the soil landscape continuum. In Bryant, R.B. and Arnold, R.W, editors, Factors of soil formation: a fiftieth anniversary retrospective. Soil Science Society of America Special Publication 39. Madison WI: Soil Science Society of America, 127-145.

Miko, L., Hošek, M. 2009. The state of nature and the landscape in the Czech Republic. First edition. Prague. Agency for Nature Conservation and Landscape Protection of the Czech Republic, 102.

Minasny, B., McBratney, A.B. 2010. Methodologies for global soil mapping. In: Boettinger, J.L., Howell, D.W., More, A.C., Hartemink, A.E., Kienast-Brown, S. (Eds.), Digital Soil Mapping: Bridging Research, Environmental Application, and Operation. Springer, London, 429-443.

Minasny, B., McBratney, A., Malone, B., Wheeler, I. 2013. Digital Mapping of Soil Carbon. Advances in Agronomy, 118, 1-47.

Minasny, B., Vrugt, J.A., McBratney, A.B. 2011. Confronting uncertainty in model-based geostatistics using Markov chain Monte Carlo simulation. Geoderma, 163, 150-162.

Mishra, U., Lal, R., Slater, B., Calhoun, F., Liu, D., Van Meirvenne, M. 2009. Predicting soil organic carbon stock using profile depth distribution functions and ordinary kriging. Soil Science Society of America Journal, 73(2), 614-621.

Moghimi, S., Parvizi, Y., Mahdian, M., Masihabad, M. 2015. Comparison of applying multilinear regression analysis and artificial neural network methods for simulating the topographic factors' effect on soil organic carbon. Journal of Watershed Engineering and Management, 6(4), 312-322.

Moore, I.D., Gressler, P.E., Nielsen, G.A., and Peterson, G.A. 1993. Soil attribute prediction using terrain analysis. Soil Science Society of America Journal, 57, 443-52.

Mosleh, Z., Salehi, M.H., Jafari, A., Borujeni, I.E. and Mehnatkesh, A. 2016. The effectiveness of digital soil mapping to predict soil properties over low-relief areas. Environmental monitoring and assessment, 188, 1-13.

Mulder, V.L., de Bruin, S., Schaepman, M.E., Mayr, T.R. 2011. The use of remote sensing in soil and terrain mapping - A review. Geoderma, 162(1-2), 1-19.

Muñoz, J., Felicísimo, Á.M. 2004. Comparison of statistical methods commonly used in predictive modeling. Journal of Vegetation Science, 15(2), 285–292.

Němeček, J. 2000. The status of soil mapping in the Czech Republic. The European Soil Information System. World Soil Resources Reports, 91, 61-65.

Němeček, T., Dubois, D., Huguenin-Elie, O., Gaillard, G. 2011. Life cycle assessment of Swiss farming systems: Integrated and organic farming. Agricultural Systems, 104. 217-232.

Němeček, J., Tomášek, M. 1983. Geography of soils of the Czech Republic. Academia, Studie ČSAV, 100.

Odgers, N.P., Sun, W., McBratney, A.B., Minasny, B., Clifford, D. 2014. Disaggregating and harmonizing soil map units through resampled classification trees. Geoderma, 214, 91-100.

Omuto, C., Nachtergaele, F., Vargas, R. 2012. State of the Art Report on Global and Regional Soil Information: Where are we? Where to go? Report, Food and Agriculture Organization of the United Nations (FAO), Available on: http://www.fao.org/docrep/017/i3161e/i3161e.pdf

Ottoy, S., De Vos, B., Sindayihebura, A., Hermy, M., Van Orshoven, J. 2017. Assessing soil organic carbon stocks under current and potential forest cover using digital soil mapping and spatial generalisation. Ecological Indicators, 77, 139-150.

Parton, W., Gutmann, M., Ojima, D. 2007. Long-term trends in population, farm income, and crop production in the Great Plains. BioScience, 57(9), 737–747.

Peters, J., De Baets, B., Verhoest, N.E., Samson, R., Degroeve, S., De Becker, P., Huybrechts, W. 2007. Random forests as a tool for ecohydrological distribution modeling. Ecological Modelling, 207(2), 304-318.

Peterson, E., Ver, H., Jay, M. 2010. A mixed-model moving-average approach to geostatistical modeling in stream networks. Publications, Agencies and Staff of the U.S. Department of Commerce, 186.

Pennock, D.J., Zebarth, B.J., De Jong, E. 1987. Landform classification and soil distribution in hummocky terrain, Saskatchewan, Canada. Geoderma 40, 297- 315.

Pimentel, D., Burgess, M. 2013. Soil erosion threatens food production. Geoderma, 3, 443-463.

Pronk, G.J., Heister, K., Ding, G.C., Smalla, K. and Kögel-Knabner, I. 2012. Development of biogeochemical interfaces in an artificial soil incubation experiment; aggregation and formation of organo-mineral associations. Geoderma, 189, 585-594.

Pospíšil, F. 1964. Fractionation of humus substances of several soil types in Czechoslovakia. Rostlinna Vyroba, 10, 567-579.

Qi, Y. 2012. Random forest for bioinformatics. Ensemble Machine Learning. Springer, 307-323.

Qi, F., Zhu, A.X. 2003. Knowledge discovery from soil maps using inductive learning. International Journal of Geographical Information Systems 17, 771-795.

Quinlan, J.R. 1992. Learning with continuous classes, 5th Australian Joint Conference on artificial intelligence. Singapore, 343-348.

Quinlan, J.R. 1993. C4. 5: Programs for Empirical Learning. Morgan Kaufmann. San Francisco, CA.

R Development Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Riggers, C., Poeplau, C., Don, A., Bamminger, C., Höper, H., Dechow, R. 2019. Multi-model ensemble improved the prediction of trends in soil organic carbon stocks in German croplands. Geoderma, 345, 17-30.

Rokach, L. 2010. Ensemble-based classifiers. Artificial Intelligence Review, 33, 1-39.

Samuel-Rosa, A., Heuvelink, G.B.M., Vesques, G.M., Anjos, L.H.C. 2015. Do more detailed environmental covariates deliver more accurate soil maps? Geoderma, 243-244: 214-227.

Sanchez, P.A., Ahamed, S., Carré, F., Hartemink, A.E., Hempel, J., Huising, J., Lagacherie, P., McBratney, A.B., McKenzie, N.J., Mendonça Santos, M.L., Minasny, B., Montanarella, L., Okoth, P., Palm, C.A., Sachs, J.D., Shepherd, K.D., Vågen, T-G., Vanlauwe, B., Walsh, M.G., Winowiecki, L.A., Zhang, G-L. 2009. Digital soil map of the world. Science, 325, 680-681.

Sarmadian, F., Keshavarzi, A., Odagiu, A., Zahedi, Gh., Javadkia, H. 2014. Digital mapping of soil organic carbon Based on topographic indicators and remotely sensed data. ProEnvironment, 7, 3-9.

Schaetzl,RJ., Anderson, S. 2005. Soils: Genesis and Geomorphology. Cambridge University Press.

Scull, P., Franklin, J., Chadwick, O. 2005. The application of classification tree analysis to soil type prediction in a desert landscape. Ecological Modeling, 181(1), 1-15.

Selvaradjou, S-K., Montanarella, L., Carre, F., Jones, A., Panagos, P., Ragunath, K-P., Kumaraperumal, R., Natarajan, S. 2007. An innovative approach to updating soil information based on digital soil mapping techniques. EUR 22545 EN. Office for Official Publications of the European Communities, Luxembourg, 34.

Šíp, V., Růžek, P., Chrpová, J., Vavera, R., Kusá, H. 2009. The effect of tillage practice, input level and environment on the grain yield of winter wheat in the Czech Republic. Field Crops Research, 113, 131-137.

Smith, S., Bulmer, C., Flager, E., Frank, G., Filatow, D. 2010. Digital soil mapping at multiple scales in British Columbia, Canada, Program and Abstracts, 4th Global Workshop on Digital Soil Mapping Rome, Italy.

Soil Survey Division Staff. 1993. Soil Survey Manual. Soil Conservation Service. US Department of Agriculture Handbook, 18.

Song, Y., Yang, L., Li, B., Hu, Y., Wang, A., Zhuo, W., Cui, X., Liu, Y. 2017. Spatial prediction of soil organic matter using a hybrid geostatistical model of an extreme learning machine and ordinary kriging. Sustainability, 9, 754.

Svetnik, V., Liaw, A., Tong, C., Culberson, C., Sheridan, R.P., Feuston, B.P. 2003. Random Forest: a classification and regression tool for compound classification and QSAR modeling. Journal of Chemical Information and Computer Sciences 43, 1947-1958.

Taghizadeh-Mehrjardi, R., Nabiollahi, K., Kerry, R. 2016. Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. Geoderma, 266, 98-110.

Tahmasebi, P. 2018. Multiple Point Statistics: A Review. In: Daya Sagar B., Cheng Q., Agterberg F. Handbook of Mathematical Geosciences. Springer, Cham.

Thompson, J.A., Nauman, T.W., Odgers, N.P., Libohova, Z., Hempel, J.W. 2012. Harmonization of legacy soil maps in North America: Status, trends, and implications for digital soil mapping efforts. In Digital Soil Assessments and Beyond. CRC Press, Leiden, Netherlands, 97-102.

Tziachris, P., Aschonitis, V., Chatzistathis, T., Papadopoulou, M., Doukas, I.D. 2020. Comparing Machine Learning Models and Hybrid Geostatistical Methods Using Environmental and Soil Covariates for Soil pH Prediction. ISPRS International Journal of Geo-Information, 9(4), 276.

Vanguelova E.I., Bonifacio E., De Vos B., Hoosbeek M.R., Berger T.W., Vesterdal L., Armolaitis K., Celi L., Dinca L., Kjønaas O.J., Pavlenda P., Pumpanen J., Püttsepp Ü., Reidy B., Simončič P., Tobin B., Zhiyanski M. 2016. Sources of errors and uncertainties in the assessment of forest soil carbon stocks at different scales – Review and recommendations. Environmental Monitoring and Assessment, 188, 630.

Victoria, R., Banwart, S., Black, H., Ingram, J., Joosten, H., Milne, E., Noellemeyer, E. 2012. The benefits of soil organic carbon. Emerging issues in our global environment. UNEP Division of Early Warning and Assessment United Nations Environment. UNION, Nairobi 00100, Kenya. 80.

Webster, R., Burrough, P.A. 1972a. Computer-based soil mapping of small areas from sample data - I: Multivariate classification and ordination. Journal of Soil Science 23, 210-221.

Webster, R., Burrough, P.A. 1972b. Computer-based soil mapping of small areas from sample data - II: Classification smoothing. Journal of Soil Science 23, 222-234.

Weier, J., Herring, D. 2000. Measuring Vegetation (NDVI & EVI). NASA Earth Observatory, Washington DC.

Wiesmeier, M., Barthold, F., Blank, B., Kögel-Knabner, I. 2011. Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. Plant and Soil, 340(1-2), 7-24.

Woods, M., Pitt, D., Penner, M., Lim, K., Nesbitt, D., Etheridge, D., Treitz, P. 2011. Operational implementation of LiDAR inventory in Boreal Ontario. The Forestry Chronicle, 87, 512-528.

Yamashita, N., Ishizuka, S., Hashimoto, S., Ugawa, S., Nanko, K., Osone, Y., Iwahashi, J., Sakai, Y., Inatomi, M., Kawanishi, A. and Morisada, K. 2022. National-scale 3D mapping of soil organic carbon in a Japanese forest considering microtopography and tephra deposition. Geoderma, 406, 115534.

Yigini, Y., Panagos, P. 2016. Assessment of soil organic carbon stocks under future climate and land cover changes in Europe. Science of the Total Environment, 557-558, 838-850.

Zhao, Y.V., Shi, X.Z. 2010. Spatial prediction and uncertainty assessment of soil organic carbon in Hebei Province, China. In Digital Soil Mapping: Bridging Research, Environmental Application, and Operations. Progress in Soil Science 2.

Zhong, Z., Chen, Z., Xu, Y., Ren, C., Yang, G., Han, X., Ren, G., Feng, Y. 2018. Relationship between Soil Organic Carbon Stocks and Clay Content under Different Climatic Conditions in Central China. Forests, 9(10), 598.

Zhu, A.X. 1997a. A similarity model for representing soil spatial information. Geoderma, 77, 217-242.

Zhu, A.X. 2000. Mapping soil-landscape as spatial continua: the neural network approach. Water Resources Research, 36, 663-677.

Zhu, A.X., Band, L.E. 1994. A knowledge-based approach to data integration for soil mapping. Canadian Journal of Remote Sensing, 20, 408-418.

Zhu, M., Feng, Q., Qin, Y., Cao, J., Zhang, M., Liu, W., Deo, R. C., Zhang, C., Li, R., Li, B. 2019. The role of topography in shaping the spatial patterns of soil organic carbon. CATENA, 176, 296-305.

Zhu, H., Hud, W., Jing, Y., Cao, Y. 2018. Soil organic carbon prediction based on scale-specific relationships with environmental factors by discrete wavelet transform, Geoderma, 330, 9-18.

Zhu, A., Hudson, B., Burt, J., Lubich, K., Simonson, D. 2001. Soil Mapping Using GIS, Expert Knowledge, and Fuzzy Logic. Soil Science Society of America Journal, 65, 1463-1472.

Žigová, A., Artemyeva, Z., Šťastný, M. 2017. Conditions of soil cover on Permo-Carboniferous rocks in the Český Brod area 45 years after afforestation. in: Šarapatka, B., and Bednář, M. (Eds.): Degradation and Revitalisation of Soil and Landscape, Olomouc, September 10-13, 2017, Proceedings: 73. Palacký University in Olomouc.

APPENDIX

Table A. 1. Summary statistics for the values of SOC, clay, and terrain parameters for sampling sites in the Liberec district

Variable	Observations	Minimum	Maximum	Mean	Standard deviation
SOC (%)	71	0.42	11.3	2.83	2.50
Clay (%)	71	2.70	24.3	8.58	3.46
Elev (m)	71	239.7	719.1	412.6	98.29
S (radian)	71	0.005	0.534	0.081	0.091
Aspect (radians)	71	0.248	6.038	3.130	1.716
Sin (Aspect)	71	-0.998	0.966	-0.063	0.674
Cos (Aspect)	71	-1.000	0.970	-0.092	0.740
Plan. Cur	71	0.000	0.001	0.000	0.000
PC	71	0.000	0.000	0.000	0.000
CI	71	-24.75	51.05	1.930	9.794
Cat. Area (Km^2)	71	43975	133017453	4156382	18522553
TWI	71	5.929	16.004	9.792	1.880
LSF	71	0.030	16.965	1.648	2.754
CNBL	71	239.04	415.98	328.89	42.69
VDCN	71	3.432	389.25	230.41	87.97
RSP	71	0.000	0.991	0.262	0.222

Table A. 2. Summary statistics for the values of SOC, clay, and terrain parameters for sampling sites in the Domažlice district

Variable	Observations	Minimum	Maximum	Mean	Standard deviation
SOC (%)	67	0.00	9.33	2.83	2.39
Clay (%)	67	2.18	23.7	11.9	4.42
Elev (m)	67	356.3	719.7	481.4	77.45
S (radian)	67	0.003	0.158	0.048	0.039
Aspect (radians)	67	0.070	6.199	2.924	1.761
Sin (Aspect)	67	-1.000	0.999	0.133	0.742
Cos (Aspect)	67	-0.998	0.998	-0.006	0.668
Plan. Cur	67	0.000	0.000	0.000	0.000
PC	67	0.000	0.000	0.000	0.000
CI	67	-26.59	32.31	-1.228	9.516
Cat. Area (Km^2)	67	40263	171505984	4720369	22355584
TWI	67	7.049	19.12	10.48	2.164
LSF	67	0.018	3.061	0.748	0.711
CNBL	67	356.33	501.67	426.39	31.703
VDCN	67	5.321	276.1	131.7	58.59
RSP	67	0.000	0.981	0.270	0.215

Table A. 3. Multiple Correlation matrix of variables of arable land in the Liberec

	Elev	S	Aspect	Cat. Area	TWI	LSF	CNBL	VDCN	RSP	Clay	SOC
Elev	1										
S	0.623	1									
Aspect	0.042	-0.010	1								
Cat. Area	-0.365	-0.133	0.044	1							
TWI	-0.417	-0.471	0.008	0.562	1						
LSF	0.525	0.940	-0.015	-0.026	-0.253	1					
CNBL	0.849	0.459	-0.080	-0.360	-0.186	0.426	1				
VDCN	-0.289	-0.323	-0.531	0.130	0.342	-0.241	-0.131	1			
RSP	0.772	0.608	0.294	-0.228	-0.538	0.466	0.407	-0.676	1		
Clay	-0.019	0.056	-0.354	-0.201	0.057	0.042	0.220	0.117	-0.207	1	
SOC	0.421	0.595	-0.047	0.009	-0.034	0.730	0.338	-0.204	0.357	-0.055	1

Table A. 4. Correlation matrix (Pearson) for the Liberec district

Variables	SOC	Clay	Elev	S	Aspect	Sin (Aspect)	Cos (Aspect)	Plan. Cur	PC	CI	Cat. Area	TWI	LSF	CNBL	VDCN	RSP
SOC	1															
Clay	-0.019	1														
Elev	0.380**	0.042	1													
S	0.525**	-0.025	0.667**	1												
Aspect	-0.115	-0.083	-0.127	-0.140	1											
Sin (Aspect)	0.129	0.039	0.213	0.185	-0.836	1										
Cos (Aspect)	-0.085	-0.063	0.139	0.158	-0.137	0.199	1									
Plan. Cur	0.049	-0.180	0.491**	0.352	-0.049	0.061	0.116	1								
PC	-0.179	-0.161	0.322	0.050	-0.037	0.013	-0.048	0.746**	1							
CI	0.079	-0.272	0.314	0.156	-0.077	0.141	0.103	0.708**	0.591	1						
Cat. Area	0.046	-0.091	-0.254	0.104	0.071	-0.146	-0.185	0.031	-0.047	-0.018	1					
TWI	-0.145	0.135	-0.598	-0.530	0.134	-0.239	-0.228	-0.521	-0.402	-0.490	0.421	1				
LSF	0.508**	-0.005	0.440	0.917**	-0.053	0.069	0.065	0.242	-0.056	0.085	0.310	-0.299	1			
CNBL	0.100	0.304	0.568**	0.249	0.021	-0.017	0.100	0.014	-0.109	-0.070	-0.248	-0.118	0.190	1		
VDCN	-0.206	0.174	-0.487	-0.307	-0.124	-0.012	-0.020	-0.498	-0.526	-0.351	0.171	0.513**	-0.129	0.125	1	
RSP	0.387**	-0.136	0.842**	0.574**	-0.077	0.180	0.077	0.568**	0.490**	0.394	-0.203	-0.667	0.341	0.121	-0.806	1

*,**,*** Correlation is significant at P < 0.05,0.01,0.001, respectively .

SOC - Soil Organic Carbon; Elev - Elevation; S - Slope; Plan. Cur - Plan Curvature; PC - Profile Curvature; CI - Convergence Index; Cat. Area – Total Catchment Area; TWI - Topographic Wetness Index; LSF - Length Slope Factor; CNBL - Channel Network Base Level; VDCN - Vertical Distance to Channel Network; RSP - Relative Slope Position

Table A. 5. Correlation matrix (Pearson) for the Domažlice district

Variables	SOC	Clay	Elev	S	Aspect	Sin (Aspect)	Cos (Aspect)	Plan. Cur	PC	CI	Cat. Area	TWI	LSF	CNBL	VDCN	RSP
SOC	1															
Clay	-0.042	1														
Elev	0.658**	-0.197	1													
S	0.444**	-0.276	0.721	1												
Aspect	0.030	0.007	-0.087	-0.156	1											
Sin (Aspect)	-0.092	-0.014	0.034	0.055	-0.833	1										
Cos (Aspect)	-0.096	-0.057	-0.012	0.097	0.002	-0.057	1									
Plan. Cur	0.253	-0.083	0.449	0.235	-0.339	0.377	-0.063	1								
PC	-0.064	0.107	0.094	0.016	-0.194	0.184	0.003	0.547	1							
CI	0.167	0.023	0.333	0.153	-0.241	0.238	-0.160	0.709	0.374	1						
Cat. Area	-0.134	-0.022	-0.265	-0.156	-0.094	0.060	0.001	-0.075	-0.055	-0.170	1					
TWI	-0.347	0.126	-0.596	-0.685	0.103	-0.124	-0.007	-0.457	-0.240	-0.508	0.569	1				
LSF	0.398**	-0.257	0.644	0.962	-0.109	-0.014	0.047	0.079	-0.097	0.040	-0.091	-0.532	1			
CNBL	0.449**	-0.206	0.718	0.416	0.076	-0.112	0.073	0.031	-0.183	0.053	-0.309	-0.324	0.410	1		
VDCN	-0.060	-0.048	-0.099	-0.103	0.082	-0.047	0.037	-0.239	-0.397	-0.114	0.031	0.093	-0.072	0.181	1	
RSP	0.556**	-0.114	0.805**	0.652**	-0.199	0.121	-0.003	0.521	0.323	0.398	-0.238	-0.588**	0.549**	0.301	-0.528	1

*,**,*** Correlation is significant at $P < 0.05, 0.01, 0.001$, respectively .

SOC - Soil Organic Carbon; Elev - Elevation; S - Slope; Plan. Cur - Plan Curvature; PC - Profile Curvature; CI - Convergence Index; Cat. Area – Total Catchment Area; TWI - Topographic Wetness Index; LSF - Length Slope Factor; CNBL - Channel Network Base Level; VDCN - Vertical Distance to Channel Network; RSP - Relative Slope Position