# Explainable machine learning in soil mapping

## Peeking into the black box

**Dissertation**

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Tobias Rentschler

aus Calw

Tübingen

2020

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

| | |
|---|---|
| Tag der mündlichen Qualifikation: | 26.04.2021 |
| Dekan: | Prof. Dr. Thilo Stehle |
| 1. Berichterstatter: | Prof. Dr. Thomas Scholten |
| 2. Berichterstatter: | Prof. Dr. Volker Hochschild |

*I propose to consider the question,*
*'Can machines think?'*

Alan M. Turing, 1950

# Table of contents

# List of Figures

## List of Publications in this thesis

### Accepted and published

Manuscript 1: **Rentschler, T.**, Gries, P., Behrens, T., Bruelheide, H., Kühn, P., Seitz, S., Shi, X., Trogisch, S., Scholten, T., Schmidt, K. (2019): Comparison of catchment scale 3D and 2.5D modelling of soil organic carbon stocks in Jiangxi Province, PR China. PLoS ONE 14(8): e0220881. doi: 10.1371/journal.pone.0220881

Manuscript 2: **Rentschler, T.**, Werban, U., Ahner, M., Behrens, T., Gries, P., Scholten, T., Teuber, S., Schmidt, K. (2020): 3D mapping of soil organic carbon content and soil moisture with multiple geophysical sensors and machine learning. Vadose Zone Journal 19, 1. doi: 10.1002/vzj2.20062

Manuscript 3: Taghizadeh-Mehrjardi, R., Schmidt, K., Amirian-Chakan, A., **Rentschler, T.**, Zeraatpisheh, M., Sarmadian, F., Valavi, R., Davatgar, N., Behrens, T., Scholten, T. (2020): Improving the Spatial Prediction of Soil Organic Carbon Content in Two Contrasting Climatic Regions by Stacking Machine Learning Models and Rescanning Covariate Space. Remote Sensing 12 1095. doi: 10.3390/rs12071095

Manuscript 5: Stumpf, F., Schneider, M.K., Keller, A., Mayr, A., **Rentschler, T.**, Meuli, R.G., Schaepman, M., Liebisch, F., 2020. Spatial monitoring of grassland management using multi-temporal satellite imagery. Ecological Indicators 113, 106201. doi: 10.1016/j.ecolind.2020.106201

### In preparation

Manuscript 4: **Rentschler, T.**, Behrens, T., Scholten, T., Teuber, S., Schmidt, K.: 3D soil quality mapping and interpretability of multi-scale contextual spatial modelling in Andalusia, Spain. (in preparation for Scientific Reports)

## Abstract

During the Anthropocene and especially in the past decades earth's environment has undergone major changes. The planetary boundaries are increasingly under pressure. Since soil affects climate as compartment of the carbon and nitrogen cycles, it is an important resource in approaching these environmental problems. Consequently, knowledge about soil, soil processes and soil functions plays an essential role in research on and solutions for these severe environmental and socio-economic challenges. The mapping and modelling of soil provides spatial knowledge of soil status and changes over time, which allows to assess and evaluate soil management practices and attempts to solve to environmental problems. Machine learning methods have proven to be suitable for spatial mapping and modelling of soil, but often are black boxes and the model decisions and prediction results remain unexplained. However, explainable soil models based on machine learning would facilitate detection of environmental changes, contribute to decision making for environmental protection and foster acceptance in science, politics, and society. Therefore, latest efforts in machine learning were to expand the conventional machine learning framework to explainable machine learning to 1) justify decisions, 2) control, and 3) improve models and 4) to discover new knowledge. The core elements for explainable machine learning are transparency, interpretability and explainability. Additionally, domain knowledge and scientific consistency are crucial. However, to date the concepts of explainable machine learning played a marginal role in soil modelling and mapping. Objective of this thesis was to explore and describe how transparency, interpretability and explainability can be achieved in the soil mapping framework. The example studies showed how scientific consistency can be evaluated with model comparison and domain knowledge was and incorporated in DSM models. The studies showed how transparency can be accomplished with reproducible sample and covariate selection, and how interpretation of the models can be linked with domain knowledge about soil formation and processes to explain the model results.

## Zusammenfassung

Während des Anthropozäns und insbesondere in den letzten Jahrzehnten hat sich die Umwelt der Erde stark verändert. Die planetarischen Grenzen stehen zunehmend unter Druck. Da der Boden als wichtiger Teil der Kohlenstoff- und Stickstoffkreisläufe das Klima beeinflusst, ist er eine wichtige Ressource bei der Bewältigung dieser Umweltprobleme. Folglich spielt das Wissen über den Boden, Bodenprozesse und Bodenfunktionen eine wesentliche Rolle bei der Erforschung und Lösung dieser schwerwiegenden ökologischen und sozioökonomischen Herausforderungen. Die Kartierung und Modellierung des Bodens liefert räumliche Kenntnis über den Zustand des Bodens und seine Veränderungen im Laufe der Zeit. Dies ermöglicht es, Methoden der Bodenbewirtschaftung und Lösungsansätze für Umweltprobleme zu beurteilen und zu bewerten. Methoden des maschinellen Lernens haben sich für die räumliche Kartierung und Modellierung des Bodens als geeignet erwiesen. Oft handelt es sich dabei aber um Black Boxes und die Modellentscheidungen und -ergebnisse werden nicht erklärt. Allerdings würden erklärbare Bodenmodelle auf der Grundlage des maschinellen Lernens die Erkennung von Umweltveränderungen erleichtern, zur Entscheidungsfindung für den Umweltschutz beitragen und die Akzeptanz von Wissenschaft, Politik in Gesellschaft fördern. Daher sind die jüngsten Bemühungen im Bereich des maschinellen Lernens darauf ausgerichtet, den konventionellen Rahmen des maschinellen Lernens auf erklärbares maschinelles Lernen zu erweitern, um 1) Entscheidungen zu begründen, 2) die Modelle besser zu steuern und 3) zu verbessern und 4) neues Wissen zu generieren. Die Kernelemente für erklärbares maschinelles Lernen sind Transparenz, Interpretierbarkeit und Erklärbarkeit. Darüber hinaus sind domain knowledge und wissenschaftliche Konsistenz entscheidend. Bei der Bodenmodellierung spielten die Konzepte des erklärbaren maschinellen Lernens jedoch bisher eine geringe Rolle. Ziel dieser Arbeit war es, zu untersuchen und zu beschreiben, wie Transparenz, Interpretierbarkeit und Erklärbarkeit im Rahmen der Bodenmodellierung erreicht werden können. Die Fallbeispiele zeigten, wie Konsistenz mit Modellvergleichen bewertet werden kann und domain knowledge in die Modelle einfließt. Ebenso zeigten die Studien, wie Transparenz mit reproduzierbarer Proben- und Variablenauswahl erreicht werden kann und wie die Interpretation der Modelle mit domain knowledge verknüpft werden kann, um die Modellergebnisse besser zu erklären und in Bezug zu bodenkundlichem Wissen zu setzen sind.

# 1. Introduction

During the Anthropocene and especially in the past decades earth's environment has undergone major changes (Crutzen, 2002; Rockström et al., 2009). Rockström et al. (2009) distinguish between nine environmental systems and propose safe operating spaces for each of them. Among these environmental systems are climate change and human inference with the biochemical flow (mainly including the nitrogen cycle; Gruber and Galloway, 2008) where the boundaries are already exceeded. Climate change induced by increasing greenhouse gas emissions as well as the interference of the nitrogen cycle through increasing use of nitrogenous fertilizer in agriculture (Foley et al., 2011; Gruber and Galloway, 2008) are closely linked to soil, soil processes and soil functions (Bouma and McBratney, 2013). Soil affects climate as compartment of the carbon and nitrogen cycles and environmental problems such as eutrophication (Gruber and Galloway, 2008; Lal, 2010). Further, ongoing population and consumption growth will increase pressure on these boundaries and the need for changes in land use and food production (Foley et al., 2011; Godfray et al., 2010). Consequently, knowledge about soil, soil processes and soil functions plays an essential role in research on and solutions for these severe environmental and socio-economic challenges (Bouma, 2014; Bouma and McBratney, 2013).

## 1.1 Soil formation

Soil formation and development depends on the interplay between decomposition, conversation, and relocation of organic and inorganic material. The concurrence of soil forming factors induce and pursue soil forming processes. These principles where first identified by Dokuchaev (1883). The soil forming factors were formalised and assembled in the soil forming equation proposed by Jenny (1941):

$$S = f(cl, o, r, p, t, \cdots)$$

where S is a soil property or class at a certain location that develops as a function f of the environmental conditions of climate (cl), organisms (o), relief (r), parent material (p), time (t), and potential other factors (···). E.g. climate influences soil formation through temperature, precipitation, and wind. Organisms contribute with bioturbation, material input and decomposition as well as protection against erosion. Relief, i.e. terrain, effects erosion, transport and reallocation with water and wind. Parent material determines the grain size distribution, the mineralogy and soil chemistry. Time represents the duration of the period in which the other factors are taking effect.

Besides this deterministic part, which represents the contextual dependency, the stochastic part, i.e. the spatial dependence which is not incorporated in the soil forming equation, needs to be considered (Behrens et al., 2019a). Combining the first law of geography (*"everything is related to everything else, but near things are more related than distant things."* Tobler, 1970) as stochastic part with the *clorpt*-model as deterministic part results in the universal model of spatial variation proposed by Matheron (1971):

$$Z(s) = Z^*(s) + \varepsilon'(s) + \varepsilon'$$

where Z(s) is a soil property or class, Z*(s) is the deterministic part, i.e. the soil forming equation (*clorpt*), ε'(s) is the stochastic part of (apparently) random variation, and ε' is random noise which a soil model cannot account for. The stochastic part of the universal model of spatial variation can be described as contextual spatial information (Behrens et al., 2010a) because the soil forming factors are relevant on different scales besides the local site characteristics in the *clorpt*-model. This concept is used to describe potential soil variation conventionally as the concept map (and the mental map) of a soil mapper comprises the landscape and it's evolution where the soil profile that is used for mapping is located (Fig. 1; Ad-hoc-AG Boden, 2005). This extends from local slopes with e.g. erosive patterns over regional geologic formations to interregional climatic relations and teleconnections.

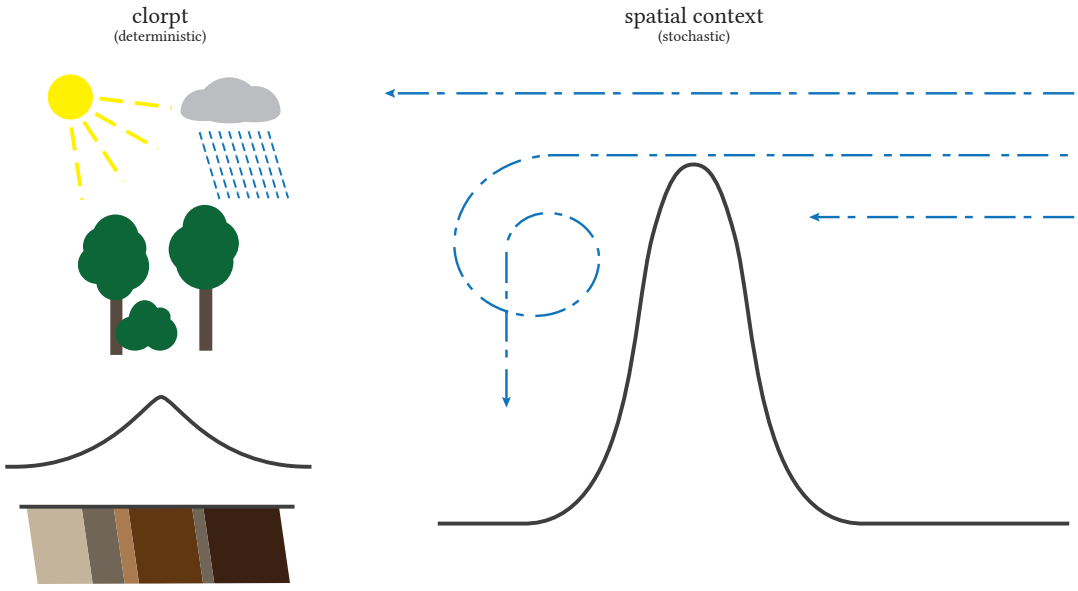


**Fig. 1.** Illustration of the deterministic and stochastic parts of the universal model of spatial variation. The *clorpt*-model reflects local processes such as weathering and autochthon material accumulation of e.g. SOC or clay minerals and catenary processes such as degradation, transport, and allochthon accumulation. The spatial context reflects large-scale processes such as climatic teleconnections and geomorphic signature.

Besides the spatial variation of soil, the vertical variation of soils, the soil profile, is determined by the soil forming factors (Jenny, 1941). This is because of various soil formation processes, such as erosion, transport, and reallocation that alter the soil thickness as well as physical and chemical soil properties with removal and burying of sediments in dependence to terrain (Milne, 1936). Further, climate and vegetation induce different vertical trends in soil organic carbon (SOC) through varying accumulation and turnover-times depending on activity of organisms temperature, and water supply (Jobbágy and Jackson, 2000). Parent material with potential stratification and varying grain sizes influence soil water percolation (Abdu et al., 2008).

### 1.2 Digital soil mapping with machine learning

The holistic *clorpt*-model and the universal model of spatial variation provide the theory for soil mapping – conventional as well as digital soil mapping (DSM). Soil mapping and spatial pedologic modelling have been important sources for soil knowledge and the spatial dimension of soil (Brevik et al., 2016). The spatial dimension of soil is crucial for land valuation and taxation, agricultural and regional decision making (Brevik et al., 2016) as well as soil assessment and monitoring (Carré et al., 2007). Therefore, spatial knowledge of soil and soil mapping can contribute to major insights in the environmental change and to solutions for environmental problems. To address environmental challenges, soil mapping is suitable to generate spatial information and data of soil, soil properties and soil functions for third-party environmental models in response to global change (Sanchez et al., 2009). Sanchez et al. (2009) identified three main steps for DSM: first the data input is assembled. The second step is model building or training by deriving quantitative relationships between the dependent target variable and the environmental covariates with the model prediction to the unsampled covariate space. In step three pedotransfer functions are used to calculate or estimate soil properties or functions that are difficult to measure (Wösten et al., 2001). Following steps comprise the development of soil management practises based on the results of the soil mapping workflow to provide the obtained information and knowledge for decision making.

Machine learning (ML) methods are able capture the deterministic and stochastic relations between soil properties and environmental properties. ML models incorporate the knowledge about soil formation processes in statistical and mathematical functions. Thus, ML methods are highly suitable to identify the relationships between soil data and environmental covariates and to predict the target variable to the unsampled covariate space. ML methods that are used for DSM are artificial neural networks (ANN; Behrens et al., 2005; McCulloch and

Pitts, 1943), random forests (RF; Breiman, 2001; Grimm et al., 2008), support vector machines (SVM; Kovačević et al., 2010; Smola and Schölkopf, 2004), Cubist (Lacoste et al., 2014; Quinlan, 1993), and deep learning (Behrens et al., 2018b; LeCun et al., 2015) amongst others.

Many DSM approaches focus on two-dimensional (2D) models. However, it is necessary to model soil as three-dimensional (3D) continuum to represent the whole solum with its vertical trends of soil properties and soil processes (Jenny, 1941; Sanchez et al., 2009). This is important to e.g. account for longer turnover times of subsoil SOC compared to topsoil SOC in respect to mitigate climate change (Meersmans et al., 2009; Trumbore, 2000) or transport processes in the vadose zone (Abdu et al., 2008; Binley et al., 2015) regarding the biochemical cycles. This can be accomplished with multi-layered 2D models. Grimm et al. (2008), Lacoste et al. (2014), Malone et al. (2009), and Mulder et al. (2016) used this approach to map SOC content and/or stocks from regional to national scale. Adhikari et al. (2013) modelled soil texture for Denmark, Gasch et al. (2015) modelled soil water content and temperature on field scale, and Hengl et al. (2014) and Viscarra Rossel et al. (2015) modelled various soil properties on continental and global scale. However, multi-layered 2D modelling has the drawback of limited vertical resolution and does not provide full 3D soil information. Soil information between the modelled layers is derived on an interpretative and subjective basis. Liu et al. (2016) denote this as pseudo-3D or 2.5D mapping, which is valuable due to computational constraints on national, continental or global scale. To model soil as 3D continuum Kempen et al. (2011) and Meersmans et al. (2009) applied soil type and land use type specific soil depth functions (Minasny et al., 2016) to estimate the vertical trends of SOC content related to categorial covariates. To enable more precise evaluations of the vertical trends on continuous spatial basis the soil depth functions can be treated as dependent soil property of the *clorpt*-model to predict SOC content and stocks (Aldana Jague et al., 2016; Liu et al., 2013; Liu et al., 2016; Minasny et al., 2006; Veronesi et al., 2014), soil compaction (Veronesi et al., 2012), and other soil properties.

### 1.3 Explainable machine learning

Many studies in DSM are target-oriented and, thus, have focused on spatial predictions with high mapping accuracy and low model error. While this is an effective approach for decision making based on spatial predictions, this approach disregards inference of new knowledge about the model itself and underlying soil processes. To extract new knowledge from DSM models it is necessary to improve model interpretability and explainability (Murdoch et al., 2019; Wadoux et al., 2020a). There are four important reasons to research on model inter-

pretability and explainability: 1) to justify decisions, 2) to control and 3) to improve models and 4) to discover new knowledge (Adadi and Berrada, 2018; Roscher et al., 2020). Further, confirmation of existing knowledge can be achieved and robustness and reliability can be evaluated better (Lipton, 2016). Roscher et al. (2020) identified three core of explainable ML:

- transparency,
- interpretability, and
- explainability.

Besides these core elements, domain knowledge and scientific consistency are relevant to achieve new scientific insights. Model transparency is given if the process to extract model parameters and make predictions can be described and motivated by the researcher. E.g. kernel-based methods such as support vector machines are transparent through the kernel functions and the convex optimization problem which is reproducible mathematically. In contrast deep neural networks are not fully transparent since the data relationships are expressed mathematically and the network architecture may be justified, but the hyper-parameters used for model training are chosen heuristically due to multiple possible local minima. Further, the plethora of hyper-parameters may be non-transparent for direct interpretation by humans. It is rather unrealistic to achieve full model transparency (Murdoch et al., 2019). According to Lipton (2016) transparent models are the opposite of black boxes.

Model interpretability is based on the presentation of the model, model properties, and model results in understandable terms to humans, i.e. "*the mapping of an abstract concept (e.g. a predicted class) into a domain that the human can make sense of*" (Montavon et al., 2018). This can be achieved through feature importance in decision tree models, saliency masks and heatmaps in computer vision, sensitivity analysis or relevance scores to interpret model decisions. Analysing large numbers of individual interpretations may be very time consuming and, therefore, possibly restricted (Murdoch et al., 2019).

Model explainability comprises the assemblage of interpretations made from a model together with further contextual information derived from domain knowledge related to the problem. Explainability cannot be achieved by the algorithm on its own but requires interpretation by experts. Yet, there is no joint understanding of this concept (Roscher et al., 2020). The term theory-guided data science (Karpatne et al., 2017) is closely related and includes the interconnection between domain knowledge and explainability.

For domain knowledge the type of knowledge, the representation and transformation of knowledge, and the integration into the ML approach are relevant. In each of the ML compartments, training data, hypothesis and training algorithm, domain knowledge can be incorporated. Domain knowledge may be given as mathematic equation (analytical expression,

differential equation, or relation between dependent and independent variable). With choosing a certain model architecture and restricting the model structure to existing real-world relations between the variables domain knowledge can be integrated into the hypothesis (Roscher et al., 2020). Domain knowledge in ML driven DSM is knowledge of the deterministic soil forming factors and soil formation as well as the stochastic landscape context that are used for conventional soil mapping. The soil forming factors of the *clorpt*-model account for the deterministic part, local environmental properties, and processes that determine soil formation. Local factors and processes comprise parent material, land cover and climate with grain size distribution, weathering, and autochthon material input and accumulation, e.g. of SOC or clay minerals. Catenary factors and processes are slope, curvature, and contributing area with soil erosion, transport, and allochthon accumulation, e.g. of colluvial deposits. Catenary processes can be depicted by terrain analysis as cause and effect are adjacent. Regional and interregional teleconnected factors and processes are climate and large-scale geomorphic signature with e.g. leeward accumulation of aeolian sediments. The contextual spatial information needs to be described with the surrounding landscape to account for the stochastic part of the universal model of spatial variation of the soil model. Additional contextual spatial information such as the teleconnections can be derived via feature engineering from other environmental covariates to cover regional and continental environmental processes (Murdoch et al., 2019; Roscher et al., 2020). Consequently, many aspects of domain knowledge regarding the *clorpt*-model were used for DSM approaches in the past (Behrens et al., 2005; Grimm et al., 2008) and DSM without soil specific domain knowledge is not scientifically consistent and appropriate (compare Wadoux et al., 2020b). However, research on contextual spatial information is rare (Behrens et al., 2010a; Behrens et al., 2018b; Behrens et al., 2019a; Behrens et al., 2019b).

By comparing the final model with existing models and domain knowledge, e.g. theories, paradigms or other models, the model can be checked for plausibility and consistency with scientific principles and other scientific studies (Reichstein et al., 2019; Roscher et al., 2020). Along with model interpretability and explainability scientific consistency with existing scientific knowledge is essential for model reliability and performance besides training accuracy (Karpatne et al., 2017).

Based on the additional elements of ML described above, Roscher et al. (2020) extended the three-step-approach of DSM (Sanchez et al., 2009) with model transparency, interpretability and explainability as well as domain knowledge and scientific consistency for "*Peeking Inside the Black Box*" (Fig. 2; Adadi and Berrada, 2018). Thus, ML in DSM is not replacing pedology

or pedologic knowledge, "*but is instead augmenting it for the twenty-first century, resulting in more of a renaissance than a revolution*" (Brunton and Kutz, 2019). This perspective highlights the use of ML as a toolbox for solving scientific questions.

### 1.4 Explainable machine learning for DSM

The extended ML framework implies for DSM that only credible covariates related to pedogenesis (*clorpt*-model) should be used. Also, model restrictions, such as unrealistic sums of grain size fractions, can improve model plausibility which may facilitate ML acceptance in soil sciences. Model interpretations are related to and limited by scientific knowledge from various disciplines (including chemistry, physics, and geosciences) in specific aspects of the model, such as distribution of silt and leeward slope positions. Further, new conclusions and hypothesis can be drawn from individual interpretations (Behrens et al., 2019a; Wadoux et al., 2020a). Behrens et al. (2019a) interpret and explain the spatial distribution of topsoil silt based on contextual spatial mapping (Behrens et al., 2010a), extracting the location of the most important feature by variable importance per soil sample, and the direction and distance of the location. This approach allows to identify locations influenced by aeolian transport of silt from the Rhine floodplain where many loess deposits in Southwest Germany have their origin. However, this is a rare example of explainable ML in DSM and research should focus on this problem to solve major problems in soil science (Behrens and Viscarra Rossel, 2020) and environmental change.

While accurate ML based soil maps are crucial for decision and policy making, researchers of DSM need to amplify the standard of producing more accurate maps by using more and more environmental data, data derivatives and the most sophisticated ML methods. ML based DSM needs to investigate why a prediction is made and how it is made to gain insights in the model structure and prediction. This can be achieved with transparency, interpretability, and explainability and helps to improve model accuracy but also helps to validate soil science theories, spatial environmental knowledge and provides solutions for urgent environmental problems. Further, transparency, interpretability and explainability may foster acceptance and trustworthiness of ML in science, politics and society (European Commission, 2020). In regard to environmental change, there is hardly need for more diagnostic studies, alarming declarations and action plans, but need for efficient framing of the environmental problems

and the implications by soil sciences and DSM (Bouma and McBratney, 2013). On the one hand, this may be achieved with conceptional frameworks to promote the perception of soil as a finite resource and link to many other environmental compartments and an important basis for human well-being (Bouma, 2014; Teuber and Schweizer, 2020). On the other hand, explainable ML in DSM, soil sciences and geosciences and corresponding visualisation techniques (mapping of feature importance and (tele-)connections, Behrens et al., 2019a; saliency masks and heatmaps, Roscher et al., 2020) can contribute to the framing via explaining and highlighting the relations between soil and environmental change.

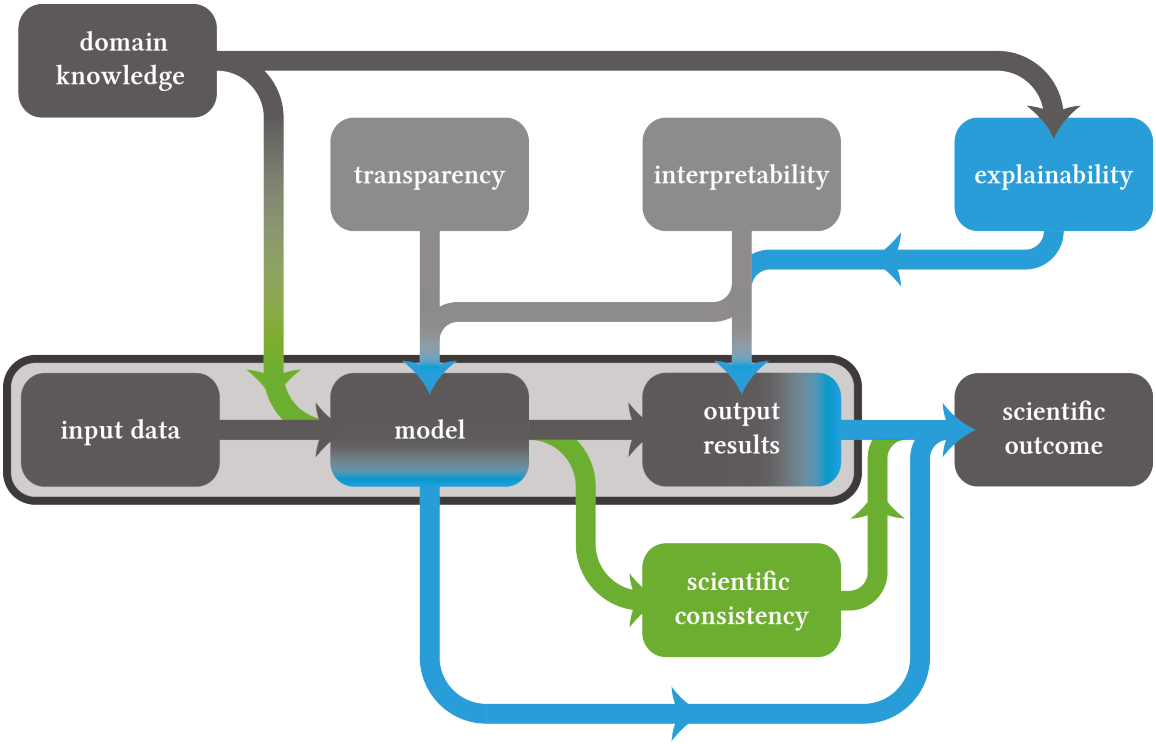


**Fig. 2.** The extended machine learning (ML) framework as described by Roscher et al. (2020) to gain new scientific insights to a certain problem. The light grey box depicts the conventional ML approach often used in DSM. Domain knowledge can generate new scientific knowledge along with transparency, interpretability and explainability of the model and model outcome. Further, domain knowledge fosters scientific consistency of the model and model results.

## 2. Objectives

As described above it is crucial to develop 3D soil models for various soil properties to provide data input for other environmental models, soil assessment and monitoring as well as decision and policy making. Most 2D and 3D DSM approaches focused on producing accurate maps and limited the domain knowledge incorporated in the model to the *clorpt*-model without explicitly discussing domain knowledge. Interpretation of the underlying soil processes and connections to environmental characteristics is an important aspect of most 3D studies but disregarded as well. According to Behrens and Viscarra Rossel (2020) model interpretability and explainability should be integrated in the DSM framework. The extended ML framework provides a starting point to achieve insights in model structure, model decisions, and potentially new scientific outcome.

Based on this, the objective of this thesis was to expand the three-step ML workflow commonly used for 3D DSM approaches with model transparency, interpretability and explainability as well as domain knowledge and scientific consistency for better understanding of soils. This was accomplished by

- incorporating domain knowledge of the *clorpt*-model in 3D soil modelling (manuscript 1, 2, 3 and 4),
- comparison between 3D and 2.5D models trained with 2D environmental covariates to evaluate the scientific consistency of 3D models (manuscript 1),
- using 3D environmental covariates as domain knowledge for 3D soil models (manuscript 2),
- using (semi-)reproducible sampling designs covering the covariate space comprehensively to increase model transparency (manuscript 2, 3 and 4),
- interpreting the variable importance of environmental covariates (manuscript 3),
- comparing 2.5D models with increasing spatial context from feature engineering (manuscript 4), and
- interpretation and explanation of the models in respect to domain knowledge (manuscript 1, 2, 3 and 4).

Manuscript 5 deals with feature engineering and is related to the temporal aspects of soil formation (t). It will be discussed briefly in the conclusion sections as outlook.

# 3. Material and methods

## 3.1 Machine learning methods

ML provides a wide toolbox for data analysis, model calibration and predictions. There are three main categories of ML: unsupervised machine learning, supervised machine learning, and reinforcement learning. While models for unsupervised learning are trained without example inputs, but with the whole dataset and the desired target variable, supervised learning models are trained with prelabelled example data. Reinforcement learning is based on dynamic datasets in which the algorithm performs a specific task and earns rewards which are tried to maximize. In soil sciences and DSM, supervised learning is used in most cases. With supervised learning a linear or non-linear function is derived from labelled training data. The data labels can be categorial such as soil type or continuous such soil texture, SOC, SM, and pH. In soil sciences the function between the target variable and related soil forming processes is usually non-linear. Examples for non-linear models used in DSM are artificial neural networks (ANN; Behrens et al., 2005; McCulloch and Pitts, 1943) and deep learning (Behrens et al., 2018b; LeCun et al., 2015), decision trees such as classification and regression trees (CART; Breiman et al., 1984) and random forests (RF; Breiman, 2001; Grimm et al., 2008), rule-based models as Cubist (Lacoste et al., 2014; Quinlan, 1993), multivariate adaptive regression splines (MARS; Friedman, 1991) and support vector machines (SVM; Kovačević et al., 2010; Smola and Schölkopf, 2004). For details on the methods used in this thesis see the respective chapter of the manuscript in the appendix.

Supervised ML models learn the holistic universal model of spatial variation (*clorpt*-model and spatial context) to make the dependent soil variables representable with environmental measurements. The increasing amount of publicly available soil data sets (e.g. LUCAS topsoil survey, Jones et al., 2020) and environmental data measured by remote sensing systems (e.g. SRTM, Farr et al., 2007; Landsat, Masek et al., 2020; Roy et al., 2014; Sentinel-2, Drusch et al., 2012) and proximal soil sensing (Viscarra Rossel et al., 2010; Viscarra Rossel et al., 2011), derived from measured data (e.g. terrain analysis of digital elevation models, [Amatulli et al., 2018, Jasiewicz and Stepinski, 2013] and vegetation indices [Bannari et al., 1995]) and modelled environmental data (e.g. climate data and models Brown et al., 2018; Fick and Hijmans, 2017; Kaufman et al., 2020) provides a comprehensive basis for DSM.

## 3.2 Domain knowledge and scientific consistency (manuscript 1)

The BEF-China Project was a biodiversity experiment on land leased and managed by the Institute of Botany of the Chinese Academy of Sciences (Bruelheide et al., 2014). The study site is a small catchment located near Xingangshan, Jiangxi Province, PR China (UTM/WGS84: 50R 588000 3222000), about 400 km south-west of Shanghai in the highly species-rich subtrop-

ics with summer monsoon, mean temperature of 17 ℃ and mean precipitation of 1800 mm. The catchment of 26.7 ha has an elevation ranging from 105 to 275 m a.s.l., typically convex slopes of 29° in average and 45° at maximum. The bedrock consists of non-calcareous slates with varying sand and silt content and grey-green sandstone. The soil types in the catchment are Endoleptic Cambisols with Anthrosols at the hillsides and Gleysols at the valley bottom and the mean soil depth is 0.6 m with underlaying saprolite (isomorphic weathered slate). The soil texture is silt loam to silty clay loam (Goebes et al., 2015; Scholten et al., 2017).

The soil data sets used in this study are part of the legacy database of BEF-China. The soil sampling design was based on the design of the biodiversity experiment with 271 experimental plots in total covering 18 ha of the catchment. The squared plots had a size of ≈665 $m^2$ (traditional Chinese unit of 1 mu) and were planted with trees in monocultures and mixtures of 2, 4, 8, 16 and 24 species. For soil sampling and other studies, 67 very intensively studied plots (VIPs) were chosen representing all diversity richness levels (Bruelheide et al., 2014). Samples for SOC (n=67) were taken with nine soil cores (3 cm in diameter) at a regular grid at each VIP and bulked for the depth increments of 0–5 cm, 5–10 cm, 10–20 cm, 20–30 cm and 30–50 cm. SOC content was determined with dry combustion CNS-analysis (Vario EL III; Elementar). SOC content ranged from 0.35 % in subsoil to 5.06 % in topsoil. Bulk density samples (n=55) were taken with soil sample rings (100 $cm^3$) in fife replicates for each depth increment. Bulk density (BD) was determined gravimetrically and ranged from 0.75 g $cm^{-3}$ in topsoil to 1.84 g $cm^{-3}$ in subsoil. The digital elevation model (DEM) of the study site was generated by ordinary kriging based on 1956 points measured with a differential GPS and had a resolution of 5 m (Krige, 1951).

We calculated slope, aspect and subsequently eastness and northness, plan, profile, longitudinal, tangential and flowline curvature, vertical distance to channel network, sky visibility, sky view factor, direct and diffusive insolation, catchment area, topographic wetness index and slope length and the steepness factor. Most terrain attributes can be derived with different methods and equations Since it is unknown which method is most suitable for modelling SOC and BD, we used several well-established methods for terrain analysis if available. Further, the most relevant scale of the derived terrain attributes is unknown due to the non-representative and fixed DEM resolution (Behrens et al., 2018a). Consequently, we applied simple smoothing with low-pass filters to incorporate spatial contextual knowledge. The used filters had circular neighbourhood sizes with radii of 1, 2, 4, 6 and 8 pixels. The maximum radius of 8 pixels represents the local catena scale of 90 m. There were 290 environmental covariates based on terrain analysis in total.

For modelling of the soil properties, we used MARS, RF and SVM. These methods are robust against multi-collinearity of the environmental covariates, that are likely to be highly correlated, since all covariates are derived from the same DEM and some terrain attributes describe the same feature calculated with various methods.

All 290 environmental covariates were used for model training according to the conventional and basic ML workflow as independent variables of the models. SOC content and BD were the dependant variables. For the multi-layered 2.5D predictions the models were trained for each sampled depth increment individually, were the central point of the depth increment was assigned as depth of the layer, which results in 15 cm vertical resolution (Hengl et al., 2014). The models were trained with exhaustive grid search (Schmidt et al., 2008) for optimal parameter selection while preserving the models ability to generalise to new data. The model performance was evaluated with 10-fold cross-validation. For the 3D modelling of SOC content and BD, we modelled soil depth functions spatially to predict the vertical trend of the soil properties in space. The vertical trend was represented by 3$^{rd}$ degree polynomial, logarithmic, and exponential functions (Aldana Jague et al., 2016; Liu et al., 2016; Veronesi et al., 2012). The dimensionless coefficients of these functions were treated as abstract soil property, the models were trained with the function coefficients as dependent variable and the coefficients were predicted in space. Subsequently, for each grid location the vertical trends of SOC content and BD were calculated from the soil depth functions with the predicted coefficients with 5 cm resolution. From this 3D array the SOC stocks were calculated with SOC content and BD.

### 3.3 Domain knowledge and model transparency (manuscript 2)

The site of this case study is an agricultural field of 58 ha at the Elbe floodplain ≈70 km north of Leipzig, Saxony, Germany. The field is bordered by the creeks Altes Fliet and Fließgraben and has no visible terrain variation. The present soil types are Gleysols and Gleyic Cambisols developing from alluvial loam over Holocene sediments of fluvial sands (LAGB, 2014).

The geophysical measurements were recorded with the electromagnetic induction (EMI) sensors CMD-Explorer and CMD-Mini-Explorer (GF Instruments). EMI sensors measure the apparent electrical conductivity ($EC_a$ in mS $m^{-1}$). The penetration depth of the signal is controlled by the sensor settings (intercoil spacing, orientation of the magnetic dipoles, and signal frequency). The penetration depth of the signal increases with intercoil spacing. Further, the signal of vertically oriented magnetic dipoles (VDP) has double penetration depths of horizontally oriented dipoles (HDP; Martini et al., 2017; McNeill, 1980a, 1980b; von Hebel et al., 2019). Both CMD-Explorer and CMD-Mini-Explorer had one transmitter and three receiver coils with intercoil spacings of 1.48 m, 2.82 m, and 4.49 m and 0.32 m, 0.71 m, and 1.18 m, respectively. Thus, the CMD-Explorer had effective penetration depths of 2.2 m, 4.2 m, and

6.7 m and the CMD-Mini-Explorer had 0.5 m, 1 m, and 1.8 m for VDP and half of the penetration depths of VDP for HDP. Together, both sensors enabled multi-depth exploration of $EC_a$ with one survey for each dipole orientation. This sensor setup enabled to detect spatial variations of $EC_a$ in the subsurface sediments of the floodplain with high spatial resolution. $EC_a$ is an important proxy for soil properties such as soil texture, horizonation, bulk density, SOC and soil moisture (SM; Doolittle and Brevik, 2014). Additionally, gamma-ray spectroscopy with measurement of $^{40}K$, $^{238}U$ and $^{232}Th$ content and the derivation of dose rate (Gy $h^{-1}$; IAEA, 2003) was applied for description of spatial variation of soil texture and SOC in high resolution of the most upper 30 cm of soil. The point measurements of the geophysical sensors were interpolated with ordinary kriging (Krige, 1951) to a cell size of 5 m to obtain spatial information about the variation of $EC_a$ in the projected depth as well as $^{40}K$, $^{238}U$, $^{232}Th$ content and dose rate.

Spatial modelling of soil requires specific sampling schemes for model training and validation (Brus et al., 2011; Schmidt et al., 2014). Besides for model training the environmental covariates also served as basis for the sampling design to cover the full range of environmental variability and potential drivers of soil formation while reducing sampling and analysis costs. In this case study, we calculated the locations of 25 soil profiles with weighted extreme conditioned Latin Hypercube Sampling (wecLHS; Schmidt et al., 2014). wecLHS is a stratified random sampling design that covers the full range of existing environmental variation and weights the data input with the model performance ($R^2$) of the kriging models to account for noise in the spatial prediction. Additionally, 10 randomly localised profiles were taken for independent validation which requires a sampling scheme independent from the training data sampling scheme (Schmidt et al., 2014; Steyerberg and Harrell, 2016).

At each soil profile four samples of 15 cm intervals from 0-60 cm were taken. In the upper 60 cm of soil more than 80 % of many crop roots can be found (Fan et al., 2016). In total 100 samples were taken for model training and 40 samples were taken for independent validation. SOC content was measured with an ELTRA CHS-580A Helios (ELTRA) and calculated as difference of total carbon and inorganic carbon. SM was determined gravimetrically after drying at 90 ℃ for 24 h.

The soil profile data was used to calculate 2$^{nd}$ degree polynomial and exponential soil depth functions (Aldana Jague et al., 2016; Rentschler et al., 2019). The function coefficients were modelled spatially with the environmental covariates as independent variables with RF and Cubist. Subsequently, for each grid location the vertical trends of SOC and SM were calculated from the soil depth functions with the predicted function coefficients with 5 cm resolution.

The 3D model predictions were validated independently with the data of the 10 randomly selected soil profiles with Person's $R^2$, Lin's concordance correlation coefficient (CCC; Lin, 1989), and the RMSE.

### 3.4 Model interpretability (manuscript 3)

The study sites were in Yazd Province, Central Iran, and Gilan Province, Northern Iran. Yazd in Central Iran has an arid climate with precipitation of 75 mm $a^{-1}$ and Gilan at the Caspian Sea a sub-humid climate with 1200 mm $a^{-1}$. The study sites were 720 $km^2$ and 3000 $km^2$ in size. Dominant soil types are Solonchaks, Gypsisols, and Regosols in Yazd and Cambisols and Chernozems in Gilan (Taghizadeh-Mehrjardi et al., 2014). The soil sampling locations were selected with conditioned Latin Hypercube Sampling (Minasny and McBratney, 2006) to provide a representative sample distribution covering the covariate space. For the site in Yazd 154 soil profiles and for Gilan 99 soil profiles were sampled. At each profile the genetic horizons were sampled to a depth of 2 m and analysed for SOC content with wet oxidation (Nelson and Sommers, 1983). We used equal-area splines to harmonize the soil data from different soil depths and estimated the SOC content for six consistent depth intervals of 0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm (Arrouays et al., 2014; Bishop et al., 1999).

The environmental covariates comprised elevation, wetness index, catchment area, catchment slope, multi-resolution valley bottom flatness index, valley depth, plane curvature, profile curvature, general curvature, and total insolation derived from the DEM of the SRTM with 30 m resolution (Farr et al., 2007). Further, we used six spectral bands of Landsat 8 (B2, B3, B4, B5, B6, and B7; Roy et al., 2014), ten spectral bands of Sentinel-2 (B2, B3, B4, B5, B6, B7, B8, B8a, B11, and B12; Drusch et al., 2012) and calculated the normalised difference vegetation index (NDVI) for both remote sensing platforms.

The main focus of this publication in respect to this thesis was the investigation of covariate selection and variable importance. RF (Breiman, 2001) with its random permutation was combined with the Boruta algorithm (Kursa and Rudnicki, 2010) to rank the environmental covariates by importance for the model. This procedure is summarised as follows:

i) The covariate space is extended by adding randomly permuted existing covariates (pC) to remove their correlation with SOC content,

ii) SOC content is predicted at six standard depths with RF using the extended covariate space (i.e., covariates and permuted covariates),

iii) The Z-score, which is an indicator of the importance of all covariates, is computed,

iv) The maximum Z-score (MZSA) among the pC's is defined,

v) A hit is assigned to all covariates that scored better than MZSA,

vi) A two-test of equality is performed for undetermined important covariates,

vii) The original covariates are respectively flagged as “unimportant” or “important” if they have significant lower or higher scores than MZSA,

viii) All permuted covariates are removed.

The Z-score is the indicator for the ability of an environmental covariate to explain the variability of SOC content. The Z-score was grouped in four classes from weakly relevant ($Z<5$), slightly relevant ($5\leq Z<10$), moderately relevant ($10\leq Z<15$) to relevant ($Z>15$).

### 3.5 Model interpretability and explainability (manuscript 4)

The study area with 1.000 km² was located 50 km North-East of Seville in the middle Guadalquivir basin, Andalusia, Spain. The area is divided in three subareas with the Sierra Morena mountain range with Paleozoic granite, gneiss, and slate in the North, the Guadalquivir flood plain with Pleistocene marl, calcarenite, and coarse sand and Holocene sands and loams in the centre running from ENE to WSW, and tertiary terraces in the South (Civis et al., 2004; Wolf and Faust, 2015). The most frequent soil types are Alfisols, Entisols, Inceptisols, and Vertisols (Gómez-Miguel, 2005). The study area is a heterogenous intensely used agricultural landscape of cropland, orange and olive plantations, pastures and the Dehesa, an agrosylvopastoral system.

Soil samples were taken according to a stratified random sampling design based on four geomorphic positions (geomorphons: Iwahashi and Pike, 2007; Jasiewicz and Stepinski, 2013; namely footslope, slope, shoulder, and flat areas) and the predominant land-cover classes “arable land”, “permanent crops”, “pastures”, “forest”, and “shrub and/or herbaceous vegetation associations” of the CORINE Land Cover level 2 classes (Büttner and Kosztra, 2010). The point density was proportional to the area of the strata with a minimum of three points for the smallest stratum. In total 130 soil profiles with five depth increments of 0-10, 10-20, 20-30, 40-60, and 70-100 cm (or less for shallow soils) were sampled with three replicates. The 506 samples in total were analysed with NIR and MIR spectroscopy (833–2500 nm, Tensor II, Bruker Optics; 2500–25.000 nm, GladiATR, Pike Technlogies). A subset of 97 samples was analysed for SOC and total nitrogen content (Nt; Vario EL III), pH in KCl solution ($pH_{KCl}$), and the grain size fractions clay (<2 µm), silt (2-50 µm), and sand (50-2000 µm) with a SediGraph III (Micromeritics). This subset was used to train a partial least squares regression (PLSR) model with the NMIR spectra to predict the measured soil properties for the samples not analysed with wet chemistry. PLSR is a common predictive model in chemometrics (Viscarra Rossel and Behrens, 2010; Wold et al., 2001). The RMSE of the PLSR models was 0.5 % SOC, 0.02 % Nt, 0.4 $pH_{KCl}$, 4 % clay, 5 % silt and 5 % sand content. Subsequently, we calculated water content at field capacity ($\theta_{FC}$ in $cm^3\ cm^{-3}$) and cation exchange capacity (CEC in cmol $kg^{-1}$) with pe-

do-transfer functions according to Tóth et al. (2015) and Khaledian et al. (2017) as suggested for soil properties that are difficult to measure (Sanchez et al., 2009). With $pH_{KCl}$, $\theta_{FC}$, and CEC a simple weighted soil quality index was calculated (Pulido et al., 2017).

A DEM of 5 m resolution was the basis for terrain analysis (CNIG, 2018). For contextual spatial modelling, gaussian mixed scaling was used to derive the environmental covariates (Behrens et al., 2019b). The Gaussian pyramid is a multi-scale processing method to decompose the scales of environmental covariates (Behrens et al., 2018a) with down-scaling (removing every second row and column of the DEM) and Gaussian filtering to remove artifacts of the down-scaling process. The outputs of the repeated procedure are called octave, which are up-sampled after deriving the terrain attributes to the original resolution with interpolation so that all environmental covariates are given on the same resolution. The terrain attributes in this study were slope, sine and cosine transformed aspect, mean, profile and planform curvature, flow accumulation, and topographic wetness index.

Based on Behrens et al. (2019b) we used additive and subtractive multi-scale RF models based on mixed scaling of the DEM. Model performance was evaluated with 10 times repeated 10-fold cross-validation. With additive and subtractive multi-scale modelling series of models are trained by sequentially adding or removing sets of covariates of the same octave. Additive multi-scale modelling starts with the original scale and sequentially adds the coarser scales. Subtractive multi-scale modelling starts with the full set of covariates and removes the respective coarsest scale. With this approach it was possible to identify the relevant range of scales and if there was an upper and a lower limit of spatial range, which helped to interpret the spatial context of a landscape and its relation to soil (Behrens and Viscarra Rossel, 2020). Sequential model training is computational exhausting. Final analysis for this study was completed at the time of submission of this thesis, but visualisation and discussion of the results were pending.

# 4. Results and discussion

## 4.1 Domain knowledge and scientific consistency (manuscript 1)

The cross-validation results of the 2.5D predictions showed mean $R^2$ of 0.34 for MARS, 0.41 for RF and 0.39 for SVM for SOC over all depth increments. The mean $R^2$ for BD was 0.43, 0.39 and 0.39 for BD. This means that the environmental covariates used for model training explain the spatial distribution of SOC and BD in the investigated depth increments of 0–5 cm, 5–10 cm, 10–20 cm, 20–30 cm and 30–50 cm in the study area to up to 40 %. This was because the terrain attributes derived on different scales represent the terrain characteristics that induce and control erosion, sediment transport, and reallocation. Consequently, the domain knowledge of erosion processes and their influence on soil profile development presented by Milne (1936) was incorporated in the model via representative environmental covariates. The unexplained variation in the model may be attributed to the local processes of SOC input through vegetation and weathering of the underlying slates of the bedrock. Some studies suggested that multi-scale terrain covariates may be a proxy for other environmental covariates such as microclimate, solar insulation, and subsequently primary production of vegetation as well as parent material through strike and dip of geologic layers (Behrens et al., 2010b; Behrens et al., 2018a). However, this may be not the case in a small catchment with little variation in vegetation of the biodiversity experiment with semi-random tree species composition and the previous fir plantation. To account for small-scale spatial variation in vegetation optical remote sensing systems and vegetation indices may be more suitable. For variation in parent material, proximal sensing systems as applied in hydrogeophysics may be better (Binley et al., 2015). Among many other studies (e.g. Aldana Jague et al., 2016; Grimm et al., 2008; Veronesi et al., 2014), this example shows that pedological domain knowledge often was used for covariate selection and that the model output is assessed and discussed in respect to soil properties, processes and pedological knowledge. However, few studies addressed pedological domain knowledge in respect to explainable ML (e.g. Behrens and Viscarra Rossel, 2020).

Besides the consideration of pedologic domain knowledge for a ML model, the scientific consistency is important for explainable ML and new scientific outcomes. Scientific consistency also requires the incorporation of domain knowledge. New model approaches, such as the spatial modelling of soil depth functions, require comparison to well-established ML

approaches adopted by DSM researchers to evaluate the scientific consistency of the new methods. Further, different ML methods should be tested and compared to provide scientific consistent results. The direct comparison of the 3D prediction to the well-established 2.5D predictions showed very weak correlation of MARS models for SOC and BD with all tested depth functions in all depth increments ($R^2 \leq 0.06$; Fig. 3). The 3D predictions of MARS were not consistent with the 2.5D predictions. For the polynomial depth function RF and SVM models had high correlation for the upper two and three depth increments (0-5 and 5-10 cm) with $R^2 \approx 0.7$ to 0.9 and low to intermediate correlation with $R^2 \approx 0.1$ to 0.4 for the lower three and two depth increments for RF and SVM, respectively. For the exponential depth function for SOC content $R^2$ ranged from 0.84 to 0.96 for RF and from 0.67 to 0.93 for SVM. The $R^2$ values of models with exponential function were distributed more homogeneous over the depth increments than of the models with the polynomial function. The correlation of the models with logarithmic function for BD was similar to the models with exponential depth functions for SOC content. Therefore, the consistency of 3D models based on multi-scale terrain covariates and spatially predicted soil depth functions in this study was high for depth functions modelled with exponential and logarithmic functions. The consistency of spatially predicted polynomial soil depth functions with the well-established 2.5D predictions is high for topsoil (0-20 cm) and low for subsoil (>20 cm). Topsoil SOC content was modelled with its relation to the terrain covariates. However, subsoil SOC content is less related to erosion, transport, and accumulation processes controlled by terrain, but by bioturbation and vertical transport in the liquid phase or others. The high flexibility of the polynomial depth functions was able to model these vertical trends, which could not be reproduced in spatial modelling in subsoil with high correlation like with exponential and logarithmic depth functions. For 3D soil mapping with polynomial depth functions terrain covariates may not be sufficient. 3D models may also require covariates representing the vertical characteristics of soil to be consistent with the more common 2.5D predictions.

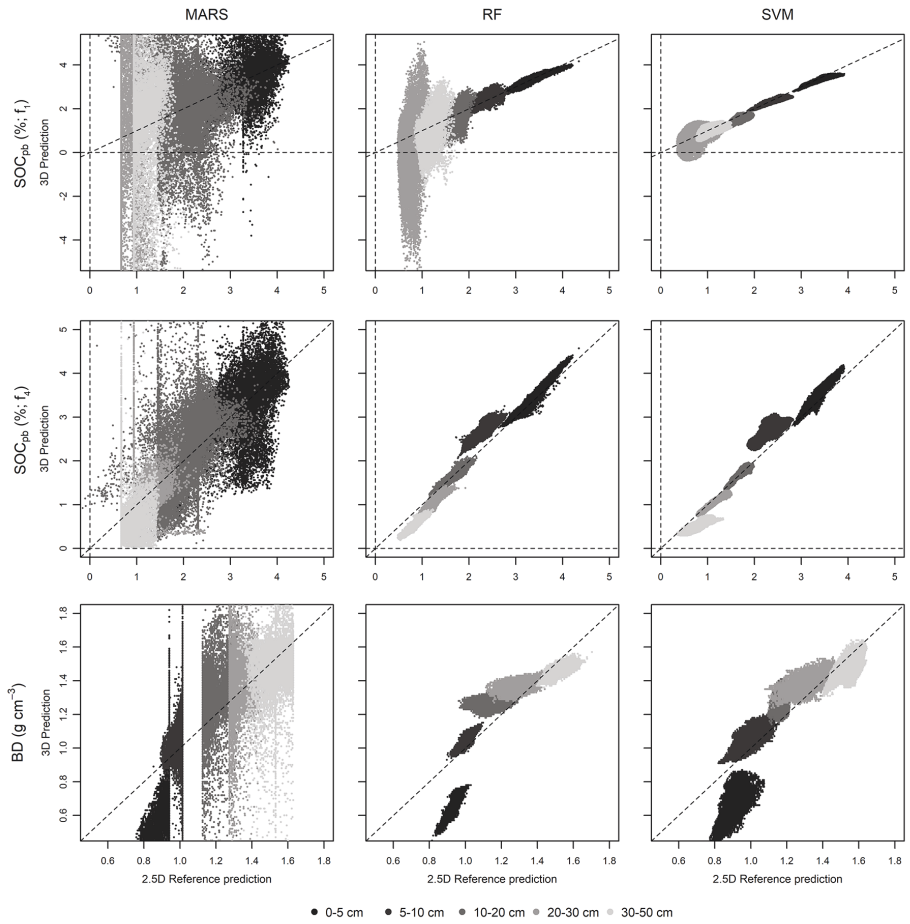


**Fig. 3.** Consistency evaluated as correlation ($R^2$) of the 3D model predictions (y-axis) with the well-established 2.5D model predictions (x-axis).

### 4.2 Domain knowledge and model transparency (manuscript 2)

The results of the independent validation of the 3D mapping approach with polynomial soil depth functions with multi-sensor hydrogeophysical covariates showed high correlation for both SOC and SM content. Both, Cubist and RF showed comparable validation results in $R^2$ and RMSE. The $R^2$ for Cubist models of SOC and SM ranged from 0.86 to 0.88 and the $R^2$ of the RF models ranged from 0.84 to 0.89. The RMSE was about 0.55-0.6 % SOC content and 2 % SM for both Cubist and RF. The high model consistency with the independent measurements showed that the hydrogeophysical covariates represent the soil and subsoil sediment characteristics adequately and enabled models with low error. The validation results of the model for each sampled depth increment were similar. The normalised RMSE which is necessary because of the different ranges of SOC content in the depth increments showed that the more flexible polynomial depth function was modelled spatially more accurate than the exponential function. The model error of Cubist and RF for the most upper depth increment (0-15 cm) was similar (SOC: 0.15-0.17; SM: 0.14-0.16), but the nRMSE for the depth increment

(15-30 cm) with the exponential function was higher (>0.3) as with polynomial function (0.23 and 0.25) for SOC content. This pattern is similar for the third increment with similar error and fourth depth increment with lower error of the polynomial function. This was because the more flexible polynomial depth function was able to model the vertical trend of SOC content with lower error. This may be related to the plough horizon, that causes an abrupt decrease in SOC content in the second depth increment. For SM there was hardly any difference between the polynomial and exponential functions and between Cubist and RF. The low model error was due to the high correlation of EMI with soil moisture, soil texture, salinity, pH amongst others, and gamma-ray spectroscopy with soil texture and mineralogy. Subsequently, corresponding soil processes such as weathering, mineralogy, and percolating water could be modelled spatially with multi-depth $EC_a$ and $^{40}K$, $^{238}U$, $^{232}Th$ content, and dose rate. Thus, incorporating multi-sensor hydrogeophysical covariates makes use of domain knowledge and underlying processes which enabled high-resolution 3D models with low error. The final prediction is shown in Fig. 4.

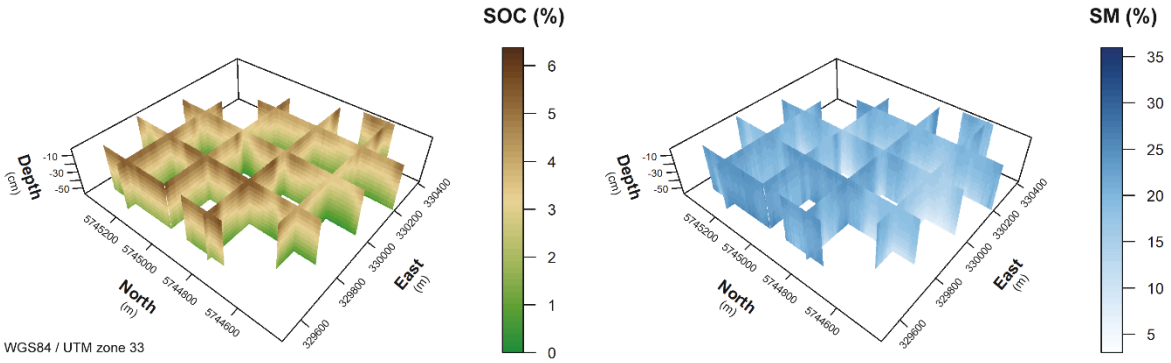


**Fig. 4.** 3D predictions of SOC content and SM.

Further, this study showed that hydrogeophysical covariates are highly suitable to prepare a comprehensive and representative sampling design. As suggested by Schmidt et al. (2014) weighted extreme conditioned Latin Hypercube Sampling is suitable to achieve a representative sampling design that covers the covariate space with all relevant characteristics comprehensively and sparsely. In respect to the extended ML framework (Roscher et al., 2020) this sampling design is based on domain knowledge and the model input of environmental and soil data. Consequently, an adaption of the extended ML framework by linking domain knowledge with the input data is appropriate for DSM (Fig. 5). Sampling design methods such as the stratification of LHS and expert-based stratification in respect to the soil forming factors are (semi-)reproducible whereas randomly or grid-based located soil profiles and from legacy datasets randomly selected subsets for model training and validation are not.

Reproducible sampling designs justify why a specific sample is relevant for the model and, thus, reproducible calculated sampling locations contribute to transparency of the ML framework (Fig. 5).

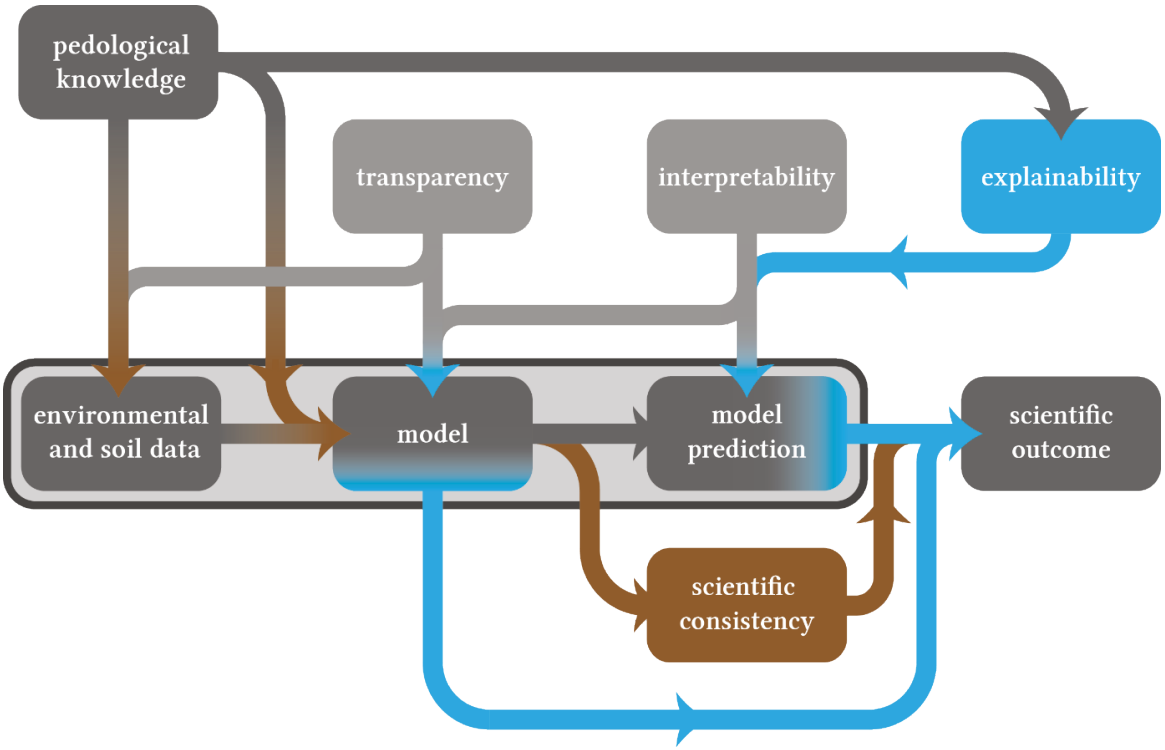


**Fig. 5.** Adaption of the extended ML workflow to account for model transparency introduced with reproducible sampling schemes and domain knowledge-based covariate selection.

### 4.3 Model interpretability (manuscript 3)

The environmental covariates of the two study sites showed a varying variable importance (Z-score; Fig. 6). At the arid site, the covariates were weakly to moderately relevant (0.4<Z<10.6). The terrain covariates were weakly (catchment area, valley depth, curvature, total insulation) to moderately (elevation, wetness index, catchment slope, valley bottom flatness index) relevant. There was hardly any difference in variable importance across the depth increments. Only general curvature and total insulation showed that they were less relevant in topsoil and tended to be more relevant in subsoil. Conversely, the bands of Landsat 8 and Sentinel-2 as well as the NDVI were slightly relevant in topsoil (<30 cm) and weakly relevant in subsoil (60-200 cm). There was a transition zone from 30-60 cm were the bands of the red edge and short-wave infrared of Sentinel-2 (B7, B8, B8a and B12; 783 nm-865 nm and around 2190 nm) and vegetation indices were slightly and moderately relevant and most Landsat 8 bands and Sentinel-2 bands 2-6 are weakly relevant (Fig. 6). At the sub-humid site most terrain covariates were weakly relevant. Valley depth and total insulation were weakly to

slightly relevant in topsoil (<30 cm) and weakly relevant in subsoil (>30 cm). Catchment area showed an increasing trend from weakly relevant in topsoil (<30 cm) over slightly relevant (30-100 cm) to moderately relevant (100-200 cm). The bands of Landsat 8 and Sentinal-2 were slightly relevant to relevant in topsoil (>30 cm) except for B5. Optical remote sensing covariates were more relevant for topsoil than for of models of subsoil SOC.

The calculation of variable importance enabled the interpretation of the contribution of each covariate to the model and increases the interpretability of a model. The model transparency is increased since the interpretation is based on objective and reproducible values. The distinction of optical remote sensing covariates between topsoil and subsoil was because they are proxies for SOC input to the soil as they reflect primary production. Especially red edge and near infrared bands as well as vegetation indices were highly suitable as vegetation proxy. Therefore, the spatial distribution of SOC at the sub-humid site was well characterised by vegetation proxies. Terrain covariates that reflect erosion processes were less relevant since ground covering vegetation prevents soil erosion mostly. Low erosion was not reflected by most of the terrain covariates except for catchment area in buried subsoil. In topsoil this effect was less visible due to the higher relevance of vegetation. In the arid study site erosion processes and reallocation of SOC are reflected by elevation, wetness index, catchment slope, and valley bottom flatness index. These covariates represented important geomorphological units that control erosion at unvegetated slopes and deposition at footslopes and the valley bottom. Further, the soil colour of unvegetated areas was slightly relevant for SOC content. Red edge and B12 (short-wave infrared) bands and NDVI of Sentinel-2 were slightly to moderately relevant to distinguish between cultivated areas in the centre of the study area and unvegetated bare soil.

This example showed, that domain knowledge and its interpretation related to organisms of the *clorpt*-model is important to distinguish between climatic regions with different vegetation patterns. Further, the relative effect of the covariates on topsoil and subsoil SOC content varied with depth and allows conclusions on the different soil processes involved in SOC cycling in the whole soil profile.

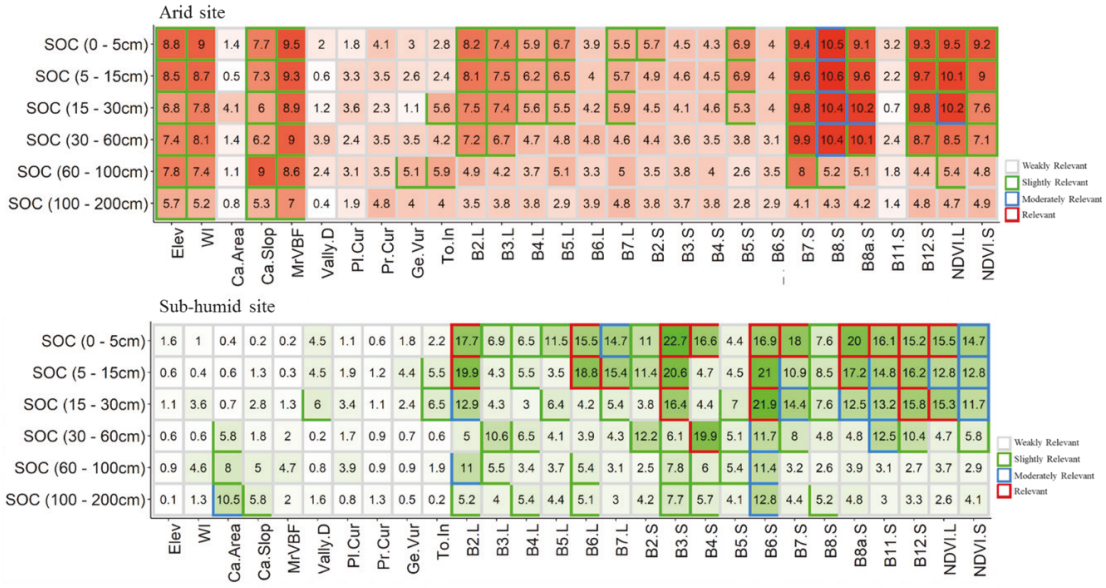


**Fig. 6.** Z-Scores for each environmental covariate calculated with the Boruta algorithm (figure by R. Taghizadeh-Mehrjardi).

### 4.4 Model interpretability and explainability (manuscript 4)

The results of the repeated cross-validation with sequentially adding of larger scales of the terrain covariates showed a general increase $R^2$ and a decrease of RMSE for all depth increments (Fig. 7). $R^2$ was double with all scales incorporated compared to the original resolution. The RMSE decreased by roughly 0.1 of the dimensionless soil quality index (SQI). However, the increase of $R^2$ and the decrease of RMSE showed different patterns for the depth increments. The $R^2$ of the models for the first depth increment was similar with the first 5 octaves (scale 5-160 m) and increased from around 0.08 to 0.18 from octaves 5 to 11 (scale 160-10240 m). The models of the second depth increment had an $R^2$ increasing from 0.8 to 0.15 at octave 4 (80 m) and increasing from 0.15 to 0.26 from octaves 7-11 (scale 240-10240 m). The $R^2$ of the models for the third depth increment were around 0.15 for the octaves 1-9 (scale 5-1280 m) and increased with subsequent octaves (scale 1920-10240 m) to 0.27. $R^2$ of the models for the fourth depth increment increased constantly from 0.13 to 0.28. $R^2$ of the models for the fifth depth increment increased from 0.13 to 0.28 (octave1-3, scale 5-40 m), were similar but very noisy with adding octaves 5-8 (scale 80-1280 m) and increased with octaves 9-11 (scale 2560-10240 m). The RMSE for all depth increments decreased with similar patterns but noisier, which makes interpretation difficult.

The constant $R^2$ values for topsoil (0-30 cm) for small scales showed that the scales 10-160 m and 10-1280 m were weakly relevant for the models, since model performance did not improve. This may be because of the agricultural use of most of the soils and homogenisation

of the soil quality through decades or even centuries of cultivation through ploughing, fertilization, and soil erosion. Larger scales (>160 and >1280 m) may be relevant to distinguish between the three main areas (Sierra Morena mountain range, Guadalquivir flood plain, and tertiary terraces), with the individual dominant type of land cover (Dehesa, forest/shrub; olive plantations; orange plantations and cotton; cereals). The individual land cover and use may influence CEC as part of the soil quality index used in this study. Further, the parent material of the three regions is shale and gneiss in the Sierra Morena and fluvial sediments at the floodplain and tertiary terraces, which may influence soil texture and, thus, water holding capacity ($\theta_{FC}$) and $pH_{KCl}$ as parts of the soil quality index. The models of subsoil (40-100 cm) showed a more constant increase in model performance. This contrast may be related to the small impact of agricultural practises that homogenise soil quality through ploughing, fertilization, and soil erosion. This differences in model performance with the different relevance of the spatial context of the landscape may refer to the decoupling of natural soil processes and human-influenced development of topsoil.

Thus, the domain knowledge about the landscape (local, catenary, and large-scale spatial context), environmental data, and soil management practices was crucial to interpret the model performance and explain effects of the soil forming factors. Explaining the effects of the soil forming factors on current soil status may help to maintain soil quality and soil functions.

(This study was work in progress at the time of printing of this thesis.)

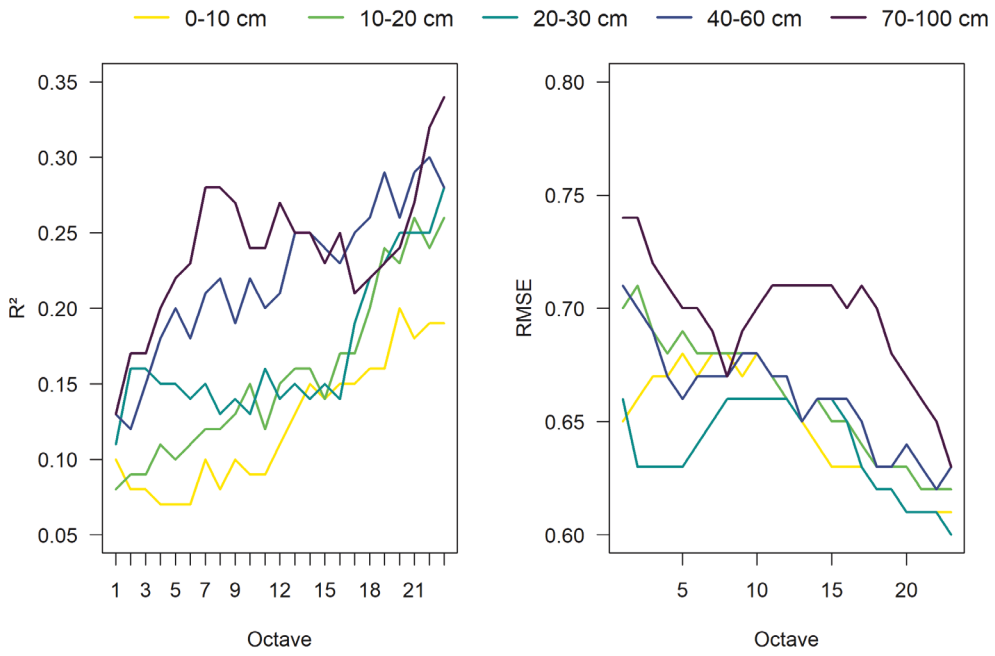


**Fig. 7.** Results of the repeated 10-fold cross-validation of the additive model training with sequentially adding scales (octaves in this figure are half octaves).

## 5. Conclusions

The objectives of this thesis were to develop 3D soil models and expand the three-step ML framework based on these models. All models of the individual studies had in common that they used domain knowledge of the *clorpt*-model. With manuscript 1 it was shown that multi-scale terrain data derived from a DEM can represent soil erosion, transport, and reallocation processes adequately for topsoil SOC. Subsoil SOC, however, was modelled less accurate and, therefore, other processes of SOC input and redistribution, such as vegetation and vertical SOC transport with bioturbation and the liquid phase, should be considered. The latter was shown with manuscript 2. For that, an agricultural field without visible terrain variation was chosen to investigate the relations of multi-sensor EMI and gamma-ray spectroscopy to SOC and soil moisture. The 3D models had constant low error throughout the investigated soil depth. Manuscript 3 contrasted the use of terrain covariates and optical remote sensing covariates that were used as proxy for photosynthetic activity and biomass. In the different climatic regions both soil forming factors, terrain and organisms, had different variable importance and relevance for the soil forming processes of SOC. Manuscript 4 showed the importance of the spatial context for local soil information. Due to agricultural practices topsoil soil quality was homogenised by ploughing, fertilization, and soil erosion in relation to the natural conditions of the cultivated landscape. Subsoil models showed constant increase in model performance, indicating that the spatial relation of soil quality is closer to the natural environmental characteristics. Many studies in DSM used domain knowledge regarding the *clorpt*-model, but without discussing it explicitly. This should be considered and discussed in more detailed in DSM studies in respect to explainable ML.

Model transparency and scientific consistency were discussed regarding sampling designs (manuscripts 2, 3, and 4) and comparison of new modelling approaches to well-established DSM approaches. Transparency of the approaches was increased with (semi-)reproducible sampling methods, i.e. conditioned Latin Hypercube Sampling (manuscript 3) and extensions of that (manuscript 2) or stratification of the covariate space based on domain knowledge (manuscript 4). Both methods justify why a sample is relevant for the model and why it was taken. Hence, the extended ML approach was adapted as domain knowledge and transparency are linked and should be included in ML based DSM. The scientific consistency of the 3D models based on spatial modelling of soil depth functions with 2.5D multi-layered models showed that the 3D models are partly consistent when the soil forming processes of the whole modelled solum are related environmental covariates used for the model, i.e. terrain covariates mainly effect topsoil properties, whereas hydrogeophysical covariates can represent subsoil as well.

The interpretability and explainability of the models were always related to domain knowledge. Domain knowledge was incorporated in the model through the selection of environmental covariates as well as through expert knowledge about underlying soil forming processes. The calculation of the variable importance (manuscript 3) could clearly show which covariate is relevant in which area and under which environmental conditions that were not represented by environmental covariates, i.e. climatic conditions of the contrasting regions. Consequently, knowledge about the study area that was not or could not be represented by environmental covariates is important for explanations. Manuscript 4 used environmental covariates created with feature engineering to incorporate the spatial context of the study area in the model and to evaluate the relevant range of the spatial context. For homogenised topsoil, the small and partly also intermediate scales were less relevant than the large scales, whereas for subsoil the model performance increased constantly with adding scales. Thus, one important outcome is that for topsoil and subsoil modelling different environmental covariates and proxies for environmental covariates are necessary, and, consequently, different soil forming factors and spatial contexts are relevant.

The soil forming factor time (t) was hardly considered in DSM to date as it is difficult to measure and translate into environmental covariates. However, time may potentially explain some remaining questions regarding spatial variation since measurements of actual soil properties always are a result of past soil forming processes and their spatial and vertical variation. Manuscript 5 describes, how time can be introduced to soil models with feature engineering of time series.

*We can only see a short distance ahead,*
*but we can see plenty there that needs to be done.*

Alan M. Turing, 1950

## 6. References

Abdu, H., Robinson, D.A., Seyfried, M., Jones, S.B., 2008. Geophysical imaging of watershed subsurface patterns and prediction of soil texture and water holding capacity. Water Resour. Res. 44 (4), 189.

Adadi, A., Berrada, M., 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access 6, 52138–52160.

Adhikari, K., Kheir, R.B., Greve, M.B., Bøcher, P.K., Malone, B.P., Minasny, B., McBratney, A.B., Greve, M.H., 2013. High-Resolution 3-D Mapping of Soil Texture in Denmark. Soil Science Society of America Journal 77 (3), 860–876.

Ad-hoc-AG Boden, 2005. Bodenkundliche Kartieranleitung, 5. verbesserte und erweiterte Auflage ed., Hannover.

Aldana Jague, E., Sommer, M., Saby, N.P.A., Cornelis, J.-T., van Wesemael, B., van Oost, K., 2016. High resolution characterization of the soil organic carbon depth profile in a soil landscape affected by erosion. Soil and Tillage Research 156, 185–193.

Amatulli, G., Domisch, S., Tuanmu, M.-N., Parmentier, B., Ranipeta, A., Malczyk, J., Jetz, W., 2018. A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. Scientific data 5, 180040.

Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B.M., Hong, S.Y., Lagacherie, P., Lelyk, G., McBratney, A.B., McKenzie, N.J., Mendonca-Santos, M.d.L., Minasny, B., Montanarella, L., Odeh, I.O.A., Sanchez, P.A., Thompson, J.A., Zhang, G.-L., 2014. Chapter Three - GlobalSoilMap: Toward a Fine-Resolution Global Grid of Soil Properties, in: Sparks, D.L. (Ed.), Advances in Agronomy, vol. 125. Academic Press, pp. 93–134.

Bannari, A., Morin, D., Bonn, F., Huete, A.R., 1995. A review of vegetation indices. Remote Sensing Reviews 13 (1-2), 95–120.

Behrens, T., Förster, H., Scholten, T., Steinrücken, U., Spies, E.-D., Goldschmitt, M., 2005. Digital soil mapping using artificial neural networks. Z. Pflanzenernähr. Bodenk. 168 (1), 21–33.

Behrens, T., MacMillan, R.A., Viscarra Rossel, R.A., Schmidt, K., Lee, J., 2019a. Teleconnections in spatial modelling. Geoderma 354, 113854.

Behrens, T., Schmidt, K., MacMillan, R.A., Viscarra Rossel, R.A., 2018a. Multiscale contextual spatial modelling with the Gaussian scale space. Geoderma 310, 128–137.

Behrens, T., Schmidt, K., MacMillan, R.A., Viscarra Rossel, R.A., 2018b. Multi-scale digital soil mapping with deep learning. Scientific reports 8 (1), 15244.

Behrens, T., Schmidt, K., Zhu, A.X., Scholten, T., 2010a. The ConMap approach for terrain-based digital soil mapping. Eur J Soil Sci 61 (1), 133–143.

Behrens, T., Viscarra Rossel, R.A., 2020. On the interpretability of predictors in spatial data science: the information horizon. Scientific reports 10 (1), 16737.

Behrens, T., Viscarra Rossel, R.A., Kerry, R., MacMillan, R., Schmidt, K., Lee, J., Scholten, T., Zhu, A.-X., 2019b. The relevant range of scales for multi-scale contextual spatial modelling. Scientific reports 9 (1), 14800.

Behrens, T., Zhu, A.-X., Schmidt, K., Scholten, T., 2010b. Multi-scale digital terrain analysis and feature selection for digital soil mapping. Geoderma 155 (3-4), 175–185.

Binley, A., Hubbard, S.S., Huisman, J.A., Revil, A., Robinson, D.A., Singha, K., Slater, L.D., 2015. The emergence of hydrogeophysics for improved understanding of subsurface processes over multiple scales. Water Resour. Res. 51 (6), 3837–3866.

Bishop, T.F.A., McBratney, A.B., Laslett, G.M., 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91 (1-2), 27–45.

Bouma, J., 2014. Soil science contributions towards Sustainable Development Goals and their implementation: linking soil functions with ecosystem services. Z. Pflanzenernähr. Bodenk. 177 (2), 111–120.

Bouma, J., McBratney, A., 2013. Framing soils as an actor when dealing with wicked environmental problems. Geoderma 200-201, 130–139.

Breiman, L., 2001. Random Forests. Machine Learning 45 (1), 5–32.

Breiman, L., Friedman, J.H., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees. Chapman and Hall, New York, NY.

Brevik, E.C., Calzolari, C., Miller, B.A., Pereira, P., Kabala, C., Baumgarten, A., Jordán, A., 2016. Soil mapping, classification, and pedologic modeling: History and future directions. Geoderma 264, 256–274.

Brown, J.L., Hill, D.J., Dolan, A.M., Carnaval, A.C., Haywood, A.M., 2018. PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. Scientific data 5, 180254.

Bruelheide, H., Nadrowski, K., Assmann, T., Bauhus, J., Both, S., Buscot, F., Chen, X.-Y., Ding, B., Durka, W., Erfmeier, A., Gutknecht, J.L.M., Guo, D., Guo, L.-D., Härdtle, W., He, J.-S., Klein, A.-M., Kühn, P., Liang, Y., Liu, X., Michalski, S., Niklaus, P.A., Pei, K., Scherer-Lorenzen, M., Scholten, T., Schuldt, A., Seidler, G., Trogisch, S., Oheimb, G., Welk, E., Wirth, C., Wubet, T., Yang, X., Yu, M., Zhang, S., Zhou, H., Fischer, M., Ma, K., Schmid, B., 2014. Designing forest biodiversity experiments: general considerations illustrated by a new large experiment in subtropical C hina. Methods Ecol Evol 5 (1), 74–89.

Brunton, S.L., Kutz, J.N., 2019. Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control. Cambridge University Press, Cambridge.

Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. European Journal of Soil Science 62 (3), 394–407.

Büttner, G., Kosztra, B., 2010. CLC2018Technical Guidelines. European Environment Agency. https://land.copernicus.eu/user-corner/technical-library/clc2018technicalguidelines_final.pdf.

## References


Carré, F., McBratney, A.B., Mayr, T., Montanarella, L., 2007. Digital soil assessments: Beyond DSM. Geoderma 142 (1-2), 69–79.

Civis, J., Dabrio, C.J., Gonzcilez-Delgado, J.A., Gay, J.l., Ledesma, S., Pais, J., Sierra, F.J., Zazo, C., 2004. Cuenza del guadalquvir, in: Vera, J.A. (Ed.), Geológica de España, Madrid, pp. 543–550.

CNIG, 2018. Digital Surfaces Model - DSM05. http://centrodedescargas.cnig.es/CentroDescargas/index.jsp.

Crutzen, P.J., 2002. Geology of mankind. Nature 415 (6867), 23.

Dokuchaev, V.V., 1883. The Russian Chernozem: Report to the free economic society, St. Petersburg.

Doolittle, J.A., Brevik, E.C., 2014. The use of electromagnetic induction techniques in soils studies. Geoderma 223-225, 33–45.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. Remote Sensing of Environment 120, 25–36.

European Commission, 2020. White Paper on Artificial Intelligence - a European approach to excellence and trust. European Commission. https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

Fan, J., McConkey, B., Wang, H., Janzen, H., 2016. Root distribution by depth for temperate agricultural crops. Field Crops Research 189, 68–74.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D., 2007. The Shuttle Radar Topography Mission. Rev. Geophys. 45 (2).

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol 37 (12), 4302–4315.

Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., Mueller, N.D., O'Connell, C., Ray, D.K., West, P.C., Balzer, C., Bennett, E.M., Carpenter, S.R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., Tilman, D., Zaks, D.P.M., 2011. Solutions for a cultivated planet. Nature 478 (7369), 337–342.

Friedman, J.H., 1991. Multivariate Adaptive Regression Splines. The Annals of Statistics 19 (1), 1–67.

Gasch, C.K., Hengl, T., Gräler, B., Meyer, H., Magney, T.S., Brown, D.J., 2015. Spatio-temporal interpolation of soil water, temperature, and electrical conductivity in 3D + T: The Cook Agronomy Farm data set. Spatial Statistics 14, 70–90.

Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., Toulmin, C., 2010. Food Security: The Challenge of Feeding 9 Billion People. Science 327 (5967), 812–818.

Goebes, P., Seitz, S., Kühn, P., Li, Y., Niklaus, P.A., Oheimb, G. von, Scholten, T., 2015. Throughfall kinetic energy in young subtropical forests: Investigation on tree species richness effects and spatial variability. Agricultural and Forest Meteorology 213, 148–159.

Gómez-Miguel, V., 2005. Mapa de Suelos de España. Centro Nacional de Información Geográfica (CNIG), Madrid.

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 (1-2), 102–113.

Gruber, N., Galloway, J.N., 2008. An Earth-system perspective of the global nitrogen cycle. Nature 451 (7176), 293–296.

Hengl, T., Jesus, J.M. de, 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.

IAEA, 2003. Guidelines for radioelement mapping using gamma ray spectrometry data. IAEA Technical Documents 1363, 179 pp. https://www-pub.iaea.org/mtcd/publications/pdf/te_1363_web.pdf. Accessed 9 July 2019.

Iwahashi, J., Pike, R.J., 2007. Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. Geomorphology 86 (3-4), 409–440.

Jasiewicz, J., Stepinski, T.F., 2013. Geomorphons — a pattern recognition approach to classification and mapping of landforms. Geomorphology 182, 147–156.

Jenny, H., 1941. Factors of soil formation: A System of Quantitative Pedology. McGraw-Hill, New York.

Jobbágy, E.G., Jackson, R.B., 2000. The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecological Applications 10 (2), 423–436.

Jones, A., Fernandez-Ugalde, O., Scarpa, S., 2020. LUCAS 2015 Topsoil Survey: Presentation of dataset and results. European Commission.

Karpatne, A., Atluri, G., Faghmous, J.H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., Kumar, V., 2017. Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data. IEEE Trans. Knowl. Data Eng. 29 (10), 2318–2331.

Kaufman, D., McKay, N., Routson, C., Erb, M., Davis, B., Heiri, O., Jaccard, S., Tierney, J., Dätwyler, C., Axford, Y., Brussel, T., Cartapanis, O., Chase, B., Dawson, A., Vernal, A. de, Engels, S., Jonkers, L., Marsicek, J., Moffa-Sánchez, P., Morrill, C., Orsi, A., Rehfeld, K., Saunders, K., Sommer, P.S., Thomas, E., Tonello, M., Tóth, M., Vachula, R., Andreev, A.,

## References

Bertrand, S., Biskaborn, B., Bringué, M., Brooks, S., Caniupán, M., Chevalier, M., Cwynar, L., Emile-Geay, J., Fegyveresi, J., Feurdean, A., Finsinger, W., Fortin, M.-C., Foster, L., Fox, M., Gajewski, K., Grosjean, M., Hausmann, S., Heinrichs, M., Holmes, N., Ilyashuk, B., Ilyashuk, E., Juggins, S., Khider, D., Koinig, K., Langdon, P., Larocque-Tobler, I., Li, J., Lotter, A., Luoto, T., Mackay, A., Magyari, E., Malevich, S., Mark, B., Massaferro, J., Montade, V., Nazarova, L., Novenko, E., Pařil, P., Pearson, E., Peros, M., Pienitz, R., Płóciennik, M., Porinchu, D., Potito, A., Rees, A., Reinemann, S., Roberts, S., Rolland, N., Salonen, S., Self, A., Seppä, H., Shala, S., St-Jacques, J.-M., Stenni, B., Syrykh, L., Tarrats, P., Taylor, K., van den Bos, V., Velle, G., Wahl, E., Walker, I., Wilmshurst, J., Zhang, E., Zhilich, S., 2020. A global database of Holocene paleotemperature records. Scientific data 7 (1), 115.

Kempen, B., Brus, D.J., Stoorvogel, J.J., 2011. Three-dimensional mapping of soil organic matter content using soil type–specific depth functions. Geoderma 162 (1-2), 107–123.

Khaledian, Y., Brevik, E.C., Pereira, P., Cerdà, A., Fattah, M.A., Tazikeh, H., 2017. Modeling soil cation exchange capacity in multiple countries. Catena 158, 194–200.

Kovačević, M., Bajat, B., Gajić, B., 2010. Soil type classification and estimation of soil properties using support vector machines. Geoderma 154 (3), 340–347.

Krige, D.G., 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of the Southern African Institute of Mining and Metallurgy 52 (6), 119–139.

Kursa, M.B., Rudnicki, W.R., 2010. Feature Selection with the Boruta Package. Journal of Statistical Software; Vol 1, Issue 11 (2010).

Lacoste, M., Minasny, B., McBratney, A., Michot, D., Viaud, V., Walter, C., 2014. High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape. Geoderma 213, 296–311.

LAGB, 2014. Bodenbericht Sachsen-Anhalt 2014 - Grundlagen, Parameter und Hintergrundwerte. Mitteilungen zu Geologie und Bergwesen von Sachsen-Anhalt 18, Halle (Saale), Germany, 74 pp. Accessed 9 July 2019.

Lal, R., 2010. Managing Soils and Ecosystems for Mitigating Anthropogenic Carbon Emissions and Advancing Global Food Security. BioScience 60 (9), 708–721.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Lin, L.I.-K., 1989. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics 45 (1), 255.

Lipton, Z.C., 2016. The Mythos of Model Interpretability, 9 pp. Accessed 12 October 2020.

Liu, F., Rossiter, D.G., Song, X.-D., Zhang, G.-L., Yang, R.-M., Zhao, Y.-G., Li, D.-C., Ju, B., 2016. A similarity-based method for three-dimensional prediction of soil organic matter concentration. Geoderma 263, 254–263.

Liu, F., Zhang, G.-L., Sun, Y.-J., Zhao, Y.-G., Li, D.-C., 2013. Mapping the Three-Dimensional Distribution of Soil Organic Matter across a Subtropical Hilly Landscape. Soil Science Society of America Journal 77 (4), 1241–1253.

Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154 (1), 138–152.

Martini, E., Werban, U., Zacharias, S., Pohle, M., Dietrich, P., Wollschläger, U., 2017. Repeated electromagnetic induction measurements for mapping soil moisture at the field scale: validation with data from a wireless soil moisture monitoring network. Hydrol. Earth Syst. Sci. 21 (1), 495–513.

Masek, J.G., Wulder, M.A., Markham, B., McCorkel, J., Crawford, C.J., Storey, J., Jenstrom, D.T., 2020. Landsat 9: Empowering open science and applications through continuity. Remote Sensing of Environment 248, 111968.

Matheron, G., 1971. The Theory of Regionalized Variables and Its Applications. École national supérieure des mines, Paris.

McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics 5 (4), 115–133.

McNeill, J.D., 1980a. Electrical conductivity of soils and rocks: Technical note TN-5, Mississauga, ON, Canada, 22 pp. http://www.geonics.com/pdfs/technicalnotes/tn5.pdf. Accessed 9 July 2019.

McNeill, J.D., 1980b. Electromagnetic Terrain Conductivity Measurement at Low Induction Numbers: Technical note TN-6, Mississauga, ON, Canada, 15 pp. http://www.geonics.com/pdfs/technicalnotes/tn6.pdf. Accessed 9 July 2019.

Meersmans, J., van Wesemael, B., Ridder, F. de, van Molle, M., 2009. Modelling the three-dimensional spatial distribution of soil organic carbon (SOC) at the regional scale (Flanders, Belgium). Geoderma 152 (1-2), 43–52.

Milne, G., 1936. Normal Erosion as a Factor in Soil Profile Development. Nature 138 (3491), 548–549.

Minasny, B., McBratney, A.B., 2006. Chapter 12 Latin Hypercube Sampling as a Tool for Digital Soil Mapping, in: Lagacherie, P., McBratney, A.B., Voltz, M. (Eds.), Digital soil mapping. An introductory perspective, vol. 31, 1. ed. ed. Developments in Soil Science 31. Elsevier, Amsterdam, pp. 153–606.

Minasny, B., McBratney, A.B., Mendonça-Santos, M.L., Odeh, I.O.A., Guyon, B., 2006. Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley. Soil Res. 44 (3), 233.

Minasny, B., Stockmann, U., Hartemink, A.E., McBratney, A.B., 2016. Measuring and Modelling Soil Depth Functions, in: Hartemink, A.E., Minasny, B. (Eds.), Digital Soil Morphometrics. Springer International Publishing, Cham, pp. 225–240.

## References


Montavon, G., Samek, W., Müller, K.-R., 2018. Methods for interpreting and understanding deep neural networks. Digital Signal Processing 73, 1–15.

Mulder, V.L., Lacoste, M., Richer-de-Forges, A.C., Martin, M.P., Arrouays, D., 2016. National versus global modelling the 3D distribution of soil organic carbon in mainland France. Geoderma 263, 16–34.

Murdoch, W.J., Singh, C., Kumbier, K., Abbasi-Asl, R., Yu, B., 2019. Definitions, methods, and applications in interpretable machine learning. Proceedings of the National Academy of Sciences of the United States of America 116 (44), 22071–22080.

Nelson, D.W., Sommers, L.E., 1983. Total Carbon, Organic Carbon, and Organic Matter. Agronomy Monographs, 539-579.

Pulido, M., Schnabel, S., Contador, J.F.L., Lozano-Parra, J., Gómez-Gutiérrez, Á., 2017. Selecting indicators for assessing soil quality and degradation in rangelands of Extremadura (SW Spain). Ecological Indicators 74, 49–61.

Quinlan, J.R., 1993. Combining Instance-based and Model-based Learning, in: Utgoff, P. (Ed.), Proceedings of the Tenth International Conference on International Conference on Machine Learning. Morgan Kaufmann, San Mateo, pp. 236–243.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566 (7743), 195–204.

Rentschler, T., Gries, P., Behrens, T., Bruelheide, H., Kühn, P., Seitz, S., Shi, X., Trogisch, S., Scholten, T., Schmidt, K., 2019. Comparison of catchment scale 3D and 2.5D modelling of soil organic carbon stocks in Jiangxi Province, PR China. PloS one 14 (8), e0220881.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F.S., Lambin, E.F., Lenton, T.M., Scheffer, M., Folke, C., Schellnhuber, H.J., Nykvist, B., Wit, C.A. de, Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P.K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R.W., Fabry, V.J., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P., Foley, J.A., 2009. A safe operating space for humanity. Nature 461 (7263), 472–475.

Roscher, R., Bohn, B., Duarte, M.F., Garcke, J., 2020. Explainable Machine Learning for Scientific Insights and Discoveries. IEEE Access 8, 42200–42216.

Roy, D.P., Wulder, M.A., Loveland, T.R., C.E., W., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R., Scambos, T.A., Schaaf, C.B., Schott, J.R., Sheng, Y., Vermote, E.F., Belward, A.S., Bindschadler, R., Cohen, W.B., Gao, F., Hipple, J.D., Hostert, P., Huntington, J., Justice, C.O., Kilic, A., Kovalskyy, V., Lee, Z.P., Lymburner, L., Masek, J.G., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R.H., Zhu, Z., 2014. Landsat-8: Science and product vision for terrestrial global change research. Remote Sensing of Environment 145, 154–172.

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.d.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 (5941), 680–681.

Schmidt, K., Behrens, T., Daumann, J., Ramirez-Lopez, L., Werban, U., Dietrich, P., Scholten, T., 2014. A comparison of calibration sampling schemes at the field scale. Geoderma 232-234, 243–256.

Schmidt, K., Behrens, T., Scholten, T., 2008. Instance selection and classification tree analysis for large spatial datasets in digital soil mapping. Geoderma 146 (1-2), 138–146.

Scholten, T., Goebes, P., Kühn, P., Seitz, S., Assmann, T., Bauhus, J., Bruelheide, H., Buscot, F., Erfmeier, A., Fischer, M., Härdtle, W., He, J.-S., Ma, K., Niklaus, P.A., Scherer-Lorenzen, M., Schmid, B., Shi, X., Song, Z., Oheimb, G. von, Wirth, C., Wubet, T., Schmidt, K., 2017. On the combined effect of soil fertility and topography on tree growth in subtropical forest ecosystems—a study from SE China. Journal of Plant Ecology 10 (1), 111–127.

Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Statistics and Computing 14 (3), 199–222.

Steyerberg, E.W., Harrell, F.E., 2016. Prediction models need appropriate internal, internal-external, and external validation. Journal of Clinical Epidemiology 69, 245–247.

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.

Teuber, S., Schweizer, B., 2020. Resources Redefined: Resources and ResourceComplexes, in: Teuber, S., Scholz, A.K., Scholten, T., Bartelheim, M. (Eds.), Waters. Conference Proceedings for "Waters as a Resource" of the SFB 1070 ResourceCultures and DEGUWA (Deutsche Gesellschaft zur Förderung der Unterwasserarchäologie e.V.), vol. 11. RessourcenKulturen 11, 9-20.

Tobler, W.R., 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography 46, 234.

Tóth, B., Weynants, M., Nemes, A., Makó, A., Bilas, G., Tóth, G., 2015. New generation of hydraulic pedotransfer functions for Europe. Eur J Soil Sci 66 (1), 226–238.

Trumbore, S., 2000. Age of soil organic matter and soil respiration: radiocarbon constraints on belowground C dynamics. Ecological Applications 10 (2), 399–411.

Turing, A.M., 1950. Computing machinery and intelligence. Mind LIX (236), 433–460.

Veronesi, F., Corstanje, R., Mayr, T., 2012. Mapping soil compaction in 3D with depth functions. Soil and Tillage Research 124, 111–118.

Veronesi, F., Corstanje, R., Mayr, T., 2014. Landscape scale estimation of soil carbon stock using 3D modelling. The Science of the total environment 487, 578–586.

## References

Viscarra Rossel, R.A., Adamchuk, V.I., Sudduth, K.A., McKenzie, N.J., Lobsey, C., 2011. Chapter Five - Proximal Soil Sensing: An Effective Approach for Soil Measurements in Space and Time, in: Sparks, D.L. (Ed.), Advances in Agronomy, vol. 113. Academic Press, pp. 243–291.

Viscarra Rossel, R.A., Behrens, T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma 158 (1-2), 46–54.

Viscarra Rossel, R.A., Chen, C., Grundy, M.J., Searle, R., Clifford, D., Campbell, P.H., 2015. The Australian three-dimensional soil grid: Australia's contribution to the GlobalSoilMap project. Soil Res. 53 (8), 845–864.

Viscarra Rossel, R.A., McBratney, A.B., Minasny, B. (Eds.), 2010. Proximal Soil Sensing. Springer, Dordrecht.

von Hebel, C., van der Kruk, J., Huisman, J.A., Mester, A., Altdorff, D., Endres, A.L., Zimmermann, E., Garré, S., Vereecken, H., 2019. Calibration, Conversion, and Quantitative Multi-Layer Inversion of Multi-Coil Rigid-Boom Electromagnetic Induction Data. Sensors 19 (21).

Wadoux, A.M.J.-C., Minasny, B., McBratney, A.B., 2020a. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. Earth-Science Reviews 210, 103359.

Wadoux, A.M.J.-C., Samuel-Rosa, A., Poggio, L., Mulder, V.L., 2020b. A note on knowledge discovery and machine learning in digital soil mapping. Eur J Soil Sci 71 (2), 133–136.

Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58 (2), 109–130.

Wolf, D., Faust, D., 2015. Western Mediterranean environmental changes: Evidences from fluvial archives. Quaternary Science Reviews 122, 30–50.

Wösten, J.H.M., Pachepsky, Y.A., Rawls, W.J., 2001. Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. Journal of Hydrology 251 (3), 123–150.

# Appendix

**Manuscript 1**

## Comparison of catchment scale 3D and 2.5D modelling of soil organic carbon stocks in Jiangxi Province, PR China

*PLoS ONE 14(8): e0220881*

Tobias Rentschler[1,2,*], Philipp Gries[1], Thorsten Behrens[1], Helge Bruelheide[3,4], Peter Kühn[1], Steffen Seitz[1], Xuezheng Shi[5], Stefan Trogisch[3,4], Thomas Scholten[1] and Karsten Schmidt[1,2]

[1] Chair of Soil Science and Geomorphology, Department of Geosciences, University of Tübingen, Tübingen, Germany
[2] SFB 1070 ResourceCultures, University of Tübingen, Tübingen, Germany
[3] Department of Geobotany and Botanical Garden, Institute of Biology, Martin Luther University Halle-Wittenberg, Halle, Germany.
[4] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
[5] Institute of Soil Science, State Key Laboratory of Soil and Sustainable Agriculture, Chinese Academy of Sciences, Nanjing, China

*Corresponding author
E-mail: t.rentschler@uni-tuebingen.de

Submitted: December 23, 2018; Accepted: July 25, 2019; Published: August 20, 2019

# Appendix

## Abstract

As limited resources, soils are the largest terrestrial sinks of organic carbon. In this respect, 3D modelling of soil organic carbon (SOC) offers substantial improvements in the understanding and assessment of the spatial distribution of SOC stocks. Previous three-dimensional SOC modelling approaches usually averaged each depth increment for multi-layer two-dimensional predictions. Therefore, these models are limited in their vertical resolution and thus in the interpretability of the soil as a volume as well as in the accuracy of the SOC stock predictions. So far, only few approaches used spatially modelled depth functions for SOC predictions. This study implemented and evaluated an approach that compared polynomial, logarithmic and exponential depth functions using non-linear machine learning techniques, i.e. multivariate adaptive regression splines, random forests and support vector machines to quantify SOC stocks spatially and depth-related in the context of biodiversity and ecosystem functioning research. The legacy datasets used for modelling include profile data for SOC and bulk density (BD), sampled at five depth increments (0-5, 5-10, 10-20, 20-30, 30-50 cm). The samples were taken in an experimental forest in the Chinese subtropics as part of the biodiversity and ecosystem functioning (BEF) China experiment. Here we compared the depth functions by means of the results of the different machine learning approaches obtained based on multi-layer 2D models as well as 3D models. The main findings were (i) that 3rd degree polynomials provided the best results for SOC and BD ($R^2$=0.99 and $R^2$=0.98; RMSE=0.36 % and 0.07 g $cm^3$). However, they did not adequately describe the general asymptotic trend of SOC and BD. In this respect the exponential (SOC: $R^2$=0.94; RMSE=0.56 %) and logarithmic (BD: $R^2$=84; RMSE=0.21 g $cm^{-3}$) functions provided more reliable estimates. (ii) random forests with the exponential function for SOC correlated better with the corresponding 2.5D predictions ($R^2$: 0.96 to 0.75), compared to the 3$^{rd}$ degree polynomials ($R^2$: 0.89 to 0.15) which support vector machines fitted best. We recommend not to use polynomial functions with sparsely sampled profiles, as they have many turning points and tend to overfit the data on a given profile. This may limit the spatial prediction capacities. Instead, less adaptive functions with a higher degree of generalisation such as exponential and logarithmic functions should be used to spatially map sparse vertical soil profile datasets. We conclude that spatial prediction of SOC using exponential depth functions, in conjunction with random forests is well suited for 3D SOC stock modelling, and provides much finer vertical resolutions compared to 2.5D approaches.

## Introduction

Soils are a fundamental part of ecosystem functioning and services [1]. As finite resources, soils contribute to food production, nutrient cycling, biodiversity and freshwater quality [2]. Furthermore, they are interconnected with other ecosystem functions and services, such as local and global climate alteration; and therefore, contribute indirectly to human well-being [3]. Among soil properties, soil organic carbon (SOC) plays an important role in this context. SOC increases the water-holding capacity (e.g. important for agriculture, forest and flood management), improves the physical properties of soils, such as nutrient availability for plants in agriculture and forestry, and accounts for carbon sequestration to mitigate climate change [4–6]. In forestry, there is strong interest in the effects of tree species and tree diversity on soil carbon input and mineralization as well as the net effects of these processes [7]. Knowledge about the interconnection between SOC, forests and the diversity of tree species as well as SOC stock degradation by soil erosion [8,9] and land cover change [10,11] can also help to implement countermeasures to reduce global warming [7]. Consequently, the implementation of a credible soil carbon auditing and monitoring to verify changes in SOC is crucial regarding soil security and carbon sequestration [7,12].

To preserve the functions and services provided by soils, a good quantitative understanding of the SOC stocks is required – both in the vertical domain of a soil profile as well as in the spatial domain over landscapes [13,14]. However, conventional soil maps use soil classes in horizontal dimension and soil horizons in vertical dimension. This categorical setup is often not precise enough and not well suited for interpreting soil functions and processes as well as for decision-making, since soil properties mostly vary continuous in space and time [15,16].

For the spatial prediction of continuous soil properties, such as SOC, methods of digital soil mapping (DSM) are suitable [17–19]. DSM is based on the soil forming factor concept [20] and the *scorpan* model introduced by McBratney et al. [21]. Both approaches illustrate soil information as a function of environmental covariates, influencing the process of soil formation. Terrain parameters, describing the shape of the land surface, are used widely as an environmental covariate in DSM. Terrain is an essential factor of soil formation and controls the effects of gravity, climate, lithology, water and biota [22–24]. Hence, models that are based on terrain parameters reproduce displacement and reallocation of soil (i.e. mass movements and soil erosion) and are of particular interest when modelling SOC at catchment scale [25]. Furthermore, terrain can not only be used to estimate or model soil displacement and reallocation, but also as a proxy for environmental covariates, which are not used as predictors, or inaccessible *scorpan*-factors. For instance, slope and aspect can serve as proxy for microclimate through its influence on local solar insulation [24]. The catchment area can

serve as a proxy for soil fertility because of terrain driven water and SOC accumulation [19] and elevation, slope and aspect can act as proxy for parent material, tectonics and periglacial climate through strike and dip of the geological sediments and down-cutting processes [22,23,26].

For spatially modelling soil properties, different approaches have been established to derive relationships between soil properties and environmental covariates. However, for a reliable estimation of SOC stocks, the vertical dimension is crucial [13]. A common way of three-dimensional mapping is to consider the vertical dimension as multiple two-dimensional predictions, which can be interpreted in a three-dimensional way [17,27–29]. Because, multi-layered predictions do not provide full 3D soil information, since they are limited to the mapped depth increments. Information of the space between the mapped depth increments has to be derived on an interpretative and subjective basis. One approach is to vertically interpolate the single layers to construct a volumetric model, which is computationally intensive [30,31].

Therefore, multi-layered models are referred to as pseudo-3D mapping or 2.5D mapping [32]. To overcome these drawbacks, it is favourable to map soil properties as continuous depth function in the spatial domain [13,18], where the vertical distribution of soil properties is represented by depth functions, that are predicted spatially. These predictions allow the calculation of SOC stocks over the integral of the functions [33] as well as the calculation of fully three-dimensional maps at any vertical resolution [32,34–37].

Besides geostatistical frameworks [38,39], different depth functions have been applied for 3D modelling: power, logarithmic [32,40], exponential decay [32,33], polynomial [34,36] and equal-area spline functions [31,41].

While with 2.5D mapping soil properties are directly predicted at specific depth levels using the environmental covariates [17,29], 3D approaches use environmental covariates to predict parameters of the depth functions [34], which are abstract soil properties. According to the *scorpan* model, soil properties can be spatially mapped with neighbourhood relations solely [21], which also have been used for 3D modelling [36,40,42,43]. Over the past years, machine learning techniques have become a standard technique in DSM due to several advantages like dealing with non-linearity or the handling of large datasets. Aldana Jague et al. [33] used multiple linear regression (MLR) to model SOC incorporating terrain covariates, while Gasch et al. [43] compared spatial and terrain covariates using random forests (RF) and regression kriging for mapping SOC at different depth layers. Piikki et al. [27] used multivariate adaptive regression splines (MARS) to model clay and sand fractions as well as organic matter based on proximal soil sensing data. Several other studies also suggest

that machine learning techniques, such as artificial neural networks (ANN; [41,44]), random forests (RF; [17]) and support vector machines (SVM; [45]), can be applied successfully in DSM.

The objectives of this study were to test the spatial prediction of four soil profile depth functions for modelling SOC content and bulk density with different machine learning methods based on multi-scale terrain covariates. The tested soil profile depth functions are polynomials of 2nd and 3rd degree, natural logarithmic and exponential functions. The machine learning methods used to model the depth functions spatially were multivariate adaptive regression splines (MARS), random forests (RF) and support vector machines (SVM) with radial basis functions. We validated the machine learning models with 10-fold cross-validation and evaluated the results of the 3D mapping approach by comparing it with the predictions of the more common multi-layered 2.5D modelling approach based on five layers.

## Material and methods

### Study area and sampling design

The BEF-China study sites are artificial biodiversity experiments on property leased and managed by the project partner Institute of Botany, Chinese Academy of Sciences, 20 Nanxincun, Xiangshan, Bejing, 100093, PR China. Field studies did not involve endangered or protected species and no specific permissions for field research were required.

The biodiversity and ecosystem functioning (BEF) China project [46] is located near Xingangshan, Jiangxi Province, PR China (UTM/WGS84: 50R 588000 3222000), about 400 km south-west of Shanghai (Fig 1). The study site is a topographically heterogeneous environment in a small catchment of 26.7 ha leased by the Institute of Botany of the Chinese Academy of Sciences (CAS). It features an elevation ranging from 105 to 275 m a.s.l., slopes inclined 29° in average and a maximum slope inclination of 45°, which are typically convex [19]. Non-calcareous slates with varying sand and silt content and grey-green sandstone constitute the bedrock. Predominant soil types are Endoleptic Cambisols with Anthrosols at the hillsides and Gleysols at the valley bottom. The mean soil depth is 0.6 m with underlaying isomorphic weathered slate (saprolite; [19]). Soil texture ranges from silt loam to silty clay loam [47]. The climate is typically subtropical with monsoons in summer, a mean annual temperature of about 17 ℃ and long-term average annual rainfall of about 1800 mm [48] but with a drier period from 2009 to 2012 [49].

About 18 ha were covered with 271 experimental plots. In total 8.7 ha at the valley bottom were not part of the experimental design due to paths and rivulets. Plots had a size of 25.8 m × 25.8 m (traditional Chinese unit of 1 mu, 1/15 ha) and were replanted in 2008 after clear-cut of a commercial Chinese fir plantation. One plot comprised 400 (20 × 20) trees in monocultures

and mixtures of 2, 4, 8, 16 and 24 species. Species composition of the plots was based on random as well as non-random (plant trait-oriented) extinction scenarios, where all species were represented equally (broken-stick design). The datasets used in this study comprised soil samples from random subsets of all species and species richness levels referred to as VIPs (Very Intensively Studied Plots). For details on the experimental design, see Bruelheide et al. [46] and Trogisch et al. [50].

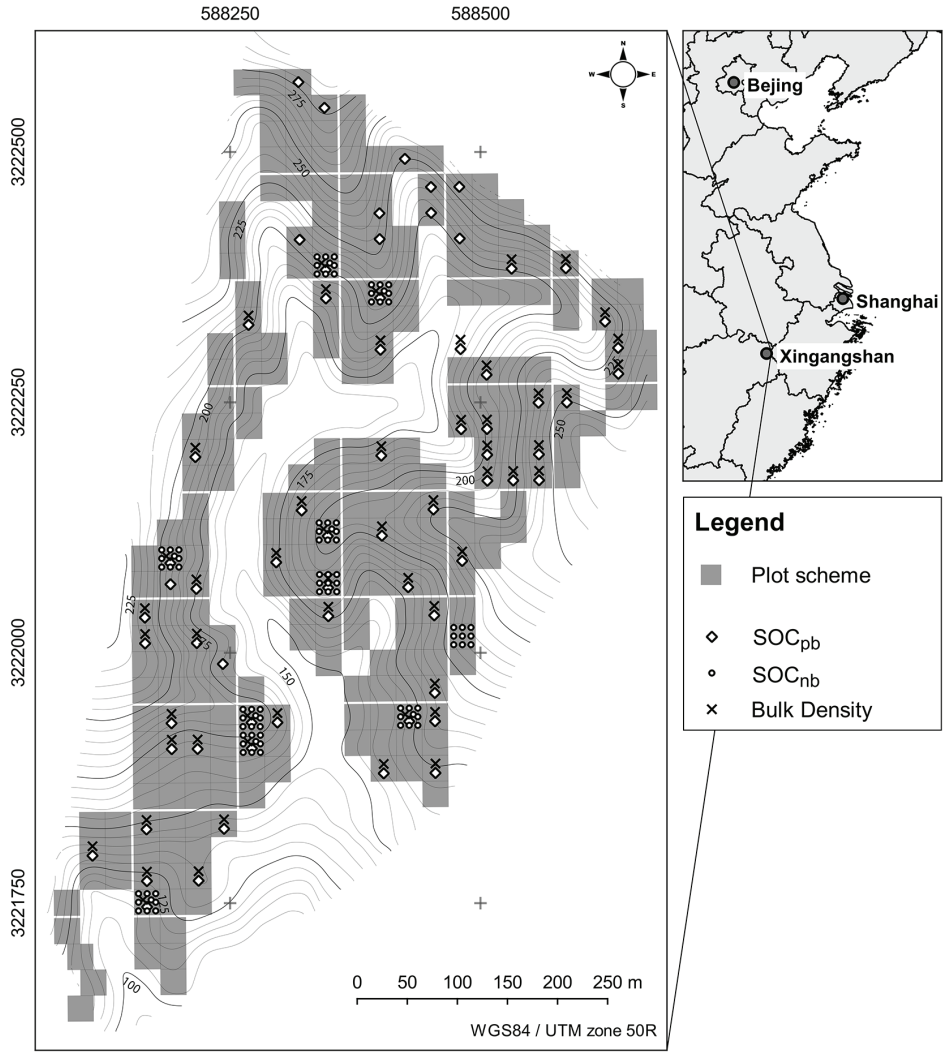


**Fig. 1**. Study area in mainland China with BEF-China plot scheme and indication of sampled plots. Upper right panel with permission by R. Hijmans; https://gadm.org/.

### Datasets

All described datasets are part of the legacy database of BEF-China. Soil sampling was conducted in 2014. Nine cores on a regular grid basis (3 cm in diameter) were taken at each of the 67 VIPs according to the BEF-China experimental design (Fig 1; [46]). The samples were bulked for each depth increment (0-5 cm, 5-10 cm, 10-20 cm, 20-30 cm and 30-50 cm)

and were referred to as dataset SOC (n=67; Fig 2). Fine roots and charcoal were sorted out manually. For dry combustion CNS-analysis, a Vario EL III (Elementar, Hanau, Germany) was used. Due to acidic soil conditions there was no detectable carbonate fraction, and thus total carbon represented SOC [19]. SOC content ranged from 5.06 to 0.35 % decreasing with depth.

Bulk density samples (n=55) were taken in April 2015 with soil sample rings (100 $cm^3$) and five replicates for each depth increment at the VIPs. Bulk density was determined gravimetrically and was referred to as dataset BD (Fig 2). Bulk density ranged from 0.75 to 1.84 g $cm^{-3}$ increasing with depth.

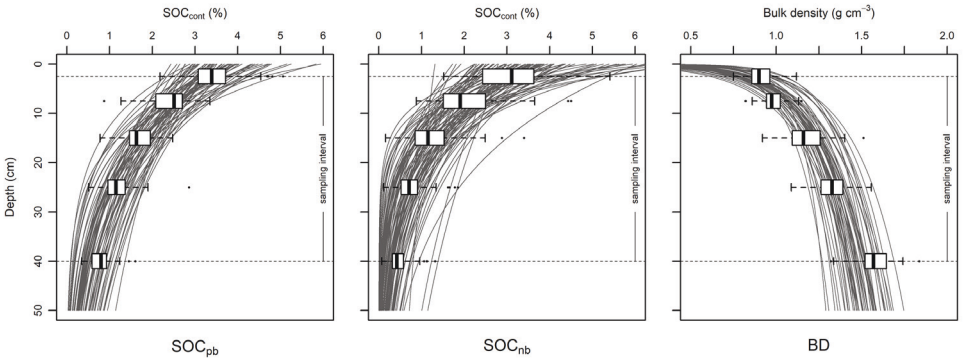


**Fig. 2.** Datasets for SOC and BD used in this study summarized in boxplots.

The boxplots show the variation of the SOC and BD values for each depth increment. SOC and BD samples were taken in five depth increments and 9 cores per plot were bulked (Note that depth increments do not increase linearly). The grey lines show model depth functions (3rd degree polynomial for SOC and natural logarithmic function for BD; see subsection "3D mapping with soil depth functions").

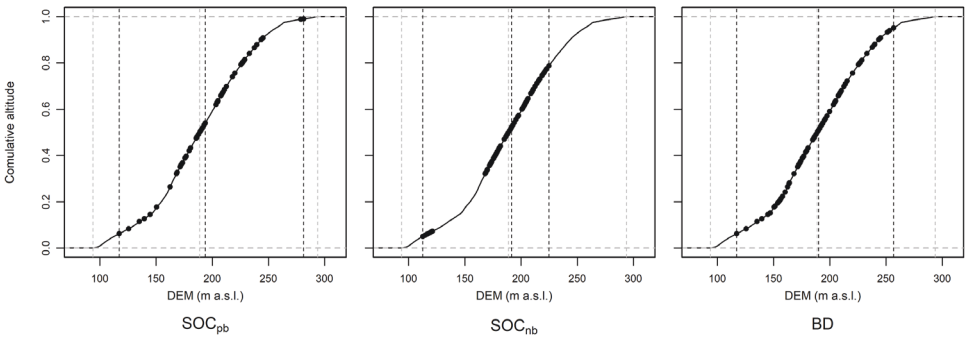


**Fig. 3.** Empirical cumulative distribution functions (ECDF) for SOC and BD datasets.

The ECDFs show the locations of the sampling sites in the state space of the elevation (DEM) in metres above sea level (m a.s.l.). The aim is to show the coverage of the DEM feature space by the samples. It can be seen that most samples are located in the mid-range of the elevation values. Therefore, predictions at grid locations which are only sparsely covered by the samples (i.e. locations close to the minimum and maximum values of the

DEM) may be less accurate. The minimum, median and maximum values of both datasets (DEM and sampling locations) are shown with vertical lines (dashed grey: DEM, dashed black: sampling locations) to compare the full range of the respective feature spaces.

Since some plots with SOC samples did not have BD data (Fig 1), both soil properties were modelled individually instead of calculating and modelling the SOC stocks directly. This 'model-then-calculate' approach is a useful alternative to the 'calculate-then-model' approach. Both were compared by Orton et al. [51].

The digital elevation model (DEM) had a resolution of 5 m and was generated by ordinary kriging [52] based on differential global positioning system data (DGPS) with 1956 points (73 points per ha; [19]). The distribution of datasets SOC and BD over the DEM is shown in Fig 3. Dataset SOC covered the elevation data more comprehensively compared to the dataset BD.

### Digital terrain analysis

Environmental covariates that describe the morphometry of a landscape are grouped in four major classes of terrain attributes: local, regional, combined (i.e. combinations of local and regional) and solar morphometric variables. Given that many terrain attributes can be calculated based on different equations or modelling approaches and because it is unknown which version would be most suitable for modelling SOC and BD within the study area, we used multiple established methods to derive single terrain attributes, if available. Given the circular nature of aspect, we used sine and cosine transformations to derive eastness and northness. Overall, we calculated 58 terrain attributes (Table 1) with SAGA GIS 2.3.1 [53].

**Table 1.** Terrain attributes used for SOC and bulk density modelling.

| | Covariates | Method | Author(s) |
|---|---|---|---|
| Local | Slope and aspect | Fitted 2nd degree polynomial | [54] |
| | | Fitted 3rd degree polynomial | [55] |
| | | Least squares fitted plane | [56] |
| | | Maximum triangle slope | [57] |
| | | Fitted 2nd degree polynomial | [58] |
| | Plan, profile, longitudinal, tangential and flowline curvature | Fitted 2nd degree polynomial | [54] |
| | | Fitted 3rd degree polynomial | [55] |
| | | Fitted 2nd degree polynomial | [58] |
| | Vertical distance to channel network | | [53] |
| | Sky visibility, sky view factor, direct and diffusive insolation | | [59] |
| Regional | Catchment area | Top-down<br>Recursive | [60] |
| Combined | Topographic Wetness Index (TWI) | Any combination of slope and catchment area | [61] |
| | Slope length and steepness factor (LS-Factor) | Any combination of slope and catchment area | [61,62] |

Terrain attributes derived from a DEM with a given resolution may not be suitable for landscape characterization and for digital soil mapping due to a non-representative DEM resolution [63], since the terrain attributes are not derived on the most relevant scale [64,65]. To examine the influence of scale, [65] applied simple smoothing (mean) filters with different neighbourhood sizes. This approach was applied on every terrain attribute used in this study with five circular neighbourhoods (radii of 1, 2, 4, 6 and 8 pixels), resulting in 290 terrain attributes in total. The maximum radius was set to 8 pixels to represent the local catena scale of 90 m.

**Machine learning techniques**

We compared three data mining methods to test the 3D prediction of soil profile depth functions for SOC and BD based on terrain covariates. Given the large number of 290 covariates (instances) and sample sizes of n=67 and n=55, not all available techniques could be applied. For example, the interpretable multiple linear regression (MLR) analysis used for spatial modelling of polynomial depth functions by Aldana Jague [34] requires more samples (n) than instances (p; [66]). Furthermore, we have to account for multi-collinearity. Many terrain covariates in this study are calculated by different algorithms for the same terrain attribute and on different spatial scales with the same algorithm, which is often seen as a constraint in machine learning [66]. To reduce the covariate space to either enable MLR or handle the 'curse of dimensionality', principal component analysis (PCA) is often applied. However, feature reduction with PCA can have negative effects on model accuracy with multi-scale terrain data and models with the full set of covariates have higher accuracies [65]. Other feature reduction methods increase accuracy only marginally [65]. In this study, we applied multivariate adaptive regression splines (MARS), random forests (RF) and support vector machine (SVM). These machine learning methods are robust against multi-collinearity, can handle n<p [66] and select the most informative covariates without expert knowledge. Further, we omitted feature reduction.

For modelling, R version 3.3.1 was used [67]. For accessing the machine learning packages, the uniform interface *caret* [68] was used, which also offers data handling and model validation methods.

***Multivariate adaptive regression splines (MARS)***

MARS was introduced by Friedman [69] and is a generalisation of recursive partitioning regression approaches using piecewise linear models. With its linear basis functions, it overcomes the discontinuous response of other recursive partitioning models like Classification and Regression Trees (CART; [70]) and can generate continuous surfaces. Therefore, prediction accuracy of MARS is expected to be higher [69]. MARS is a partial linear

function, where each new part is added with an exhaustive search for best fit and models a finite quantity of the regression. Thus, the model measures variable importance by its nature and is insensitive to non-informative instances. MARS require very little pre-processing and are non-affected by collinearity, since the predictor selection is random during iteration and redundant features are used equally [66]. This may affect measurement of variable importance and interpretation, which, however, is out of scope in this study. For modelling using MARS, the *earth* package version 4.4.6 [71] was used.

***Random forests (RF)***

RF is a widely used machine learning technique in digital soil mapping [17,22,64,72]. It was introduced by Breiman [73] and is an ensemble technique with CART [70] as a base learner. The single decision tree uses binary splits to create more homogenous groups in respect to the response. To grow an ensemble of trees, different random subsets of covariates (bootstrap sampling) and features (random set of features for every split) are used to build a single tree. The final prediction is created by averaging all individual tree outputs. Breiman [73] has proven that random forests with a large number of trees is robust against overfitting. Moreover, it is robust against noise, non-informative and correlated features. RF also returns feature importance measures (affected by correlation as MARS; [66]) and there is little need for fine-tuning [74]. The *randomForest* package version 4.6-12 [75] was used for modelling with RF.

***Support vector machine (SVM)***

Originally, SVM has been developed for classification problems [76]. It is a kernel method and uses hyperplanes to linearly separate classes of objects. For regression problems, Drucker et al. [77] developed support vector regression machines (SVR), which are an extension of SVM. Therefore, the term SVM is often used in both cases. The kernel function defines a transformation of the input data into a high dimensional feature space. In this feature space, it is possible to derive a linear regression hyperplane for non-linear relationships. Afterwards, it is back-transformed to non-linear space. Smola and Schölköpf [78] provide a comprehensive and detailed insight into SVR. The kernel used in this study is a radial basis function, where the scaling parameter $\sigma$ is estimated by *caret* after a method by Caputo et al. [79]. In contrast to MARS, Drucker et al. [77] suggest that SVM should be used when the number of features is larger than the number of instances, since its optimisation does not depend on the dimensionality of feature space. Furthermore, SVM is partially insensitive to outliers (depending on cost factor) and does not require feature reduction to reduce multi-collinearity [66]. The *kernlab* package version 0.9-25 [80] was used for radial support vector regression modelling.

### *Data pre-processing*

Some algorithms are sensitive to the scale and the range of the covariate space (e.g. SVM). To reduce effects of small values and little variance, SVM needs centred and scaled covariates [66], which was computed using the scale and centre-option in *caret*. To make all models comparable, this was also done for MARS and RF.

## Spatial 2.5D and 3D models

### *Differences between 2.5D models and spatial prediction of depth functions*

The environmental covariates were used to train regression models (MARS, RF and SVM) to predict SOC and BD. For 2.5D predictions this was done for each sampled depth increment individually, were we assigned the mid-depth of the sampled increments as depth of the respective layer. This method to obtain volumetric soil information has several advantages. For modelling of each standard depth individually, there are no further requirements to abstract soil information in terms of vertical variability, i.e. a soil profile function. Furthermore, there is no error propagation through secondary models that describe depth functions. On the other hand, in contrast to 3D modelling, 2.5D modelling has the disadvantage that the individual model outcomes are purely two-dimensional. Soil properties of the depth increments between the standard depths are not used in the model and have to be derived on an interpretative basis [15] or through further processing [30] after spatial prediction. However, this is a well-established and well-documented approach. Therefore, we compare the results of the 3D approach described below directly with the 2.5D results.

### *3D mapping with soil depth functions*

For the spatial modelling of depth functions, which we handled similar to the soil properties in terms of modelling, we applied 3rd degree polynomial functions proposed by Aldana Jague [34] and less flexible 2nd degree polynomials as well as logarithmic and exponential functions [32]. The workflow of the 3D mapping (Fig 4) of this study involved five main steps:

- Mathematical approximation of depth functions to the five depth increments with a linear least squares approach. These were
  - $f_2(x) = c_0 + c_1x + c_2x^2 + c_3x^3$ [34] (1)
  - $f_2(x) = c_0 + c_1x + x_2x^2$ (2)
  - $f_3(x) = c_1 \times ln(c_2x)$ [32] (3)
  - $f_4(x) = exp^{c_1+c_2x}$ cf. [32] (4)

- where $f_{1,2,3,4}(x)$ is SOC and BD at a specific depth x (depth of the lower corner of a voxel in cm), $c_0$ is the intercept that equals SOC and BD at depth 0 (cm) and the function coefficients $c_1$, $c_2$ and $c_3$ are dimensionless. This altogether described the vertical distribution of SOC in respect to depth x at a certain location.
- Evaluation of model error for all equations in (i).
- Spatial modelling of the function coefficients $c_1$, $c_2$, $c_3$ and $c_0$ (analogous to twodimensional modelling of SOC and BD) of the depth function with the lowest error (ii) with MARS, RF and SVM. The depth function parameters were treated and evaluated similar to a soil property.
- Evaluation of the cross-validation results for MARS, RF and SVM models of the depth function coefficients.
- Solving the depth functions with spatially modelled coefficients (iii) at each grid location to generate a three-dimensional model.

The depth functions were solved for depths from 0 cm to 50 cm in 5 cm increments. The resulting 11 depth layers (matrices) were stacked to two three-dimensional models (one for SOC and BD each), where individual values are represented by voxels, which are the volumetric 3D analogue of 2D pixels. Due to the nature of the polynomial depth functions, negative SOC predictions in the profiles are possible. Consequently, the values of these voxels had to be set to zero. This is not required for logarithmic and exponential functions.

Compared to the standard depth method, the main advantages of spatially modelled depth functions are a higher vertical resolution and the fact that the result can be interpreted as volumetric structure. Instead of pixels with SOC and BD information in multiple layers, volumetric elements – so called voxels – in a three-dimensionally georeferenced stack of matrices with user-defined vertical resolution are obtained. Since the depth functions are secondary models, the error which is propagated by the depth function model to the spatial model depends on the chosen function. Due to the limited number of samples per profile, cross-validation of the depth functions was omitted.

The final models for SOC and BD were validated internally against the measured values of the input datasets.

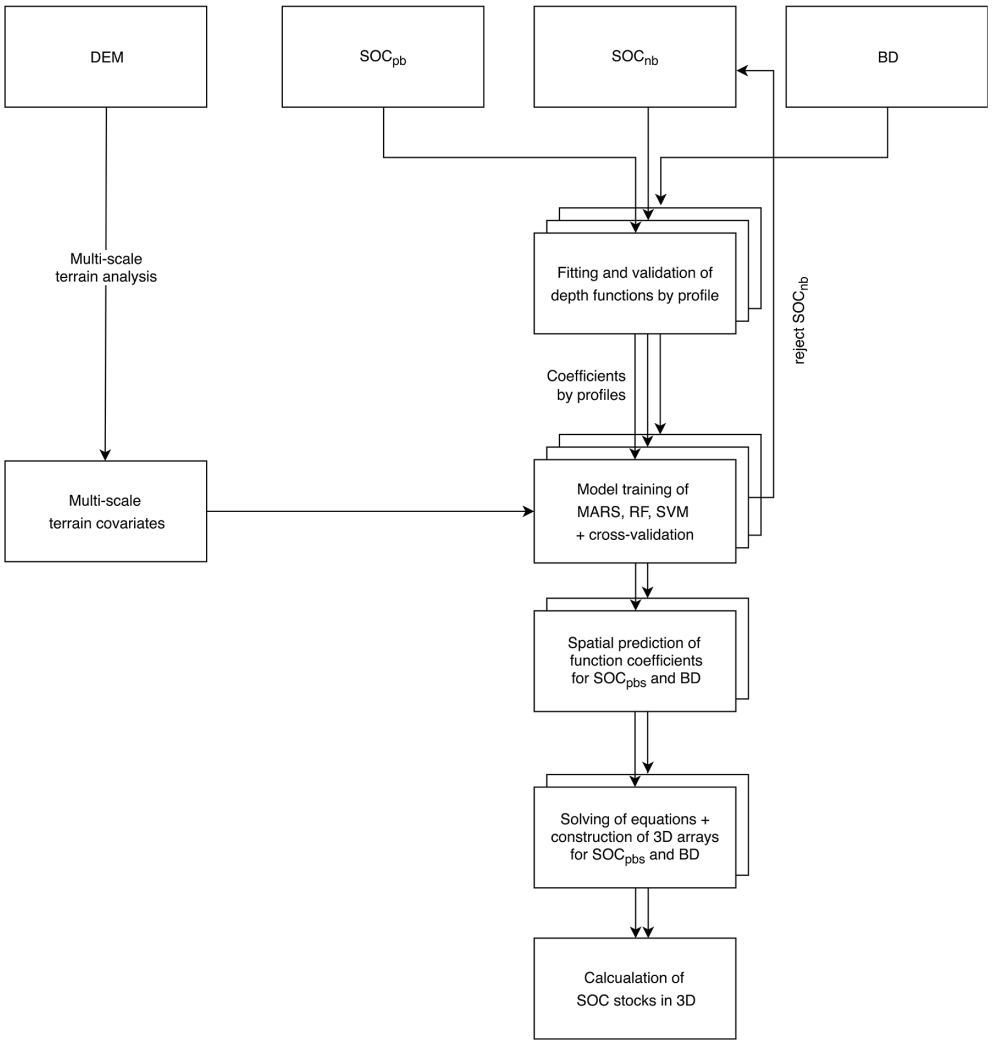


**Fig. 4.** Flow chart summarizing the methodology steps of the 3D mapping and the used datasets at each step.

### Validation and evaluation

The evaluation consists of two independent steps for the 2.5D multi-layered model predictions and the volumetric 3D model predictions of SOC and BD, where we treat the depth function parameters as soil properties.

In a first step, we evaluated each model of the soil properties SOC and BD as well as the spatial models of the depth function parameters, by using a 10-fold cross-validation with the coefficient of determination ($R^2$) and the root mean square error (RMSE) as quality criterion. In this step, the models were tuned over the default grid- or hyper-learning sequence of parameters [81] using the tune grid function of *caret* to identify the most suitable combination of tuning parameters with the lowest RMSE and to reduce the model error, while preserving the models ability to generalise. The tuning parameters are *degree* and *nprune* for MARS, *mtry* for RF and *cost* for SVM. For RF *ntree* was set to the default value and σ for SVM was

calculated by a method after Caputo et al. [79]. All models used the same set of folds to make cross-validation results comparable. The final models were selected from this sequence by the lowest RMSE.

To estimate the effect of overfitting of the depth function models based on grid learning, we evaluated the 3D model results with the datasets SOC and BD by $R^2$ and RMSE (observed-predicted-evaluation). Overfitting is indicated by large differences in the prediction error between the training and the validation sets [81].

Further, we compared the 3D models against the 2.5D predictions of the same datasets to evaluate the performance of the 3D models. We chose this approach, because the legacy datasets are too small to hold out a larger subset for independent validation. The model results should be similar, if the spatial prediction of depth function parameters is reproducing the spatial distribution of the soil properties. This means that independently from the modelling framework (modelling of SOC and BD or modelling depth function as soil property) the results of the 3D model are reasonable, if both models are similar.

We see this comparison as a valid method for the evaluation of the 3D models, since Brus et al. [38] report strong correspondence between 2.5D and 3D geostatistical models and MARS, RF and SVM are well established for 2D and 2.5D soil mapping and in data science [17,27,66]. Therefore, we use the 2.5D layered predictions at the specific mid-depth of the increments as reference predictions. For the comparison between the 2.5D models and the corresponding depths of the 3D models, we used the coefficient of determination $R^2$, Lin's concordance correlation coefficient ($\rho_c$; [82]), which validates the models against the 1:1 line, and the RMSE.

**Estimation of SOC stocks**

The three-dimensional array of SOC stocks was calculated by

$$SOC_{stocks} = \frac{SOC}{100} \times BD \times 500^2 \times 5 \qquad (5)$$

where $SOC_{stocks}$ (g voxel$^{-1}$) is the soil organic carbon storage, SOC is SOC content (%), BD is bulk density (g cm$^{-3}$), $500^2$ is the base area of a voxel (cm$^2$) related to the DEM resolution of 500 cm and 5 is the vertical resolution in cm. Consequently, 1 voxel represented 1.25 m$^3$ of soil. Adjustment with the fraction of coarse material (> 2 mm) was omitted, since the coarse fraction was negligible low (< 5 vol.-%) at the VIPs and cannot be determined precisely by coring. According to Orton et al. [51] calculating the SOC stocks from two models of SOC and BD is an useful alternative when the samples are not taken at the same locations.

## Results

### 2.5D predictions of standard depths as reference

For the models of SOC, the mean cross-validation $R^2$ of MARS was 0.33 with a root mean square error of 0.39%, compared to RF with an $R^2$ of 0.41 (RMSE 0.34%) and SVM with an $R^2$ of 0.39 (RMSE 0.35%; cf. Table 2). Models for BD showed a mean $R^2$ of 0.43 (MARS), 0.39 (RF) and 0.39 (SVM) and mean RMSE values of 0.09 g $cm^{-3}$ (MARS), 0.08 g $cm^{-3}$ (RF) and 0.08 g $cm^3$ (SVM). In addition to the mean values, Table 2 shows the prediction accuracies and the RMSE's for each depth increment and all three machine learning techniques of both SOC and BD.

**Table 2.** Performance of 10-fold cross-validation for MARS, RF and SVM applied on the sampled standard depths of SOC and BD.

| | depth | $R^2$ | | | RMSE | | |
|---|---|---|---|---|---|---|---|
| | (cm) | MARS | RF | SVM | MARS | RF | SVM |
| SOC (%) | 0 - 5 | 0.28 | 0.41 | 0.37 | 0.59 | 0.48 | 0.51 |
| | 0 - 10 | 0.25 | 0.41 | 0.42 | 0.46 | 0.4 | 0.4 |
| | 10 - 20 | 0.31 | 0.31 | 0.26 | 0.37 | 0.32 | 0.34 |
| | 20 - 30 | 0.46 | 0.47 | 0.46 | 0.3 | 0.28 | 0.29 |
| | 30 - 50 | 0.38 | 0.45 | 0.43 | 0.24 | 0.2 | 0.21 |
| | $\bar{x}$ | 0.34 | 0.41 | 0.39 | 0.39 | 0.34 | 0.35 |
| BD (g $cm^{-3}$) | 0 - 5 | 0.51 | 0.53 | 0.61 | 0.07 | 0.06 | 0.06 |
| | 0 - 10 | 0.5 | 0.52 | 0.49 | 0.07 | 0.06 | 0.06 |
| | 10 - 20 | 0.31 | 0.26 | 0.24 | 0.11 | 0.11 | 0.11 |
| | 20 - 30 | 0.41 | 0.35 | 0.33 | 0.1 | 0.1 | 0.1 |
| | 30 - 50 | 0.42 | 0.31 | 0.3 | 0.09 | 0.09 | 0.09 |
| | $\bar{x}$ | 0.43 | 0.39 | 0.39 | 0.09 | 0.08 | 0.08 |

### Soil depth functions

For SOC, all equations showed $R^2$ values higher than 0.9 (0.99 for $f_1$, 0.96 for $f_2$, 0.96 for $f_3$ and 0.94 for $f_4$) with a RMSE ranging from 0.36 ($f_1$) to 0.7 % ($f_2$). For BD, the performance in terms of $R^2$ was similar (RMSE = 0.07 g $cm^{-3}$), except for $f_3$ with $R^2$ = 0.84 (RMSE = 0.22 g $cm^{-3}$), which is the natural logarithmic function. The 3rd degree polynomial ($f_1$) resulted in the best fits for SOC and BD. However, the general trend of SOC in the profiles was exponential (Fig 2). Hence, both the 3rd degree polynomial and the exponential functions were chosen for further spatial modelling and comparison in this study. With higher errors and without being able to reproduce the general trend in the profiles profile the 2nd order polynomial ($f_2$) was omitted in the following steps.

# Appendix

## Spatial modelling of soil depth functions

The cross-validation results for the machine learning methods applied on the depth functions (c.f. Table 3) showed, that the polynomial depth functions for MARS, RF and SVM for SOC were comparable in their goodness of fit with marginal differences (mean $R^2$ from 0.3 to 0.32). $R^2$ of the exponential depth functions ranged from 0.3 for MARS to 0.44 for RF.

The models of the function coefficients could not be compared directly because $c_0$ represented the SOC in % and BD in g $cm^{-3}$, whereas $c_1$, $c_2$ and $c_3$ were dimensionless. Hence, we compared these models by the normalised RMSE (nRMSE), which is the RMSE divided by the coefficients range (Table 3). The nRMSE showed little variation of around 0.18 for all coefficient predictions of the 3rd polynomial depth function of SOC. RF had the lowest mean of nRMSE over all coefficients (0.17). The lowest nRMSE (0.09) for SOC was achieved by the exponential depth functions (RF and SVM).

The models based on the 3rd degree polynomial depth functions of BD had a mean $R^2$ of about 0.23-0.4, while the mean nRMSE was about $3\times10^4$, due to the low performance of models with $c_3$. Given such high errors, none of the models could reasonably predict the 3rd degree polynomial depth function for bulk density. The exponential function was not able to reproduce the vertical trend of BD. Thus, we used the logarithmic depth function, although it fitted the five depth increments least. However, these spatial depth function models performed better (mean $R^2$ from 0.36 to 0.45; nRMSE of about 0.16 for SVM).

**Table 3.** Performance of a 10-fold cross-validation for MARS, RF and SVM applied on function coefficients of a 3rd degree polynomial ($f_1$ for SOC and BD with four coefficients) and natural logarithmic function ($f_3$ for BD with two coefficients).

| | | $R^2$ | | | RMSE | | | nRMSE | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | MARS | RF | SVM | MARS | RF | SVM | MARS | RF | SVM |
| SOC ($f_1$) | $c_0$ | 0.29 | 0.28 | 0.26 | 0.83 | 0.75 | 0.79 | 0.20 | 0.18 | 0.19 |
| | $c_1$ | 0.36 | 0.43 | 0.46 | 0.15 | 0.13 | 0.14 | 0.22 | 0.19 | 0.20 |
| | $c_2$ | 0.29 | 0.28 | 0.24 | 0.008 | 0.007 | 0.007 | 0.2 | 0.18 | 0.18 |
| | $c_3$ | 0.3 | 0.21 | 0.31 | 0.0001 | 0.0001 | 0.0001 | 0.14 | 0.14 | 0.14 |
| | $\bar{x}$ | 0.31 | 0.3 | 0.32 | - | - | - | 0.19 | 0.17 | 0.18 |
| BD ($f_1$) | $c_0$ | 0.56 | 0.45 | 0.38 | 0.09 | 0.09 | 0.09 | 0.23 | 0.2 | 0.20 |
| | $c_1$ | 0.38 | 0.34 | 0.218 | 0.02 | 0.02 | 0.02 | 0.14 | 0.14 | 0.14 |
| | c2 | 0.38 | 0.17 | 0.26 | 0.001 | 0.001 | 0.001 | 0.2 | 0.2 | 0.2 |
| | $c_3$ | 0.25 | 0.27 | 0.31 | 0.00002 | 0.00002 | 0.00002 | $1.3\times10^5$ | $1.2\times10^5$ | $1.2\times10^5$ |
| | $\bar{x}$ | 0.39 | 0.31 | 0.28 | - | - | - | $3.2\times10^4$ | $3\times10^4$ | $3\times10^4$ |
| BD ($f_3$) | $c_1$ | 0.56 | 0.48 | 0.53 | 0.09 | 0.09 | 0.09 | 0.2 | 0.18 | 0.18 |
| | $c_2$ | 0.34 | 0.24 | 0.2 | 0.03 | 0.04 | 0.04 | 0.14 | 0.19 | 0.14 |
| | $\bar{x}$ | 0.45 | 0.36 | 0.36 | - | - | - | 0.17 | 0.19 | 0.16 |

Note that coefficients dimensions are different and specifying a mean of the RMSE is not reasonable.

### Evaluation of 3D predictions

For the comparison of 3D models against the 2.5D reference predictions, we used the coefficient of determination $R^2$, Lin's concordance correlation coefficient ρc and the RMSE in corresponding depths (Table 4).

The three-dimensional MARS prediction for SOC with the 3rd degree polynomial depth function showed the largest difference to its counterpart. The prediction at 2.5 cm ranged from close to zero to 15 % SOC compared to 1.5 to 4 % SOC in the two-dimensional prediction (Fig 5). The other depth increments showed a similar pattern with values down to 15 % SOC. For the 2.5 cm increment the performance of RF was slightly better than that of SVM, but subsequently dropped with increasing depth. Especially at 40 cm, but also at 25 cm and 15 cm, the three-dimensional prediction of RF differed more from the two-dimensional predictions than the three-dimensional predictions of SVM differed from their counterparts. There was no distinct over- or underestimation of RF, but random scattering between -4 and 4 % SOC for 40 cm (Fig 5). SVM showed lower deviation at 15 cm, 25 cm and even 40 cm. There were less predictions with negative values and less scattering. The predicted depth intersections of spatially modelled depth functions corresponded to the two-dimensional predictions by SVM largely by $R^2$ and $\rho_c$, while RMSE is low (Table 4).

In contrast, the 3D predictions of RF and SVM based on the exponential function showed good correspondence for all five depth increments (Table 4). The 3D predictions overestimated SOC for the 0-5 and 5-10 cm increments and underestimated it for 20-30 and 30-50 cm slightly due to the exponential nature of the equation, but there was no wide scattering as it was the case with the polynomial prediction for RF.

The results of the internal validation showed high correspondence between the chosen models (RF with exponential function for SOC and RF with logarithmic function for BD) and respective input data at all five sampled depth increments (Table 5). The $R^2$ and RMSE values of the internal validation were similar to the validation results of the model comparison, indicating that model overfitting of both models is similar (Table 4). This partly accounts to the propagation error of the profile depth function. The spatial prediction of the exponential function for SOC had an average $R^2$ of 0.79 with an average RMSE of 0.33 % and the prediction of the logarithmic function used for BD had a $R^2$ of 0.77 with an average RMSE of 0.14 g $cm^{-3}$.

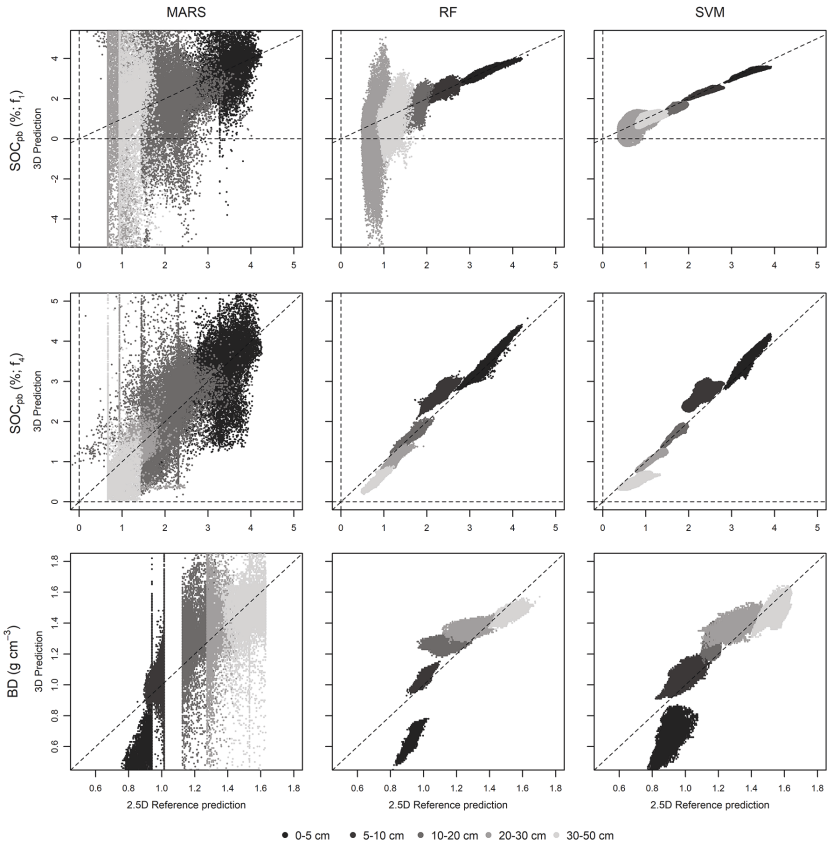


**Fig. 5.** 3D predictions of sampled depth increments plotted against corresponding 2.5D predictions.

3D prediction of SOC was calculated with 3rd degree polynomials (upper row) and exponential function (middle row). The 3D prediction for BD with logarithmic function (lower row).

### SOC stocks

The 2.5D models showed SOC stocks of 61.9 Mg ha$^{-1}$ from 0 to 40 cm, with 19, 14.7, 12, 8.9 and 7.3 Mg ha$^{-1}$ in the individual depth increments (from surface downwards).

The 3D model predicted 78.3 Mg ha$^{-1}$ over the whole interval. The upper 20 cm of soil contained about 46.4 Mg ha$^{-1}$. This depth is often designated as topsoil [83,84] and is also a critical soil depth for modelling plant productivity and community assembly [85]. 31.9 Mg ha$^{-1}$ SOC are stored in the subsoil from 20 to 40 cm. Considering that the rooting depth varies, depending on the species and individual age, a static discrimination between topsoil and subsoil may be not appropriate. The model showed that plants with shallow roots down to 5 cm mainly interacted with a carbon pool of 10.9 Mg ha$^{-1}$, whereas plants with roots in 25 cm depth interacted with a pool of 54.5 Mg ha$^{-1}$. Fig 6 shows the 3D prediction of SOC stocks as vertical intersections of the solum. The highest stocks in the upper 5 cm were predicted in the central upper slopes and at the western slopes. Predictions for this depth at the valley bottom were around 20 % lower. However, at the valley bottom the predictions for intermediate

depth increments (around 30 cm) were higher than predictions at the upslope positions. The depth function for SOC stocks was much steeper and the SOC stock decline with depth was more pronounced at upslope positions compared to downslope and valley positions.

**Table 4**. Internal validation results of the final 3D models with the exponential function for SOC and the logarithmic function for BD.

| | | $R^2$ | | | $\rho_c$ | | | RMSE | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | MARS | RF | SVM | MARS | RF | SVM | MARS | RF | SVM |
| SOC (%; $f_1$) | 2.5 cm | 0.02 | 0.92 | 0.89 | -0.03 | 0.95 | 0.79 | 4.13 | 0.09 | 0.12 |
| | 7.5 cm | 0 | 0.69 | 0.89 | -0.01 | 0.81 | 0.85 | 3.84 | 0.16 | 0.1 |
| | 15 cm | 0.02 | 0.41 | 0.72 | -0.03 | 0.43 | 0.47 | 2.42 | 0.37 | 0.22 |
| | 25 cm | 0 | 0.17 | 0.45 | 0 | 0.17 | 0.66 | 9.09 | 0.68 | 0.15 |
| | 40 cm | 0 | 0.07 | 0.15 | 0 | 0.04 | 0.31 | 46.96 | 1.73 | 0.33 |
| | $\bar{x}$ | 0.01 | 0.45 | 0.62 | -0.01 | 0.48 | 0.62 | 13.29 | 0.61 | 0.18 |
| SOC (%; $f_4$) | 2.5 cm | 0 | 0.96 | 0.93 | 0 | 0.93 | 0.79 | 19.22 | 0.11 | 0.14 |
| | 7.5 cm | 0.1 | 0.84 | 0.67 | 0 | 0.39 | 0.29 | 26.25 | 0.38 | 0.35 |
| | 15 cm | 0 | 0.89 | 0.88 | 0 | 0.94 | 0.93 | 29.42 | 0.06 | 0.05 |
| | 25 cm | 0.06 | 0.85 | 0.93 | 0 | 0.55 | 0.79 | 31.28 | 0.21 | 0.1 |
| | 40 cm | 0.02 | 0.88 | 0.75 | 0 | 0.31 | 0.26 | 32.71 | 0.31 | 0.3 |
| | $\bar{x}$ | 0.04 | 0.88 | 0.83 | 0 | 0.62 | 0.61 | 27.78 | 0.21 | 0.19 |
| BD (g $cm^{-3}$; $f_3$) | 2.5 cm | 0.02 | 0.94 | 0.87 | -0.05 | 0.39 | 0.53 | 0.48 | 0.09 | 0.07 |
| | 7.5 cm | 0 | 0.8 | 0.71 | 0 | 0.29 | 0.24 | 0.44 | 0.09 | 0.1 |
| | 15 cm | 0.01 | 0.66 | 0.5 | -0.05 | 0.48 | 0.31 | 0.46 | 0.05 | 0.07 |
| | 25 cm | 0.01 | 0.44 | 0.57 | 0.02 | 0.59 | 0.53 | 0.43 | 0.04 | 0.04 |
| | 40 cm | 0.02 | 0.76 | 0.43 | -0.12 | 0.17 | 0.08 | 0.64 | 0.1 | 0.11 |
| | $\bar{x}$ | 0.01 | 0.72 | 0.62 | -0.04 | 0.38 | 0.34 | 0.49 | 0.07 | 0.08 |

**Table 5**. Coefficient of correlation ($R^2$), Lin's concordance correlation coefficient ($\rho_c$) and RMSE of 2.5D reference predictions and correspondent depths of 3D predictions with polynomial ($f_1$), logarithmic ($f_3$) and exponential ($f_4$) depth function.

| | SOC (%) | | BD (g $cm^{-3}$) | |
|---|---|---|---|---|
| | $R^2$ | RMSE | $R^2$ | RMSE |
| 2.5 cm | 0.88 | 0.32 | 0.87 | 0.29 |
| 7.5 cm | 0.74 | 0.47 | 0.85 | 0.07 |
| 15 cm | 0.77 | 0.24 | 0.72 | 0.12 |
| 25 cm | 0.76 | 0.29 | 0.74 | 0.08 |
| 40 cm | 0.8 | 0.31 | 0.66 | 0.12 |
| $\bar{x}$ | 0.79 | 0.33 | 0.77 | 0.14 |

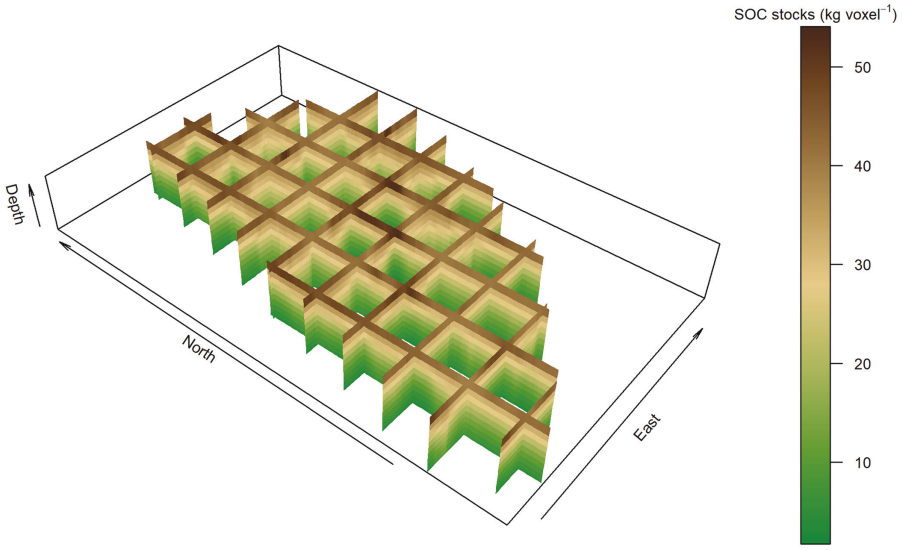


**Fig. 6.** Three-dimensional prediction of SOC stocks for the whole catchment.

The final 3D SOC stock model is shown in vertical slices 150 m apart to display the vertical variability, which is larger than the spatial variability.

## Discussion

### 2.5D predictions of standard depths as reference

As RF returned the lowest error for the 2.5D models, this was the best choice for modelling SOC. SVM ranked slightly below. Compared to the results presented by Lacoste et al. [30], who used Cubist for 2.5D SOC stock mapping, the accuracy of our results was similar and reasonable.

However, the sampled VIPs do not represent the terrain of the study site adequately, since they were chosen based on species richness levels, which were distributed randomly, and not representative for the study site. For example, a representative sampling design could be achieved with Conditioned Latin Hypercube Sampling (cLHS) [72,86].

For bulk density SVM and RF performed equal by means of $R^2$ and RMSE and showed a similar pattern, especially at 15 cm and 40 cm. MARS performed least for BD. In general, RF resulted in the most stable predictions and is therefore recommended over SVM.

### Evaluation of 3D predictions

The negative values in the prediction results and the pronounced difference between the 3D models, with predictions up to 15 % SOC, and the 2.5D models indicated that MARS is not capable of adequately predicting the depth functions in space, although the cross-validation showed similar results as for RF and SVM models. The latter showed better correspondence between the 3D and the 2.5D models (Fig 5, Table 4). According to the results of the direct

comparison between the multi-layered prediction and the corresponding depths in the 3D model, RF with exponential functions was most suitable for SOC modelling. RF and SVM with polynomials performed well at upper depth increments and less in lower increments. MARS models were not suitable of reproducing the 2.5D predictions. Lower performance of all techniques with polynomials in the lower depth increments may be referred to lower influence of the terrain as a driving factor to explain SOC accumulation and redistribution (e.g. by erosion). Other factors accounting for SOC redistribution in deeper soil horizons may be bioturbation or vertical transport in the liquid soil phase. Additionally, it is possible, that accumulation layers in the solum, that would reflect the lateral distribution, were not fully covered by the legacy dataset and, therefore, the interpretation remains difficult. All these processes and others relevant for SOC concentration as well as SOC stocks cannot be fully covered by a distinct set of terrain parameter and lead to a dilution effect by predicting the deeper horizons. Lower accordance of the models also may be referred to uncertain models of function coefficients ($c_3$) and ($c_2$), which have significant influence at greater depths (cubic and squared) and exponentiate up this error. Based on the results, we chose RF with exponential depth functions for three-dimensional mapping of SOC and the logarithmic depth function for BD.

**SOC stocks**

Compared to other studies in this area, the estimated SOC stocks were well in line. Scholten et al. [19] calculated mean SOC stocks of 70 Mg $ha^{-1}$ for the upper 50 cm with the same data but a different approach. Chen et al. [83] compared five plantations with different species in five age groups and calculated SOC stocks for the upper 20 cm. Especially the age of the trees and shrubs and their biomass have a strong impact on SOC stocks. Very young forest communities showed SOC stocks ranging from 20 to 25 Mg $ha^{-1}$ and plantations with older trees of 7 to 10 years 30-40 Mg $ha^{-1}$. The latter were slightly older than the trees of BEF-China, where 42 Mg $ha^{-1}$ were predicted. Diverse species pools in these studies may explain differences. Tang et al. [87] found SOC stocks in the top 60 cm in bamboo forests ranging from 60 to 200 Mg $ha^{-1}$.

The introduced approach is capable of summing SOC stocks at any depth interval. Since topsoil depth varies spatially, conventional static assumptions of topsoil thickness can result in inaccurate SOC stock calculations for individual horizons. Incorporating spatial models of topsoil depth into 3D SOC stock mapping can overcome this drawback and help to improve ecological and biodiversity models as conducted in the BEFChina experiment. In particular, consideration of biotic predictors like forest biomass, tree species richness and functional plant diversity might further improve model fit and accuracy of estimated SOC stocks [14].

This would allow one to quantify terrain-specific effects of changes in forest cover and composition on SOC stocks. The developed models could also help to identify areas that are especially prone to loss of SOC stocks (e.g. by soil erosion or land cover change).

Furthermore, continuous three-dimensional SOC mapping can support models of a national SOC inventory. Yang et al. [37] applied depth functions to categorical soil types and estimated SOC stocks for mainland China. Combining models with high vertical resolution by Yang et al. [37] and continuous spatial modelling like in this study can improve accuracy of SOC mapping compared to the categorical mapping approach. This combination can also help to estimate and understand carbon fluxes between topsoil and subsoil [88] as well as between soil and the atmosphere [5]. Both objectives play major roles in inventory estimation, SOC auditing and decision making in respect to ecosystem services and carbon sequestration [1,5,7,12,89].

## Conclusion

This study comprises the spatial prediction of soil depth functions for three-dimensional modelling of SOC and bulk density. The spatial prediction of the function coefficients enabled the calculation of two three-dimensional arrays by solving the depth functions for depths from 0 to 50 cm by 5 cm increments. This was used to estimate the SOC stocks in high spatial (5 m) and vertical (5 cm) resolution. The main conclusions of this study are:

- The general trend of SOC as visualised by the boxplots (Fig 2) was exponential. However, polynomial depth functions described the soil profiles for SOC with higher accuracy and the logarithmic functions for BD showed better results in spatial modelling. Therefore, we conclude that functions resulting in high accuracies based on the soil profile data may not be the most suitable for spatial modelling, as they may overfit the vertical trend of SOC content.
- The 3D RF models correspond best with the 2.5D counterparts ($R^2$ up to 0.96). Thus, RF is recommended to predict SOC based on exponential depth functions and bulk density with logarithmic depth functions in high vertical resolution. The 2.5D and 3D predictions of SOC with RF correlated much better, especially when using exponential functions, and lacked accuracy in deeper layers for SOC when modelled based on polynomial functions.
- Comparisons between conventional 2D and 2.5D predictions at the sampled depth and the corresponding depth of the three-dimensional predictions showed that MARS is not suitable for modelling corresponding 2.5D and 3D models, although cross-validation of the individual models showed similar performance in $R^2$.

Minor conclusions are: polynomial functions may be an option, when the problem of propagated errors and the ability to generalise in the horizontal domain is investigated further, however, polynomials of any degree have to be used carefully. To overcome these shortcomings, a higher sampling density in the vertical and horizontal domain and in combination with other depth functions, such as equal-area splines [90], should be considered, since exponential functions are not suitable for soil properties that do not increase or decrease continuously.

The 3D approach presented in this study is promising for SOC auditing in various disciplines and especially for decision making regarding climate and land use policies. Future work should focus on sampling design to cover valley positions outside the established plots at site A of BEF-China project. Given the dynamics of SOC stocks, we recommend the analyses of time series data and the expansion of the current database for four-dimensional models.

## Acknowledgements

We are indebted to the BEF-China members from China, Switzerland and Germany. In particular, we thank Lars Arne Meier, Philipp Goebes, Dominik Leimgruber, Chen Lin, Yang Bo and Zhengshan Song for their assistance with field and lab work.

## References

Adhikari K, Hartemink AE. Linking soils to ecosystem services — A global review. Geoderma. 2016; 262: 101–111. doi: 10.1016/j.geoderma.2015.08.009.

Montanarella L, Pennock DJ, McKenzie N, Badraoui M, Chude V, Baptista I, et al. World's soils are under threat. SOIL. 2016; 2: 79–82. doi: 10.5194/soil-2-79-2016.

Costanza R, d'Arge R, Groot R de, Farber S, Grasso M, Hannon B, et al. The value of the world´s ecosystem services and natural capital. Nature. 1997; 387: 253-360. doi: 10.1038/387253a0.

Dexter AR, Richard G, Arrouays D, Czyż EA, Jolivet C, Duval O. Complexed organic matter controls soil physical properties. Geoderma. 2008; 144: 620–627. doi: 10.1016/j.geoderma.2008.01.022.

Lal R. Soil carbon sequestration to mitigate climate change. Geoderma. 2004; 123: 1–22. doi: 10.1016/j.geoderma.2004.01.032.

Rawls WJ, Pachepsky YA, Ritchie JC, Sobecki TM, Bloodworth H. Effect of soil organic carbon on soil water retention. Geoderma. 2003; 116: 61–76. doi: 10.1016/S0016-7061(03)00094-6.

Liu X, Trogisch S, Schmid B, He J-S, Bruelheide H, Tang Z, et al. Diversity and stand age increase carbon storage and fluxes in subtropical forests. 2019: submitted.

Lal R. Soil erosion and the global carbon budget. Environment International. 2003; 29: 437–450. doi: 10.1016/S0160-4120(02)00192-7.

Song Z, Seitz S, Li J, Goebes P, Schmidt K, Kühn P, et al. Tree diversity reduced soil erosion by affecting tree canopy and biological soil crust development in a subtropical forest experiment. Forest Ecology and Management. 2019; 444: 69–77. doi: 10.1016/j.foreco.2019.04.015.

Brevik EC, Cerdà A, Mataix-Solera J, Pereg L, Quinton JN, Six J, et al. The interdisciplinary nature of SOIL. SOIL. 2015; 1: 117–129. doi: 10.5194/soil-1-117-2015.

Foley JA, Defries R, Asner GP, Barford C, Bonan G, Carpenter SR, et al. Global consequences of land use. Science. 2005; 309: 570–574. doi: 10.1126/science.1111772.

Minasny B, Malone BP, McBratney AB, Angers DA, Arrouays D, Chambers A, et al. Soil carbon 4 per mille. Geoderma. 2017; 292: 59–86. doi: 10.1016/j.geoderma.2017.01.002.

Jobbagy EG, Jackson RB. The Vertical Distribution of Soil Organic Carbon and Its Relation to Climate and Vegetation. Ecological Applications. 2000; 10: 423. doi: 10.2307/2641104.

Jackson RB, Lajtha K, Crow SE, Hugelius G, Kramer MG, Piñeiro G. The Ecology of Soil Carbon: Pools, Vulnerabilities, and Biotic and Abiotic Controls. Annual Review of Ecology, Evolution, and Systematics. 2017; 48: 419–445.

Hengl T, Jesus JM de, MacMillan RA, Batjes NH, Heuvelink GBM, Ribeiro E, et al. SoilGrids1km--global soil information based on automated mapping. PLoS ONE. 2014; 9: e105992. doi: 10.1371/journal.pone.0105992.

Ibáñez JJ, Ruiz Ramos M, Zinck JA, Brú A. Classical Pedology Questioned and Defended. Eurasian Soil Science. 2005; 38: 75–80.

Grimm R, Behrens T, Märker M, Elsenbeer H. Soil organic carbon concentrations and stocks on Barro Colorado Island — Digital soil mapping using Random Forests analysis. Geoderma. 2008; 146: 102–113. doi: 10.1016/j.geoderma.2008.05.008.

Minasny B, McBratney AB, Malone BP, Wheeler I. Chapter One - Digital Mapping of Soil Carbon. In: Sparks DL, editor. Advances in Agronomy: Academic Press; 2013. pp. 1–47.

Scholten T, Goebes P, Kühn P, Seitz S, Assmann T, Bauhus J, et al. On the combined effect of soil fertility and topography on tree growth in subtropical forest ecosystems—a study from SE China. Journal of Plant Ecology. 2017; 10: 111–127. doi: 10.1093/jpe/rtw065.

Jenny H. Factors of soil formation: A system of quantitative pedology. New York: Dover Publications, Inc.; 1941.

McBratney AB, Mendonça Santos ML, Minasny B. On digital soil mapping. Geoderma. 2003; 117: 3–52. doi: 10.1016/S0016-7061(03)00223-4.

Behrens T, Schmidt K, MacMillan RA, Viscarra Rossel RA. Multiscale contextual spatial modelling with the Gaussian scale space. Geoderma. 2018; 310: 128–137. doi: 10.1016/j.geoderma.2017.09.015.

Behrens T, Schmidt K, Ramirez-Lopez L, Gallant J, Zhu AX, Scholten T. Hyper-scale digital soil mapping and soil formation analysis. Geoderma. 2014; 213: 578–588. doi: 10.1016/j.geoderma.2013.07.031.

Eichenberg D, Pietsch KA, Meister C, Ding W, Yu M, Wirth C. The effect of microclimate on wood decay is indirectly altered by tree species diversity in a litterbag study. Journal of Plant Ecology. 2017; 10: 170–178. doi: 10.1093/jpe/rtw116.

Doetterl S, Berhe AA, Nadeu E, Wang Z, Sommer M, Fiener P. Erosion, deposition and soil carbon: A review of process-level controls, experimental tools and models to address C cycling in dynamic landscapes. Earth-Science Reviews. 2016; 154: 102–122. doi: 10.1016/j.earscirev.2015.12.005.

Pike RJ. The geometric signature: Quantifying landslide-terrain types from digital elevation models. Mathematical Geology. 1988; 20: 491–511. doi: 10.1007/BF00890333.

Piikki K, Wetterlind J, Söderström M, Stenberg B. Three-dimensional digital soil mapping of agricultural fields by integration of multiple proximal sensor data obtained from different sensing methods. Precision Agric. 2015; 16: 29–45. doi: 10.1007/s11119-014-9381-6.

Taghizadeh-Mehrjardi R, Neupane R, Sood K, Kumar S. Artificial bee colony feature selection algorithm combined with machine learning algorithms to predict vertical and lateral distribution of soil organic matter in South Dakota, USA. Carbon Management. 2017; 8: 277–291. doi: 10.1080/17583004.2017.1330593.

Viscarra Rossel RA, Chen C, Grundy MJ, Searle R, Clifford D, Campbell PH. The Australian three-dimensional soil grid: Australia's contribution to the GlobalSoilMap project. Soil Res. 2015; 53: 845. doi: 10.1071/SR14366.

Lacoste M, Minasny B, McBratney AB, Michot D, Viaud V, Walter C. High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape. Geoderma. 2014; 213: 296–311. doi: 10.1016/j.geoderma.2013.07.002.

Malone BP, McBratney AB, Minasny B, Laslett GM. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma. 2009; 154: 138–152. doi: 10.1016/j.geoderma.2009.10.007.

Liu F, Rossiter DG, Song X-D, Zhang G-L, Yang R-M, Zhao Y-G, et al. A similarity-based method for three-dimensional prediction of soil organic matter concentration. Geoderma. 2016; 263: 254–263. doi: 10.1016/j.geoderma.2015.05.013.

Minasny B, McBratney AB, Mendonça Santos ML, Odeh IOA, Guyon B. Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley. Soil Res. 2006; 44: 233. doi: 10.1071/SR05136.

Aldana Jague E, Sommer M, Saby NPA, Cornelis J-T, van Wesemael B, van Oost K. High resolution characterization of the soil organic carbon depth profile in a soil landscape affected by erosion. Soil and Tillage Research. 2016; 156: 185–193. doi: 10.1016/j.still.2015.05.014.

Kempen B, Brus DJ, Stoorvogel JJ. Three-dimensional mapping of soil organic matter content using soil type–specific depth functions. Geoderma. 2011; 162: 107–123. doi: 10.1016/j.geoderma.2011.01.010.

Veronesi F, Corstanje R, Mayr T. Mapping soil compaction in 3D with depth functions. Soil and Tillage Research. 2012; 124: 111–118. doi: 10.1016/j.still.2012.05.009.

Yang Y, Mohammat A, Feng J, Zhou R, Fang J. Storage, patterns and environmental controls of soil organic carbon in China. Biogeochemistry. 2007; 84: 131–141. doi: 10.1007/s10533-007-9109-z.

Brus DJ, Yang R-M, Zhang G-L. Three-dimensional geostatistical modeling of soil organic carbon: A case study in the Qilian Mountains, China. CATENA. 2016; 141: 46–55. doi: 10.1016/j.catena.2016.02.016.

Orton TG, Pringle MJ, Bishop TFA. A one-step approach for modelling and mapping soil properties based on profile data sampled over varying depth intervals. Geoderma. 2016; 262: 174–186. doi: 10.1016/j.geoderma.2015.08.013.

Veronesi F, Corstanje R, Mayr T. Landscape scale estimation of soil carbon stock using 3D modelling. Sci Total Environ. 2014; 487: 578–586. doi: 10.1016/j.scitotenv.2014.02.061.

Liu F, Zhang G-L, Sun Y-J, Zhao Y-G, Li D-C. Mapping the Three-Dimensional Distribution of Soil Organic Matter across a Subtropical Hilly Landscape. Soil Science Society of America Journal. 2013; 77: 1241. doi: 10.2136/sssaj2012.0317.

Chen C, Hu K, Li H, Yun A, Li B. Three-Dimensional Mapping of Soil Organic Carbon by Combining Kriging Method with Profile Depth Function. PLoS ONE. 2015; 10: e0129038. doi: 10.1371/journal.pone.0129038.

Gasch CK, Hengl T, Gräler B, Meyer H, Magney TS, Brown DJ. Spatio-temporal interpolation of soil water, temperature, and electrical conductivity in 3D + T: The Cook Agronomy Farm data set. Spatial Statistics. 2015; 14: 70–90. doi: 10.1016/j.spasta.2015.04.001.

Behrens T, Förster H, Scholten T, Steinrücken U, Spies E-D, Goldschmitt M. Digital soil mapping using artificial neural networks. Journal of Plant Nutrition and Soil Science. 2005; 168: 21–33. doi: 10.1002/jpln.200421414.

Behrens T, Scholten T. A comparison of data-mining techniques in predictive soil mapping. In: Lagacherie P, McBratney AB, Voltz M, editors. Digital soil mapping. An introductory perspective. 1st ed. Amsterdam, Boston: Elsevier; 2007. pp. 353–365.

Bruelheide H, Nadrowski K, Assmann T, Bauhus J, Both S, Buscot F, et al. Designing forest biodiversity experiments: general considerations illustrated by a new large experiment in subtropical China. Methods Ecol Evol. 2014; 5: 74–89. doi: 10.1111/2041-210X.12126.

Seitz S, Goebes P, Song Z, Bruelheide H, Härdtle W, Kühn P, et al. Tree species and functional traits but not species richness affect interrill erosion processes in young subtropical forests. SOIL. 2016; 2: 49–61. doi: 10.5194/soil-2-49-2016.

Yang X, Bauhus J, Both S, Fang T, Härdtle W, Kröber W, et al. Establishment success in a forest biodiversity and ecosystem functioning experiment in subtropical China (BEF-China). Eur J Forest Res. 2013; 132: 593–606. doi: 10.1007/s10342-013-0696-z.

Goebes P, Seitz S, Kühn P, Li Y, Niklaus PA, Oheimb G von, et al. Throughfall kinetic energy in young subtropical forests: Investigation on tree species richness effects and spatial variability. Agricultural and Forest Meteorology. 2015; 213: 148–159. doi: 10.1016/j.agrformet.2015.06.019.

Trogisch S, Schuldt A, Bauhus J, Blum JA, Both S, Buscot F, et al. Toward a methodical framework for comprehensively assessing forest multifunctionality. Ecol Evol. 2017; 7: 10652–10674. doi: 10.1002/ece3.3488.

Orton TG, Pringle MJ, Page KL, Dalal RC, Bishop TFA. Spatial prediction of soil organic carbon stock using a linear model of coregionalisation. Geoderma. 2014; 230-231: 119–130. doi: 10.1016/j.geoderma.2014.04.016.

Krige DG. A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of the Chemical Metallurgical & Mining Society of South Africa. 1951; 52: 119–139.

Conrad O, Bechtel B, Bock M, Dietrich H, Fischer E, Gerlitz L, et al. System for Automated Geoscientific Analyses (SAGA) v. 2.3.1. Geoscientific Model Development; 2015: 1991–2007.

Evans IS. An Integrated System of Terrain Analysis and Slope Mapping. Final Report (Report 6) on Grant DA-ERO-591-73-G0040. Durham: Department of Geography, University of Durham; 1979.

Haralick RM. Ridge and valley detection on digital images. Computer Vision, Graphics and Image Processing. 1983; 22: 29–38.

Horn BKP. Hill shading and the reflectance map. Proc. IEEE. 1981; 69: 14–47. doi: 10.1109/PROC.1981.11918.

Tarboton DG. A new method for the determination of flow directions and upslope areas in grid digital elevation models. Water Resources Management. 1997; 33: 309–319. doi: 10.1029/96WR03137.

Zevenbergen LW, Thorne CR. Quantitative Analysis of Land Surface Topography. Earth Surface Processes and Landforms. 1987; 12: 47–56.

Böhner J, Antonic O. Land-Surface Parameters Specific to Topo-Climatology. In: Hengl T, Reuter HI, editors. Geomorphometry. Concepts, software, applications. 1st ed. Amsterdam Netherlands, Oxford UK, Boston Mass.: Elsevier; 2009. pp. 195–226.

Freeman GT. Calculating catchment area with divergent flow based on a regular grid. Computers & Geosciences. 1991; 17: 413–422.

Moore ID, Grayson RB, Ladson AR. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. Hydrological Processes. 1991; 5: 3–30.

Wischmeier WH, Smith DD. Predicting rainfall erosion losses. A guide to conservation planning. In: United States Department of Agriculture, editor. Agriculture Handbook. Washington D. C.; 1978.

Wood J. The geomorphological characterization of digital elevation models. Dissertation, University of Leicester. 1996. Available: https://lra.le.ac.uk/handle/2381/34503.

Behrens T, Schmidt K, Zhu AX, Scholten T. The ConMap approach for terrain-based digital soil mapping. European Journal of Soil Science. 2010; 61: 133–143. doi: 10.1111/j.1365-2389.2009.01205.x.

Behrens T, Zhu AX, Schmidt K, Scholten T. Multi-scale digital terrain analysis and feature selection for digital soil mapping. Geoderma. 2010; 155: 175–185. doi: 10.1016/j.geoderma.2009.07.010.

Kuhn M, Johnson K. Applied Predictive Modeling. New York, NY: Springer New York; 2013.

R Development Core Team. R: A language and environment for statistical computing. Wien, Austria: R Foundation for Statistical Computing; 2016.

Kuhn M. Building Predictive Models in R Using the caret Package. Journal of Statistical Software, Articles. 2008; 28: 1–26. doi: 10.18637/jss.v028.i05.

Friedman JH. Multivariate adaptive regression splines. The Annals of Statistics. 1991; 19: 1–141.

Breiman L, Friedman JH, Stone CJ, Olshen RA. Classification and Regression Trees. New York, NY: Chapman and Hall; 1984.

Milborrow S. earth: Multivariate Adaptive Regression Splines. Derived from mda:mars by T. Hastie; R. Tibshirani; 2011.

Schmidt K, Behrens T, Daumann J, Ramirez-Lopez L, Werban U, Dietrich P, et al. A comparison of calibration sampling schemes at the field scale. Geoderma. 2014; 232-234: 243–256. doi: 10.1016/j.geoderma.2014.05.013.

Breiman L. Random Forests. Machine Learning. 2001: 5–32.

Díaz-Uriarte R, Alvarez de Andrés S. Gene selection and classification of microarray data using random forest. BMC Bioinformatics. 2006; 7: 3. doi: 10.1186/1471-2105-7-3.

Liaw A, Wiener M. Classification and Regression by randomForest. R News. 2002; 2: 19–22.

Vapnik VN. The Nature of Statistical Learning Theory. New York, NY: Springer; 1995.

Drucker H, Burges CJC, Kaufman L, Smola AJ, Vapnik VN. Support Vector Regression Machines. Advances in Neural Information Processing. 1997; 9: 155–161.

Smola AJ, Schölkopf B. A tutorial on support vector regression. Statistics and Computing. 2004; 14: 199–222. doi: 10.1023/B:STCO.0000035301.49549.88.

Caputo B, Sim K, Furesjo F, Smola AJ. Appearance–Based Object Recognition Using SVMs: Which Kernel Should I Use. Proceedings of NIPS Workshop on Statistical Methods for Computational Experiments in Visual Processing and Computer Vision. 2002.

Karatzoglou A, Smola AJ, Hornik K, Zeileis, A. kernlab - An S4 Package for Kernel Methods in R. Journal of Statistical Software. 2004; 11: 1–20. doi: 10.18637/jss.v011.i09.

Schmidt K, Behrens T, Scholten T. Instance selection and classification tree analysis for large spatial datasets in digital soil mapping. Geoderma. 2008; 146: 138–146. doi: 10.1016/j.geoderma.2008.05.010.

Lin LI-K. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics. 1989; 45: 255. doi: 10.2307/2532051.

Chen Y, Yu S, Liu S, Wang X, Zhang Y, Liu T, et al. Reforestation makes a minor contribution to soil carbon accumulation in the short term: Evidence from four subtropical plantations. Forest Ecology and Management. 2017; 384: 400–405. doi: 10.1016/j.foreco.2016.10.053.

Wang H, Liu S, Wang J, Shi Z, Lu L, Zeng J, et al. Effects of tree species mixture on soil organic carbon stocks and greenhouse gas fluxes in subtropical plantations in China. Forest Ecology and Management. 2013; 300: 4–13. doi: 10.1016/j.foreco.2012.04.005.

Goebes P, Schmidt K, Seitz S, Both S, Bruelheide H, Erfmeier A, et al. The strength of soil-plant interactions under forest is related to a Critical Soil Depth. Sci Rep. 2019; 9: 8635. doi: 10.1038/s41598-019-45156-5.

Minasny B, McBratney AB. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Computers & Geosciences. 2006; 32: 1378–1388. doi: 10.1016/j.cageo.2005.12.009.

Tang X, Xia M, Pérez-Cruzado C, Guan F, Fan S. Spatial distribution of soil organic carbon stock in Moso bamboo forests in subtropical China. Sci Rep. 2017; 7: 42640. doi: 10.1038/srep42640.

Rumpel C, Chabbi A, Marschner B. Carbon Storage and Sequestration in Subsoil Horizons: Knowledge, Gaps and Potentials. In: Lal R, Lorenz K, Hüttl RF, Schneider BU, Braun J von, editors. Recarbonization of the Biosphere. Ecosystems and the Global Carbon Cycle. Dordrecht: Springer; 2012. pp. 444–464.

McBratney AB, Field DJ, Koch A. The dimensions of soil security. Geoderma. 2014; 213: 203–213. doi: 10.1016/j.geoderma.2013.08.013.

Bishop TFA, McBratney AB, Laslett GM. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma. 1999; 91: 27–45. doi: 10.1016/S0016-7061(99)00003-8.

## Appendix

**Manuscript 2**

# 3D Mapping of Soil Organic Carbon Content and Soil Moisture with Multiple Geophysical Sensors and Machine Learning

*Vadose Zone Journal 19, 1.*

Tobias Rentschler[1,2], Ulrike Werban[3]*, Mario Ahner[1], Thorsten Behrens[1,4], Philipp Gries[1,2], Thomas Scholten[1,2,4], Sandra Teuber[1,2], Karsten Schmidt[1,2,4,5]

[1] Dep. of Geosciences, Chair of Soil Science and Geomorphology, Univ. of Tübingen, Rümelinstr. 19-23, Tübingen 72070, Germany
[2] SFB 1070 ResourceCultures, Univ. of Tübingen, Gartenstr. 29, Tübingen 72074, Germany
[3] Dep. of Monitoring and Exploration Technologies, Helmholtz Centre for Environmental Research–UFZ, Permoserstr. 15, Leipzig 04318, Germany
[4] Cluster of Excellence Machine Learning: New Perspectives for Science, Univ. of Tübingen, Maria von Linden Str. 6, Tübingen 72076, Germany
[5] eScience-Center, Univ. of Tübingen, Keplerstr. 2, Tübingen 72074, Germany

*Corresponding author
E-Mail: ulrike.werban@ufz.de

Submitted: May 28, 2020; Accepted: August 12, 2020 Published: September 4, 2020

## Core ideas

- Multi-depth $EC_a$ and gamma-ray spectrometry describe vertical trends of SOC and SM
- Machine learning models can predict vertical trends of SOC and SM spatially
- Cubist models of polynomial depth functions provide accurate 3D maps at field scale

## Abstract

Soil organic carbon (SOC) and soil moisture (SM) affect the agricultural productivity of soils. For sustainable food production, knowledge of the horizontal as well as vertical variability of SOC and SM at field scale is crucial. Machine learning models using depth-related data from multiple electromagnetic induction (EMI) sensors and a gamma-ray spectrometer can provide insights into this variability of SOC and SM. In this work, we applied weighted conditioned Latin Hypercube Sampling to calculate 25 soil representative profile locations based on geophysical measurements on the surveyed agricultural field, for sampling and modelling. Ten additional random profiles were used for independent model validation. Soil samples were taken from four equal depth increments of 15 cm each. These were used to approximate polynomial and exponential functions to reproduce the vertical trends of SOC and SM as soil depth functions. We modelled the function coefficients of the soil depth functions spatially with Cubist and random forests with the geophysical measurements as environmental covariates. The spatial prediction of the depth functions provides 3D maps of the field scale. The main findings are: i) the 3D models of SOC and SM had low errors; ii) the polynomial function provided better results than the exponential function, as the vertical trends of SOC and SM did not decrease uniformly; and iii) the spatial prediction of SOC and SM with Cubist provided slightly lower error than with random forests. Hence, we recommend modelling the second-degree polynomial with Cubist for 3D prediction of SOC and SM at field scale.

## Abbreviations

cLHS, conditioned Latin Hypercube Sampling; DSM, digital soil mapping; $EC_a$, apparent electrical conductivity; EMI, electromagnetic induction; HDP, horizontally oriented magnetic dipole; K, potassium; nRMSE, normalized root mean squared error; RMSE, root mean squared error; SOC, soil organic carbon; SM, soil moisture; Th, Thorium; U, uranium; VDP, vertically oriented magnetic dipole; wecLHS, weighted conditioned Latin Hypercube Sampling with extremes

# Appendix

## Introduction

The lack of knowledge about the resource soil is a major gap in agriculture (Godfray et al., 2010). Soils play a key role in sustainable agriculture (Bouma, 2014; Bouma and McBratney, 2013) and, thus, food production. Soil management practices are directly linked to the spatial knowledge about soil properties and conditions that are relevant indicators for more efficient and effective agriculture. As one of the key soil properties soil organic carbon content (SOC) is relevant for soil quality and fertility, as it influences the soil's nutrient availability and structural stability (Dexter et al., 2008). In combination with soil texture, SOC affects the soil water holding capacity, plant available water, and soil moisture (SM; Rawls et al., 2003). Understanding the spatial as well as vertical variability of SOC and SM is essential for plant cultivation, which requires fertile soils and sufficient water. Digital soil mapping (DSM) can provide high resolution information for sustainable agricultural management to facilitate food production on the field and farm scale through the spatial prediction of physical and chemical soil properties (Govers et al., 2017), such as SOC and SM. This facilitates an improvement in the decision-making processes for fertilization, irrigation, and liming amongst others, and subsequently higher productivity of food and biofuels (McBratney et al., 2005).

However, soil properties vary in the horizontal as well as in the vertical domain. Hengl et al. (2014) and Viscarra Rossel et al. (2015) mapped soil properties in multiple depths, which can be interpreted three-dimensionally, but do not actually provide continuous 3D information (Liu et al., 2016). To fully grasp the continuous character of soil, the third dimension should be included into the analysis as continuous entity, e.g. by incorporating mathematical functions that represent the vertical distribution of soil properties, so-called soil depth functions (Aldana Jague et al., 2016; Minasny et al., 2006; Rentschler et al., 2019; Veronesi et al., 2014). The spatial distribution of soil depth functions is related to the spatial distribution of environmental covariates given by the soil forming equation (Jenny, 1941; McBratney et al., 2003):

$$S = f(s, c, o, r, p, a, n)$$

where S is the soil or any soil information we would like to explain, s stands for other available soil properties at a location, c is climate, o is organisms, the factor r is the terrain, p is the parent material, a is age and n is the spatial position. The function coefficients of the soil depth functions can be treated as abstract soil property and, therefore, modelled and predicted spatially based on comprehensive sets of environmental covariates delineated from digital elevation models representing terrain (Aldana Jague et al., 2016; Minasny et al., 2006; Rentschler et al., 2019; Veronesi et al., 2014), provided by land cover maps representing organisms (Minasny et al., 2006; Veronesi et al., 2014) and gamma-ray sensing data as

indicator for soil-forming minerals (Aldana Jague et al., 2016; Cook et al., 1996; Minasny et al., 2006). Besides gamma-ray spectrometry, hydrogeophysical methods, such as electromagnetic induction (EMI) sensors, provide a widely used base in DSM in general (Binley et al., 2015; Cassiani et al., 2012; Martini et al., 2017) as well as in 3D soil mapping specifically (Moghadas et al., 2016), complementing field sampling by generating high-resolution spatial geophysical covariates. Especially geophysical sensing technologies and measurements are urgently needed at the field scale, as the distribution of soil properties (S) in the vertical domain are decreasingly linked to terrain (r) and climate (c) variability but more prone to variations in weathering, mineralogy (as parts of s and p), biological activities as well as past anthropogenic influences (o and a; Jobbágy and Jackson, 2000; Rentschler et al., 2019).

Apparent electrical conductivity ($EC_a$) from EMI sensors and dose rate, $^{40}K$, $^{238}U$ and $^{232}Th$ contents from gamma-ray spectrometers are covariates that are closely linked to numerous soil properties, such as texture, horizonation, bulk density, SOC and SM in case of EMI (Cho et al., 2016; Doolittle and Brevik, 2014; Martini et al., 2017) and in case of gamma-ray spectrometry texture and SOC. Thus, gamma-ray spectrometry and EMI are used either individually or combined (Castrignanò et al., 2012) as a proxy to the mineralogy of the parent material and other soil properties developed or inherited from the parent material (Cook et al., 1996; Jenny, 1941; McBratney et al., 2003). The geophysical measurements are interpolated with geostatistical methods like kriging (Krige, 1951) to obtain spatial information of $EC_a$, $^{40}K$, $^{238}U$ and $^{232}Th$ covering the whole field with high spatial resolution (Abdu et al., 2008; Schmidt et al., 2014). The interpolations of the geophysical measurements constitute the covariate space of the agricultural field, which is utilised in crucial modules of DSM:

- The spatial data of $^{40}K$, $^{238}U$ and $^{232}Th$ and multi-depth $EC_a$ are used to calculate the locations of a representative sampling scheme for soil sampling. The aim is to fully cover potential soil variability that influences the modelled soil property and that is found on the field at the time of measurement. For that, many approaches use conditioned Latin Hypercube Sampling (cLHS) or extensions like weighted extreme cLHS (Minasny and McBratney, 2006; Schmidt et al., 2014). The cLHS is a stratified random sampling design that provides an optimal stratification of a covariate space with a reduced number of spatially distinct sample sites (Minasny and McBratney, 2006).
- The sampled point wise soil data, i.e. SOC and SM in this work, is linked to the geophysical measurements at the soil profile locations with linear or non-linear machine learning models (Aldana Jague et al., 2016; Rentschler et al., 2019; Schmidt et al., 2014). The dependent variable of the models are the samples measured at the locations introduced above and the independent variables are the interpolated geophysical measurements at these locations.

- The model trained with sampled soil data and geophysical measurements is used to make predictions to the locations of the covariate space where no soil samples were taken.
- The model predictions are validated with additional soil samples. Ideally, the sampling scheme used for validation is independent from the scheme used for model training, which can be achieved with randomly distributed samples (Brus et al., 2011; Steyerberg and Harrell, 2016).

Based on the modules of DSM described above and the potential to measure 3D soil properties provided by geophysical measurements (other available soil properties s in the soil forming equation), we assume that EMI and gamma-ray spectroscopy are highly suitable for the spatial prediction of soil depth functions, as the combination of multi-coil EMI and gamma-ray spectrometry provides multiple penetration depths and a different sensitivity to soil parameters (Dierke and Werban, 2013; McNeill, 1980a, 1980b). To our knowledge, there are no studies on spatial prediction of soil depth functions with EMI and gamma-ray sensing data as environmental covariates.

The major objective of this study was the prediction of SOC and SM in 3D using soil depth functions based on EMI data from sensors with 12 different penetration depths and gamma-ray spectrometry by capturing the response of the parent material and overlaying soil. For model training, we used Cubist and random forests, two machine learning methods often used in DSM. The hypothesis is that for 3D modelling of SOC and SM, data from EMI and gamma-ray spectrometry will achieve low errors throughout the sampled depth increments, due to the different depth penetration of the sensors.

## Material and methods

### Study site

The study site is an agricultural field of 58 ha about 70 km north of Leipzig, Saxony, Germany (Fig. 1). The field is located on the Elbe flood plain and bordered by the creeks Altes Flieth and Fließgraben. There is no visible terrain variation in the field. Present soil types are Gleysols and Gleyic Cambisols consisting of alluvial loam (loam and clay) over Holocene sediments of fluvial sand (LAGB, 2014). At the time of sampling in August 2017, the cultivated wheat had been harvested, and the field was bare.

### Methodological overview

The workflow consisted of seven individual steps (Fig. 2). First, the geophysical measurements were taken with EMI and gamma-ray spectrometry and, subsequently, interpolated geostatistically with ordinary kriging (Krige, 1951) to receive spatial predictions of the

environmental covariates. These covariates were used to calculate representative sampling locations with an extension of conditioned Latin Hypercube Sampling (cLHS; Minasny and McBratney, 2006) and served as independent variables for the soil depth function modelling. In the next step these models were applied for spatial prediction of the soil depth functions with the independent variables and validated independently in the final step. The following subsections describe this workflow in detail.

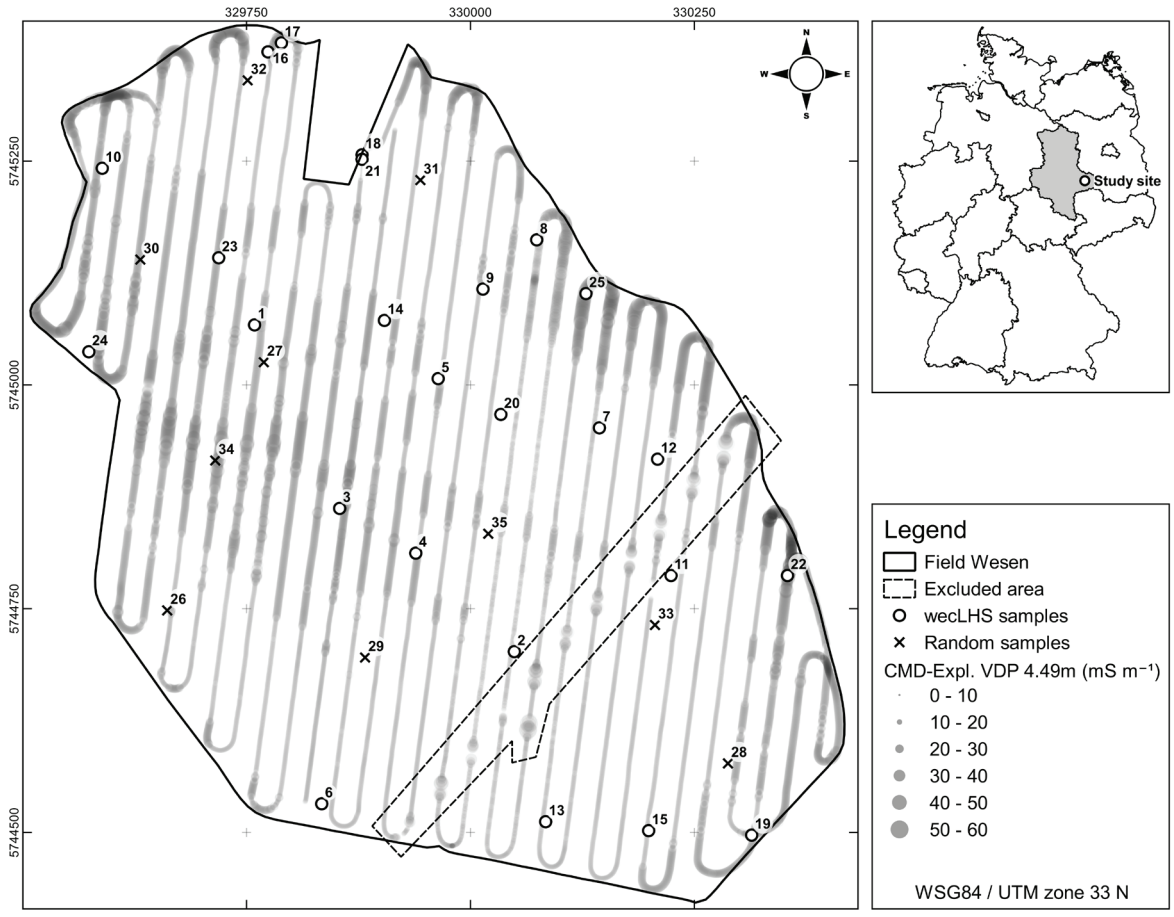


**Fig. 1.** Location of the field Wesen near Selbitz, Saxony-Anhalt, Germany, sampling scheme of the geophysical measurements with EMI and gamma-ray spectrometry, and sampled soil profiles (circle – wecLHS samples for calibration, cross – random samples for validation). The signal of the CMD-Explorer with 4.49 m intercoil spacing and vertical dipole orientation (VDP) was noisy due to a grid gas pipe and, therefore, measurements of all sensors were omitted in the marked area.

### Geophysical measurements and interpolation

The geophysical measurements were recorded with two electromagnetic induction (EMI) sensors (CMD-Explorer and CMD-Mini-Explorer, both GF Instruments, CZ) and a gamma-ray spectrometer (GMS CAR, GF Instruments, CZ) in August 2016. EMI sensors measure the apparent electric conductivity ($EC_a$ in mS $m^{-1}$). The penetration depth of the magnetic field

is mainly controlled by the intercoil spacing and the orientation of the dipoles as well as the applied frequency. The penetration depth and footprint of the sensor data increases with increasing intercoil spacing. Vertically oriented magnetic dipoles (VDP/coil axis horizontal coplanar - HCP) provide a higher depth penetration than horizontally oriented magnetic dipoles (HDP/coil axis vertical coplanar – VCP) while taking into account the different cumulative sensitivity functions of both orientations (Callegary et al., 2012; Martini et al., 2017; McNeill, 1980b; von Hebel et al., 2019). The CMD-Explorer and the CMD-Mini-Explorer enable simultaneous multi depth-exploration of $EC_a$ with either VDP or HDP. The instruments have one transmitter and three receiver coils with different intercoil spacings covering six effective penetration depths, which is defined by the manufacturer (GF Instruments) as the depth above which 70 % of the signal comes from (Table 1).This multi-sensor setup measuring $EC_a$ and penetration depths of up to 6.7 m enables to detect textural patterns of the spatially variable subsurface sediments of the Elbe floodplain indirectly (Doolittle and Brevik, 2014). This is important for SM modelling to account for subsurface sediment structures, as gravel lenses with high permeability can drain and clay lenses can retain percolate water or facilitate capillary water rise (Abdu et al., 2008).

The sledge-mounted devices (height of CMD-Explorer 80 cm, height of CMD-Mini-Explorer 10 cm) were towed by a four-wheel vehicle at less than 10 km $h^{-1}$, crossing the field in multiple parallel (track distance 27 m) and a few crossed transects. By using overlapping measurements collected at different time from crossing the field in the end, drifts in the data were assessed and a linear drift function was applied to correct the data (Martini et al., 2017). Before interpolation these quality control lines as well as outliers related to a gas pipe line were removed (Fig. 1). Within the data set negative values of $EC_a$ occurred due to the custom calibration of the instrument (von Hebel et al., 2019). We corrected the measurements with an offset of 3.44 mS $m^{-1}$ (CMD-Mini-Explorer VDP 0.32 m), 4.86 mS $m^{-1}$ (CMD-Mini-Explorer HDP 0.32 m) and 0.21 mS $m^{-1}$ (CMD-Mini-Explorer HDP 0.71 cm) to avoid confusion with these values and to make use of the containing information on spatial variability. Smoothing of the data wasn't necessary due to low noise conditions. All EMI sensors captured 5 records $s^{-1}$ in any dipole orientation. We refrained from inverting EMI data, since the reliability of the required calibration procedure is limited due to a number of fundamental issues which are not solved yet (Martini et al., 2017).

The bulk (~90 %) of aboveground measured gamma radiation is emitted in the top 30 to 50 cm of soil (Cook et al., 1996). We used a gamma-ray spectrometer with a 4 l NaI(Tl)-crystal and automatic peak-stabilization to measure the concentration of potassium ($^{40}K$), uranium ($^{238}U$) and thorium ($^{232}Th$). The device has 512 channels with an energy range from 100 keV to 3 MeV. Measurements were captured every 5 seconds. $^{40}K$, $^{238}U$ and $^{232}Th$ were measured

as counts per second. The concentration of $^{40}$K (in %) and $^{238}$U and $^{232}$Th (both in ppm) was calculated corresponding to the decay rate at specific energy levels. The concentration of $^{40}$K, $^{238}$U and $^{232}$Th was used to calculate the dose rate (Gy h$^{-1}$; IAEA, 2003).

The geophysical measurements served as a basis for the sampling design (Fig. 1) and were interpolated to a grid cell size of 5 m with ordinary kriging (Krige, 1951) using individual exponential semi-variogram functions for each data set in the gstat package version 1.1-5 (Pebesma, 2004) in R version 3.4.3 (R Development Core Team, 2017). Beforehand, measurements within 1 m range were averaged. Noisy measurements along a straight line were detected for the CMD-Explorer with the higher depth penetrations, caused by an underground grid gas pipe (Fig. 1). For reasons of continuity, all measurements from the EMI-sensors and the gamma-ray spectrometer in this area were excluded from further processing. This crucial step is to be evaluated carefully since all consecutive steps strongly depend on accurate environmental covariates. Therefore, error of the kriging predictions was assessed with a 10-fold cross-validation, which is an out-of-sample testing method to assess the ability of the model to generalize to independent data subsets. For that the dataset is partitioned in 10 folds of nearly equal size, where 9 folds are used to train a model and tested with the 10th fold. This is done 10 times to test all folds and the quality measure is the average of all models. For the final model, all folds are used.

**Soil sampling**

For the estimation of the number of soil profiles to sample, the areas under curve for the empirical cumulative distribution functions were calculated for the proposed sampling set sizes of n=10, 20, 25, 30, 35 and 40 and the geophysical measurements with the *MESS* package version 0.5.0 (Ekstrøm, 2018) in R. The mean of the differences between the areas under curve indicated the error for each sample set size and the sample set size with the lowest error is chosen (Ramirez-Lopez et al., 2014; Schmidt et al., 2014). In this case, the optimal sample set size with the lowest error was 35. However, due to costs and feasibility constraints we agreed on a sample set size of 25 for the representative sampling design as trade-off between feasibility and a slight increase in model error.

Spatial soil modelling requires specific sampling schemes or designs for sampling and validation (Brus et al., 2011; Schmidt et al., 2014). Aim of a sampling design is to cover the full range of potential driving factors that influence the modelled soil property and that are found on the field at the time of measurement while reducing soil sampling effort and analytical costs. Therefore, we calculated the locations of the soil profiles to sample for model training with a weighted conditioned Latin Hypercube Sampling with extreme values (wecLHS; Schmidt et al., 2014) based on the geophysical covariates with the lowest error in cross-validation of each sensor, to obtain representative sampling locations. The wecLHS

extends the cLHS (Minasny and McBratney, 2006) by including samples from the extrema of the used covariate space to cover the full range of data. Further, a weighting scheme according to the explained variation ($R^2$) of the kriging predictions is implemented to account for noise in the interpolation (Schmidt et al., 2014). The wecLHS design was calculated with 150,000 iterations. The settings were a temperature decrease of 0.95, an initial temperature of 1, optimization every 10 steps and an initial Metropolis value of 1.

Additionally, we sampled 10 fully randomly distributed profile locations for independent model validation. We chose a fully random sampling design for validation, since wecLHS is a stratified random sampling design and for independent validation a non-stratified sampling strategy is recommended. Further, no assumptions regarding the standard error of the estimated quality measures are required (Brus et al., 2011; Steyerberg and Harrell, 2016). The locations of the sampled profiles are displayed in Fig 1.

The soil profiles were sampled from four equal depth increments to 60 cm depth (0-15, 15-30, 30-45, 45-60 cm) with a hand auger on two days with the same weather conditions in August 2017 again after harvest under similar field conditions as during sensing. 60 cm is the depth above which about 80 % of the roots of many agricultural crops are found (Fan et al., 2016). Samples were taken for each depth increment as mixed subsamples from two corners and the centre of 1 $m^2$ (Schmidt et al., 2014), resulting in 100 samples for the training set and 40 samples for the validation set. The positions of the profiles were located with a differential GPS (Leica TPS1200+, Leica Geosystems, CH).

**Laboratory analysis**

For SOC determination, the samples were dried at 40 ℃ for 24 h, sieved (<2 mm), ground and root fragments were removed. Total carbon was determined with dry combustion using an ELTRA CHS-580A Helios analyser (ELTRA GmbH, GER). Although LAGB LSA (2014) states that soils in the flood plains of the Elbe river are mostly free of carbonates, pH of the samples ranged from 5.2 to 7.2. Consequently, inorganic carbon was determined gravimetrically with 10 % HCl solution. Then, SOC was determined as difference between total carbon and inorganic carbon.

SM was measured gravimetrically with drying at 90 ℃ for 24 h. A summary of the training and validation sets for SOC and SM is shown in Fig. 3. Both, training and validation sets were similar. The rather small differences are due to the small sampling set size of the validation set and its sensitivity to extreme values because of its random and non-stratifying nature.

## Model training, prediction and validation

### *Soil depth functions and 3D predictions*

For the 3D modelling of SOC and SM we tested a 2nd degree polynomial (equation [1]; Aldana Jague et al., 2016) and an exponential function (equation [2]; Rentschler et al., 2019).

$$f(x) = c_0 + c_1 x + c_2 x^2 \quad [1]$$

$$f(x) = exp^{c_1 + c_2 x} \quad [2]$$

where $c_0$ is the intercept with the y-axis, thus the SOC and SM at the surface, and $c_1$ and $c_2$ are dimensionless coefficients.

The functions coefficients $c_0$, $c_1$ and $c_2$ were modelled and predicted for the whole study site based on the geophysical data of the EMI sensor and gamma-ray spectrometer with Cubist and random forests. After modelling and spatial prediction of the coefficients of the soil depth functions, the respective function can be solved at every grid location of the study site and SOC and SM can be calculated with any vertical resolution (Aldana Jague et al., 2016; Liu et al., 2016; Veronesi et al., 2014). However, vertical resolution is limited by the vertical sampling of each profile that reflects the vertical variation within each profile. In this work, the soil depth functions were solved from 0 to 60 cm with a vertical resolution of 5 cm. The main advantage of this approach is that solving the soil depth functions provides data points that represent a three-dimensional entity (voxels) of the response variables instead of two-dimensional pixels. The voxels were stored in an array with the dimensions of the study area in the horizontal domain (Rentschler et al., 2019). A workflow diagram is given in Fig. 2.

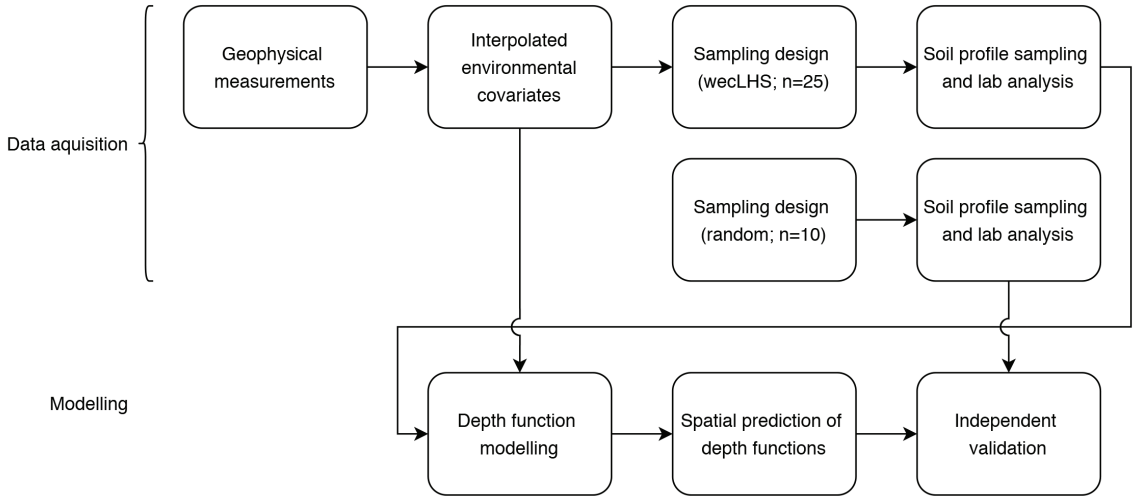


**Fig. 2.** Workflow diagram illustrating the individual working steps of this study.

## Appendix

### *Supervised machine learning*

In many applications in DSM supervised machine learning is used to train regression models with numeric values, such as the coefficients of soil depth functions. In such models, the function coefficients at each sampling location are the dependent variable (soil property S in the soil forming equation) and the geophysical measurements compose the covariate space of independent variables, i.e. $EC_a$ in the measured depth intervals, $^{40}K$, $^{238}U$ and $^{232}Th$, at the location of each sample for regression:

$$S = f(EC_a, {}^{40}K, {}^{238}U, {}^{232}Th)$$

Subsequently, the model can predict the dependent variable for each grid location at the field, since the independent variables were measured and interpolated onto the whole study area. Common supervised machine learning methods in DSM are Cubist and random forests.

Cubist uses a robust system called M5 model tree, which was established by Quinlan (1992). It applies a recursive partitioning of data to build a piecewise linear model as a decision tree, where the terminal nodes are linear models. When growing the tree, intra-subset variation is minimized at each split. A leaf of such a tree applied on continuous data contains a linear model connecting the values of the training cases to their attribute values. The procedure is based on building and applying rules. The rules generate subsets of the data according to similar characteristics of predictor and response variables. The rules are structured as: if [condition is true], then [regress], else [apply next rule], comprising single or multiple predictor variables. With the rules that fulfil the conditions, soil properties are predicted by ordinary least-squares regression. If the rule does not apply, a new rule is processed for the next node of the tree within an iterative process. These rulesets are appropriate for model interpretation (Quinlan, 1992, 1993).

Random forests were developed by Breiman (2001) as an ensemble of classification and regression trees (Breiman et al., 1984). Binary splits are used for a single decision tree to homogenize the predictor variables according to the dependent variable, thus minimizing the node impurity. Random forests use a bootstrap approach, where random predictor variables are chosen at each split of a tree. The final regression model results from averaging all decision tree outputs (Breiman, 2001). Random forests are robust against overfitting and interpretable with the feature importance calculated from its randomly permutated trees (Breiman, 2001). However, this is beyond the scope of this study as it requires more detailed analysis of the depth functions as dependent variables.

The tuning parameters for the machine learning methods used were the number of subsequently adjusted trees committing to the final decision tree (committees) and the number of neighbouring samples from the training set to adjust the model prediction (neighbours) for Cubist. The number of randomly selected covariates at each split (mtry) was used for tuning

of random forests. The number of trees (ntree) and the node size of random forests were set to default as this is not necessary when a large number of trees is computational manageable (Probst and Boulesteix, 2018). To find the best tuning parameters for the models, a grid search (Schmidt et al., 2008) with a 10 times repeated 10-fold cross validation was applied. The final models were calibrated with the tuning parameters of the models with the lowest RMSE.

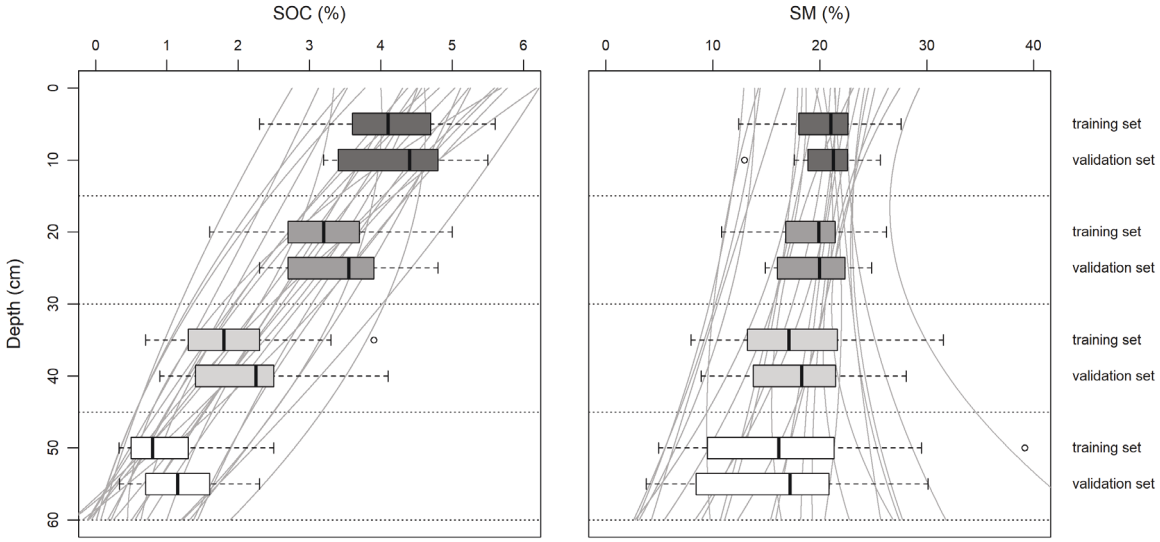


**Fig. 3.** Summary of the training and validation sets for soil organic C (SOC) and soil moisture (SM) as boxplots and the respective polynomial soil depth functions. The boxplots show the variation of the samples within the sampled depth increments. Training and validation sets have a similar range at each sampled depth increment and the validation set is suitable for model evaluation. The polynomial soil depth functions show the vertical distribution at the 25 profiles of the training set

### *Model validation*

The 3D predictions are validated independently with the coefficient of determination ($R^2$; equation [3]) as measure of correlation between the observed and predicted values, Lin's concordance correlation coefficient (CCC; equation [4]), the root mean squared error (RMSE; equation [5]) and the normalized root mean squared error (nRMSE; equation [6]) as measures of error with the 10 random profiles of 4 samples each taken at 0-15, 15-30, 30-45, 45-60 cm. The CCC is a measure of concordance of the model predictions and the measured values on the 1:1 line from the origin. The RMSE is a measure of error, which allows to compare models with observed values of the same magnitude. Since the response of SOC and SM has observed values of different range, the nRMSE is required to compare the models for each depth increment.

## Appendix

The equations for $R^2$, the CCC, the RMSE and nRMSE are:

$$R^2 = \left( \frac{\sum_{i=1}^{n}(y-\mu_y)(\hat{y}-\mu_{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y-\mu_y)^2}\sqrt{\sum_{i=1}^{n}(\hat{y}-\mu_{\hat{y}})^2}} \right)^2 \quad [3]$$

$$CCC = \frac{2\rho\sigma_y\sigma_{\hat{y}}}{{\sigma_y}^2 + {\sigma_{\hat{y}}}^2 + (\mu_y - \mu_{\hat{y}})^2} \quad [4]$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad [5]$$

$$nRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{y_{max} - y_{min}} \quad [6]$$

where y and $\hat{y}$ are the observed and predicted values and $\mu_y$ and $\mu_{\hat{y}}$ denote the means for the observed and predicted values, respectively; $\rho$ is the correlation coefficient (Pearson's r), $\sigma_y$ and $\sigma_{\hat{y}}$ are the corresponding variances; $y_{max}$ is the maximum of the observed values and $y_{min}$ is the minimum of the observed values.

## Results and Discussion

### Geophysical measurements and interpolation

The results of the 10-fold cross-validation for the geostatistical interpolation showed a high coefficient of determination ($R^2$) between the observed and the predicted values for all EMI sensors ($R^2$>0.96) and low errors (nRMSE≤0.08). The predictions for the CMD-Mini-Explorer with intercoil spacings of 0.71 and 1.18 m and vertically oriented magnetic dipoles had the highest coefficient of determination ($R^2$=0.99 and nRMSE=0.01) and the predictions for the CMD-Explorer with an intercoil spacing of 4.49 m and horizontally oriented dipole had the lowest coefficient of determination ($R^2$=0.96 and nRMSE=0.04).

The cross-validation results of the gamma-ray spectrometer showed coefficients of determination ranging from 0.75 ($^{238}$U) up to 0.92 (dose rate). The lower $R^2$ compared to EMI sensor interpolation is to be expected due to the noise prone passive nature of statistical counting gamma decays. The errors of the interpolation range from 0.05 (dose rate) to 0.07 ($^{238}$U). All results of the 10-fold cross-validation are shown in Table 2.

The spatial variation of the measured $EC_a$ varies between the sensors and sensor orientation. The highest values were measured with the CMD-Mini-Explorer VDP (0.71m; Table 2, Fig. 4b) and the lowest values were measured with the CMD-Mini-Explorer HDP (0.32m; Fig. 4d). The measures with the CMD-Mini-Explorer in the same orientation (VDP and HDP) showed considerable changes in $EC_a$ (Fig. 4a-c and d-f) with increasing intercoil spacing, while the measurements with CMD-Explorer were more alike (Fig. 4g-l). However, we used all EMI measures as independent variables, since the similar depths of investigation may contain

varying information while using different coil orientation resulting in different shapes of their sensitivity functions. The interpolations of $^{40}$K, $^{232}$Th, $^{238}$U and dose rate (Fig. 5) showed different spatial trends with some shared local minima in the centre and the South-East and a shared local maximum in the West of the field. All results of the interpolations are visualised in Figs. 4 and 5.

Given the low error (nRMSE≤8 %) of the interpolation for the EMI data and for gamma-ray measurements, the environmental covariates interpolated with ordinary kriging represented the spatial distribution of $EC_a$ at multiple depths, $^{232}$Th, $^{238}$U and dose rate adequately. Thus, they were suitable to calculate a wecLHS sampling design as well as for modelling and mapping of SOC and SM in the horizontal and vertical domain. The covariates with low cross-validation error and high coverage of the covariate space used for wecLHS were dose rate for the gamma-ray spectrometer and CMD-Mini-Explorer with VDP (0.32 m intercoil spacing), CMD-Mini-Explorer with HDP (0.32 m) and CMD-Explorer with VDP (4.49 m) for the EMI sensors.

**Soil depth functions**

The fitted soil depth functions showed the vertical trend of SOC and SM at the profile locations (Fig. 3). For both polynomial and exponential functions at each profile, the $R^2$ and RMSE values of the soil depth functions were calculated. The soil depth function with the highest $R^2$ for SOC and SM was the polynomial function (eq. 1) with a mean $R^2$ of 0.98 for SOC and 0.92 for SM (RMSE=0.14 and 0.00). The exponential soil depth function had lower mean $R^2$ and a higher error for both SOC and SM. A summary of the evaluation of the soil depth functions is shown in Table 3.

The minimum values in $R^2$ and the high standard deviation of the soil depth functions for SM showed that the soil depth functions could not depict the vertical trend in some soil profiles (Table 3). This is the case for the profiles 2 and 10, where the SM was 0.22, 0.20, 0.23 and 0.21 and 0.25, 0.21, 0.22 and 0.24, respectively. These profiles had a local maximum or minimum between 15 and 45 cm that could not be modelled with the exponential function, which cannot be explained with the geophysical measurements or additional knowledge about the field. In both cases, the polynomial function of second degree had a lower error than for all other profiles. To solve this, other functions are required that can reproduce vertical distributions with local minima and maxima (minimax; Minasny et al., 2016).

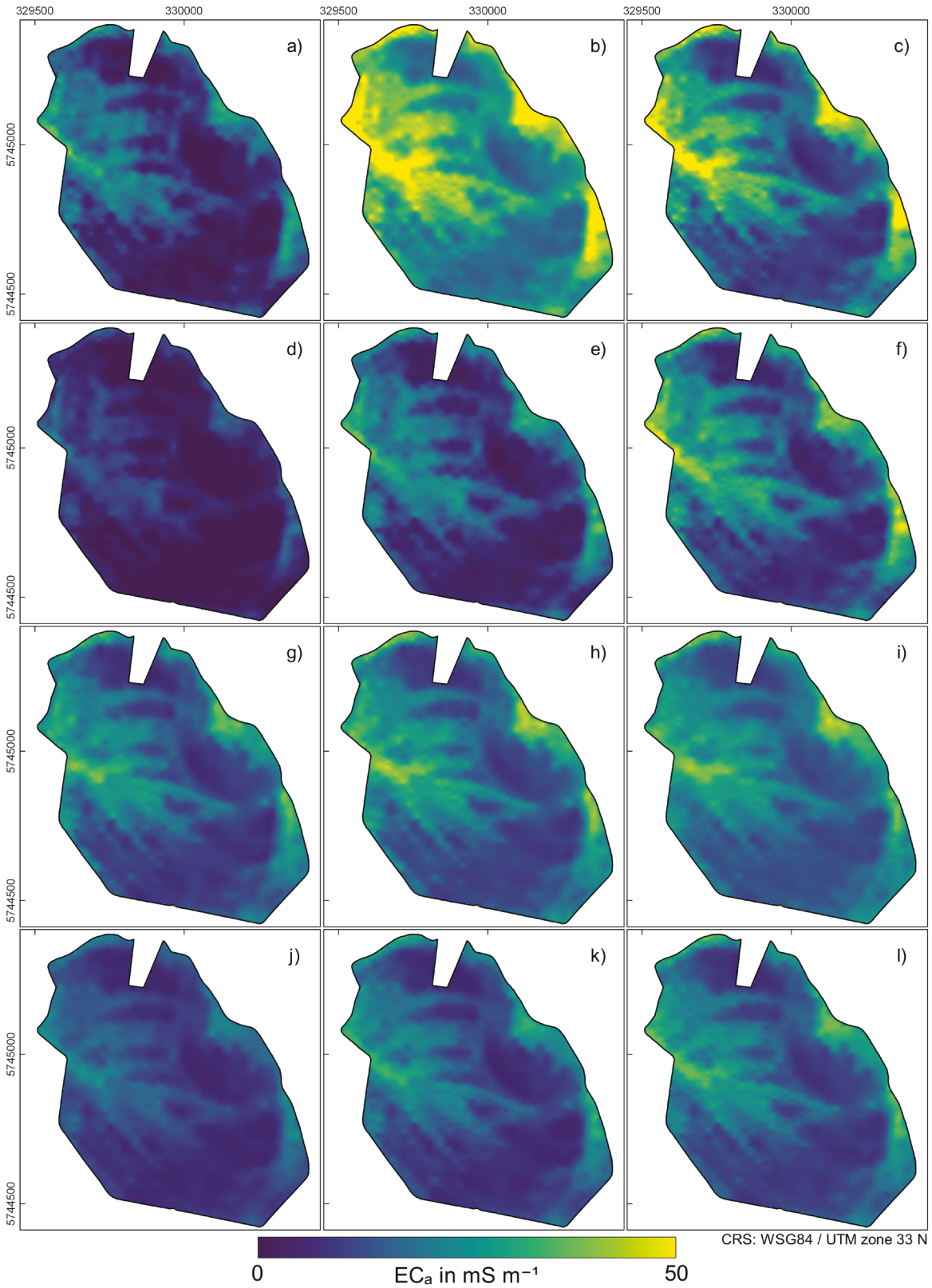


**Fig. 4.** Results of the interpolation with ordinary kriging for $EC_a$ measured with CMD-Mini-Explorer VDP (a-c), CMD-Mini-Explorer HDP (d-f), CMD-Explorer VDP (g-i) and CMD-Explorer HDP (j-l) in order of increasing effective depth range.

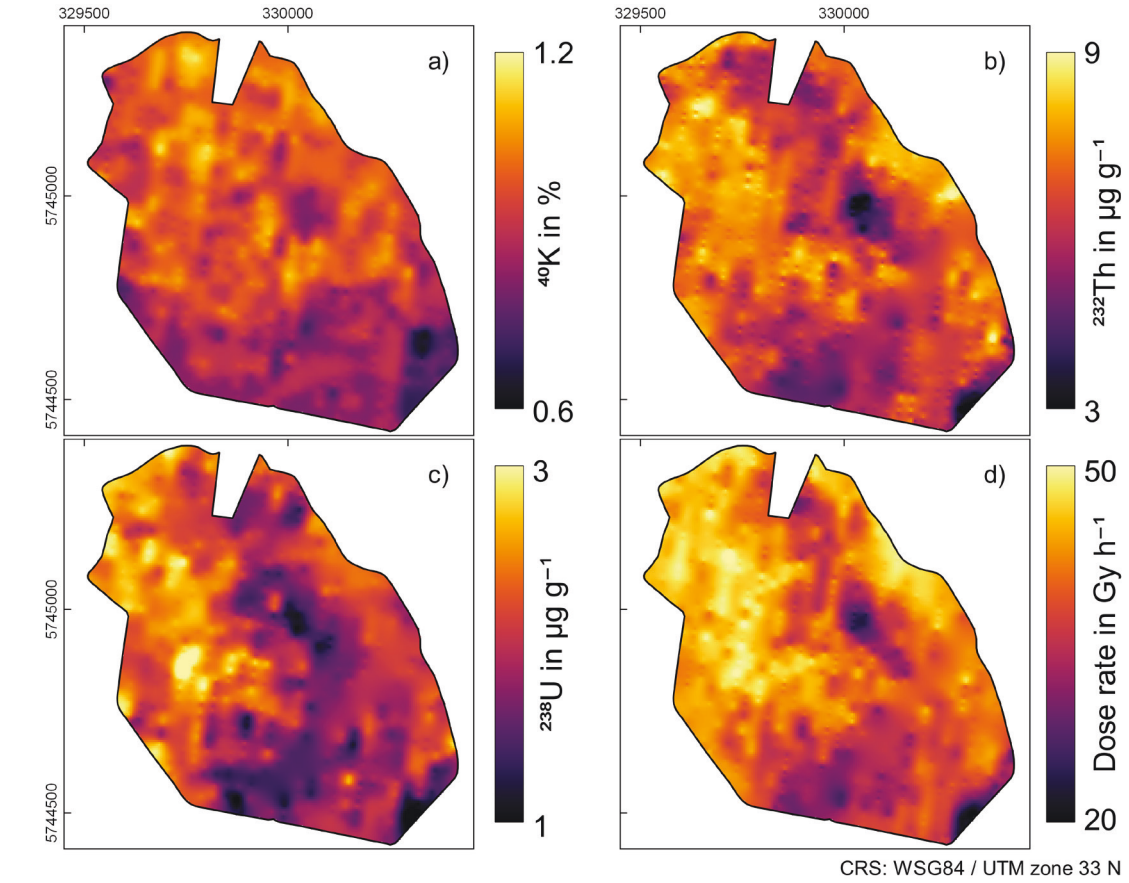


**Fig. 5.** Results of the interpolation with ordinary kriging for $^{40}$K (a), $^{232}$Th (b), $^{238}$U (c) and dose rate (d).

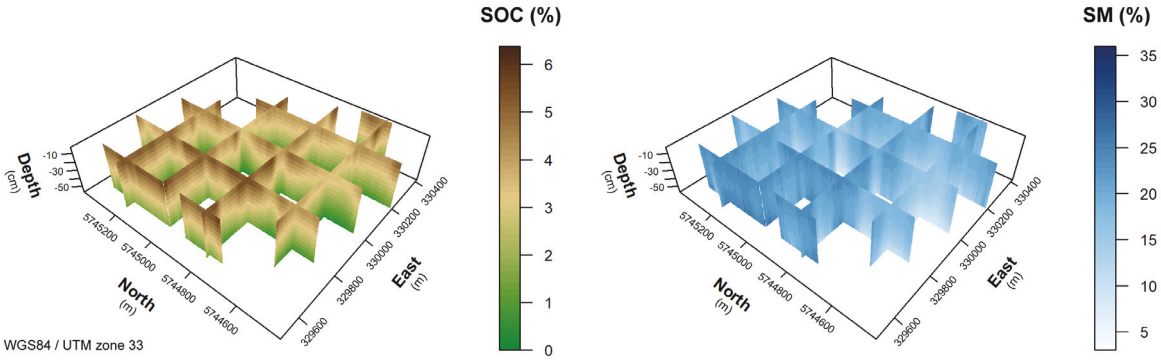


**Fig. 6.** 3D predictions of SOC and SM with polynomial depth function and Cubist (see supplements 1 to 4 for detailed animated cross sections).

### 3D predictions

The independent validation with the validation set of 10 randomly located profiles with the same sampled depth increments showed a high overall explained variation and low errors for both polynomial and exponential functions and for both Cubist and random forests. The $R^2$ of Cubist models for SOC and for SM was 0.86 and 0.88 with the polynomial function and

0.86 and 0.87 with the exponential function (Table 4). Random forest models showed a similar $R^2$ except for SM with the polynomial function ($R^2$=0.84). The CCC of all Cubist models was slightly higher than the CCC of the random forests models. Cubist models had a CCC of 0.91 for SOC and 0.91 for SM with polynomial function. The CCC for SM with exponential function was slightly higher (CCC=0.93). The error of all four models for SM was identical (RMSE=0.02). For SOC, the error of the Cubist and random forests models with polynomial function was about 0.02-0.03 % lower than the error of models with exponential function. The difference in RMSE between Cubist and random forests was less than 4 % (Table 4).

Because of the higher $R^2$ and CCC and the lower error of the random forest model with the polynomial function this was chosen for SOC modelling. Cubist in combination with the polynomial function was chosen for the final SM predictions. However, it is worth noting that the differences in error as well as $R^2$ and CCC of all models were rather small.

Comparing the model coefficient of determination, CCC and RMSE by individual depth increment showed a similar pattern. Cubist models with polynomial function for SOC had a higher $R^2$ and CCC and a lower RMSE than Cubist with exponential function and random forests with both polynomial and exponential functions. For SM, the differentiation by $R^2$ was not clear as the $R^2$ varied between the used machine learning methods. However, the Cubist models with polynomial functions also showed the lowest RMSE for soil depth functions and modelled depth of the soil (Table 5).

More important is that the nRMSE of Cubist and random forests models with the polynomial function did vary least with depth in absolute values. This showed that the model could predict SOC and SM with low error throughout the sampled interval of the soil profile (Table 5). Both Cubist and random forests with the exponential function for SOC and SM had high errors (nRMSE) in the depth increment ranging from 15 to 30 cm. Therefore, we conclude that the exponential function could not depict the vertical trend of SOC and to a lesser extent of SM within the sampled profile. We ascribe this to the 30 cm deep plough horizon, which needs to be accounted for with a less uniformly decreasing soil depth function. The flexibility of polynomial functions with third degree or higher is potentially capable of depicting local variations in the soil better than exponential functions. We recommend to be investigated in more detail.

Further, the $R^2$ of the SOC models decreased from around 0.80 to 0.50 on average with increasing depth and increased from 0.76 to 0.88 with increasing depth for SM. Lin's CCC showed a similar pattern. These differences in explained variability indicate differences in explanatory power of the geophysical measurements for SOC and SM modelling for the depth intervals used in this study. On the one hand geophysical measurements and especially EMI measurements are influenced by SM, whereas SOC content is related indirectly through

SM content and influenced by other soil and environmental processes. This may not be covered by EMI and gamma-ray sensors, where covariates of the latter have little influence on depths >30 cm. On the other hand, this may refer to the decreasing range of SOC content and the increasing range of SM (Fig. 3). The lower SOC content variability in depth may not be represented by the covariates. Therefore, these complex interactions need be investigated in more detail to make more precise conclusions about the use of geophysical measurements as covariates for 3D modelling with soil depth functions. The final predictions are sketched in Fig. 6 and shown in detail in the supplements 1 to 4.

Analysing the interaction of spatial prediction of SOC and SM and the geophysical covariates, one can see that the highest SOC values were located in areas with high $EC_a$ values (compare Fig. 1). Figure 5 shows topsoil SOC contents of up to 6% in the western and central part of the site (lower left side) and ~3 % in the south (lower right side). This range of SOC content at this particular agricultural field is similar to the range of SOC in most agricultural fields in central Europe (Tóth et al., 2013). A similar pattern can be seen in the SM prediction. In the western part SM is distributed uniformly with 20 to 25 % in the whole profile and in the South, there is much less SM in the deeper subsoil (5 %) than in the topsoil (15 to 20 %). These patterns can also be found in the sampled soil profiles (Fig. 3).

In the central part of the field, pillar-like patterns of higher SOC content values were visible. These pillars are well described and linked to old meanders of the rivulet Fließgraben or the river Elbe. In these areas, a farmer can expect better growing conditions for field crops, as SOC affects nutrient availability and soil moisture retention. Thus, these areas need less additional fertilizer than the areas with less SOC and lower $EC_a$. Employing the proposed framework can therefore contribute to a sustainable agricultural approach, e.g. precision agriculture that applies fertilizer according to soil requirements. 3D mapping is highly suitable for informing farmers as sampling based on wecLHS is fast (only few profiles are required) and the model development based on the geophysical measurements is computationally efficient, relatively fast compared to conventional soil mapping as well as potentially extendable to other soil properties such as pH, cation exchange capacity and texture (Cassiani et al., 2012; Doolittle and Brevik, 2014).

## Conclusion

In this case study, we predicted SOC and SM in the vertical as well as horizontal domain, i.e. in 3D using geophysical covariates derived from EMI and gamma-ray sensors with different intercoil spacings and, thus, different penetration depths and footprints of the signal. A weighted conditioned Latin Hypercube Sampling design was applied for calculation of the

locations of the calibration samples. We hypothesized that the used sensor setup will lead to predictions of SOC and SM with high an explained variation throughout the soil profile as well as in the spatial domain.

We showed that coefficient of determination and model error of the polynomial and exponential functions modelled and predicted with Cubist and random forests were stable over all depth increments. Thus, the data from two EMI sensors with depth dependent sensitivity and gamma-ray spectrometry are well suited for the 3D prediction of SOC and SM despite the reduced number of samples. In general, the differences between the models' error were rather small. The differences between the used machine learning methods are smaller compared to the differences between the used soil depth functions. This demonstrates the suitability of the sampling design approach for modelling with Cubist and random forests. Therefore, we conclude that the choice between soil depth functions is more important than the choice of the machine learning method for spatial prediction, if both methods are well established in DSM. The flexible polynomial function is capable of depicting local variation, which is not limited to plough horizons, but also comprises clay enriched horizons, pH drops with decreases in $CaCO_3$ content and others. We recommend the combination of second-degree polynomial soil depth function with Cubist for 3D mapping of SOC and SM with two EMI sensors and gamma-ray spectrometry covering a wide range of environmental covariates representing the horizontal and vertical domain of SOC and SM variation on the field scale. Within the scope of precision agriculture, this approach is suitable for SOC and SM estimation in similar environmental conditions as it offers a spatial evaluation that incorporates the whole soil continuum. Thus, the 3D mapping of SOC and SM with high spatial and vertical resolution can help to optimise sustainable management strategies on the field scale in respect to fertilization, irrigation, and liming and subsequently to increase food and biofuel productivity.

For future investigation and to simplify the approach for field application, the contribution and importance of the individual sensors and sensor settings are of great interest. Since both orientations of the CMD-Explorer showed similar interpolation results, these measurements may be strongly cross-correlated and redundant. This can be evaluated and solved with the feature importance calculated within random forests, but requires comprehensive and complex analysis of the interaction of the modelled depth function coefficients and the geophysical measurements, e.g. by using different slices of the target variable related to a specific depth increment compared to the depth sensitivity of the sensors. Further, the extension of the modelled depth may be of interest, since plants can uptake water from greater depth. This depends on the crop and also extends to forestry. For that purpose other soil properties, such as texture, water holding capacity, permeability, horizontation, or the extension of

the modelled depth to the depth of bedrock may be of interest. More complex soil depth functions such as polynomials of higher degree may be beneficial, when a trade-off between soil sampling costs and model benefits is found. In our study we successfully integrated depth dependent $EC_a$ data, however, we refrained from inversion of the data because we did not want to introduce additional uncertainties and ambiguities into the data analysis. As recently shown by von Hebel et al. (2019) an enhanced processing chain can provide accurate and quantitative EMI data. This opens interesting possibilities to extend the presented approach by depth-true EC values.

## Supplemental Material

The supplemental material includes animations of the 3D mapping of SOC and SM in vertical cross sections. The cross sections are provided in both north and east dimension to show the vertical variation within cross sections 10 m apart.

## Data Availability

Data sets from this study are available through the PANGAEA open access repository under https://doi.org/10.1594/PANGAEA.910272 (Pohle and Werban, 2019).

## Conflict of Interest Statement

The authors have no conflicts of interest to declare.

## Author Contributions Statement

Conceptualization: T. Rentschler, U. Werban, M. Ahner, K. Schmidt; Data curation: T. Rentschler, U. Werban, M. Ahner; Formal analysis: T. Rentschler, U. Werban, M. Ahner, K. Schmidt; Funding acquisition: U. Werban; Investigation: T. Rentschler, U. Werban, M. Ahner, K. Schmidt; Methodology: T. Rentschler, U. Werban, M. Ahner, T. Behrens, P. Gries, K. Schmidt; Project administration: U. Werban, K. Schmidt; Resources: U. Werban, K. Schmidt; Software: T. Rentschler, M. Ahner, P. Gries, K. Schmidt; Supervision: U. Werban, T. Behrens, T. Scholten, K. Schmidt; Validation: T. Rentschler, M. Ahner, P. Gries, K. Schmidt; Visualization: T. Rentschler; Writing – original draft: T. Rentschler, U. Werban, M. Ahner, T. Behrens, P. Gries, T. Scholten, S. Teuber, K. Schmidt; Writing – review & editing: T. Rentschler, U. Werban, S. Teuber, K. Schmidt

## Acknowledgements

We thank the German Research Foundation (DFG) for supporting this research through the Collaborative Research Centre (SFB 1070) ResourceCultures (subprojects Z, S and B02). We further thank Marco Pohle for assistance with the field measurements.

# Appendix

## References


Abdu, H., Robinson, D.A., Seyfried, M., Jones, S.B., 2008. Geophysical imaging of watershed subsurface patterns and prediction of soil texture and water holding capacity. Water Resour. Res. 44 (4), 189.

Aldana Jague, E., Sommer, M., Saby, N.P.A., Cornelis, J.-T., van Wesemael, B., et al., 2016. High resolution characterization of the soil organic carbon depth profile in a soil landscape affected by erosion. Soil and Tillage Research 156, 185–193.

Binley, A., Hubbard, S.S., Huisman, J.A., Revil, A., Robinson, D.A., et al., 2015. The emergence of hydrogeophysics for improved understanding of subsurface processes over multiple scales. Water Resour. Res. 51 (6), 3837–3866.

Bouma, J., 2014. Soil science contributions towards Sustainable Development Goals and their implementation: linking soil functions with ecosystem services. Journal of Plant Nutrition and Soil Science 177 (2), 111–120.

Bouma, J., McBratney, A., 2013. Framing soils as an actor when dealing with wicked environmental problems. Geoderma 200-201, 130–139.

Breiman, L., 2001. Random Forests. Machine learning 45, 5–32.

Breiman, L., Friedman, J.H., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees. Chapman and Hall, New York, NY.

Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. European Journal of Soil Science 62 (3), 394–407.

Callegary, J.B., Ferre, P.A., Groom, R.W., 2012. Three-dimensional sensitivity distribution and sample volume of low-induction-number electromagnetic-induction instruments. Soil Science Society of America Journal 76 (1), 85–91.

Cassiani, G., Ursino, N., Deiana, R., Vignoli, G., Boaga, J., et al., 2012. Noninvasive Monitoring of Soil Static Characteristics and Dynamic States: A Case Study Highlighting Vegetation Effects on Agricultural Land. Vadose Zone Journal 11 (3), 1–16.

Castrignanò, A., Wong, M.T.F., Stelluti, M., Benedetto, D. de, Sollitto, D., 2012. Use of EMI, gamma-ray emission and GPS height as multi-sensor data for soil characterisation. Geoderma 175-176, 78–89.

Cho, Y., Sudduth, K.A., Chung, S.-O., 2016. Soil physical property estimation from soil strength and apparent electrical conductivity sensor data. Biosystems Engineering 152, 68–78.

Cook, S.E., Corner, R.J., Groves, P.R., Grealish, G.J., 1996. Use of airborne gamma radiometric data for soil mapping. Australian Journal of Soil Research 34 (1), 183–194.

Dexter, A.R., Richard, G., Arrouays, D., Czyż, E.A., Jolivet, C., et al., 2008. Complexed organic matter controls soil physical properties. Geoderma 144 (3-4), 620–627.

Dierke, C., Werban, U., 2013. Relationships between gamma-ray data and soil properties at an agricultural test site. Geoderma 199, 90–98.

Doolittle, J.A., Brevik, E.C., 2014. The use of electromagnetic induction techniques in soils studies. Geoderma 223-225, 33–45.

Ekstrøm, C.T., 2018. MESS: Miscellaneous Esoteric Statistical Scripts.

Fan, J., McConkey, B., Wang, H., Janzen, H., 2016. Root distribution by depth for temperate agricultural crops. Field Crops Research 189, 68–74.

Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., et al., 2010. Food Security: The Challenge of Feeding 9 Billion People. Science 327 (5967), 812–818.

Govers, G., Merckx, R., van Wesemael, B., van Oost, K., 2017. Soil conservation in the 21st century: why we need smart agricultural intensification. SOIL 3 (1), 45–59.

Hengl, T., Jesus, J.M. de, MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., et al., 2014. SoilGrids1km—global soil information based on automated mapping. PLOS ONE 9 (8), e105992.

IAEA, 2003. Guidelines for radioelement mapping using gamma ray spectrometry data. IAEA Technical Documents 1363, 179 pp. https://www-pub.iaea.org/mtcd/publications/pdf/te_1363_web.pdf. Accessed 9 July 2019.

Jenny, H., 1941. Factors of soil formation: A system of quantitative pedology. Dover Publications, Inc., New York.

Jobbágy, E.G., Jackson, R.B., 2000. THE VERTICAL DISTRIBUTION OF SOIL ORGANIC CARBON AND ITS RELATION TO CLIMATE AND VEGETATION. Ecological Applications 10 (2), 423–436.

Krige, D.G., 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of the Southern African Institute of Mining and Metallurgy 52 (6), 119–139.

LAGB, 2014. Bodenbericht Sachsen-Anhalt 2014 - Grundlagen, Parameter und Hintergrundwerte. Mitteilungen zu Geologie und Bergwesen von Sachsen-Anhalt 18, Halle (Saale), Germany, 74 pp. Accessed 9 July 2019.

Liu, F., Rossiter, D.G., Song, X.-D., Zhang, G.-L., Yang, R.-M., et al., 2016. A similarity-based method for three-dimensional prediction of soil organic matter concentration. Geoderma 263, 254–263.

Martini, E., Werban, U., Zacharias, S., Pohle, M., Dietrich, P., et al., 2017. Repeated electromagnetic induction measurements for mapping soil moisture at the field scale: validation with data from a wireless soil moisture monitoring network. Hydrol. Earth Syst. Sci. 21 (1), 495–513.

McBratney, A., Whelan, B., Ancev, T., Bouma, J., 2005. Future Directions of Precision Agriculture. Precision Agriculture 6 (1), 7–23.

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

McNeill, J.D., 1980a. Electrical conductivity of soils and rocks: Technical note TN-5, Mississauga, ON, Canada, 22 pp. http://www.geonics.com/pdfs/technicalnotes/tn5.pdf. Accessed 9 July 2019.

McNeill, J.D., 1980b. Electromagnetic Terrain Conductivity Measurement at Low Induction Numbers: Technical note TN-6, Mississauga, ON, Canada, 15 pp. http://www.geonics.com/pdfs/technicalnotes/tn6.pdf. Accessed 9 July 2019.

Minasny, B., McBratney, A.B., 2006. Chapter 12 Latin Hypercube Sampling as a Tool for Digital Soil Mapping, in: Lagacherie, P., McBratney, A.B., Voltz, M. (Eds.), Digital soil mapping. An introductory perspective, vol. 31, 1. ed. ed. Developments in Soil Science 31. Elsevier, Amsterdam, pp. 153–606.

Minasny, B., McBratney, A.B., Mendonça-Santos, M.L., Odeh, I.O.A., Guyon, B., 2006. Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley. Australian Journal of Soil Research 44 (3), 233–244.

Minasny, B., Stockmann, U., Hartemink, A.E., McBratney, A.B., 2016. Measuring and Modelling Soil Depth Functions, in: Hartemink, A.E., Minasny, B. (Eds.), Digital Soil Morphometrics, vol. 163. Progress in Soil Science. Springer International Publishing, Cham, pp. 225–240.

Moghadas, D., Taghizadeh-Mehrjardi, R., Triantafilis, J., 2016. Probabilistic inversion of EM38 data for 3D soil mapping in central Iran. Geoderma Regional 7 (2), 230–238.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Computers & Geosciences 30 (7), 683–691.

Pohle, M., Werban, U., 2019. Near surface geophysical data (Electromagnetic Induction - EMI, Gamma-ray spectrometry), August 2017, Selbitz (Elbe), Germany.

Probst, P., Boulesteix, A.-L., 2018. To Tune or Not to Tune the Number of Trees in Random Forest. Journal of Machine Learning Research 18 (181), 1–18.

Quinlan, J.R., 1992. Learning with continuous classes. Proceedings of the 5th Australian Joint Conference On Artificial Intelligence, 343–348.

Quinlan, J.R., 1993. Combining Instance-based and Model-based Learning, in: Utgoff, P. (Ed.), Proceedings of the Tenth International Conference on International Conference on Machine Learning. Morgan Kaufmann, San Mateo, pp. 236–243.

R Development Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Wien, Austria.

Ramirez-Lopez, L., Schmidt, K., Behrens, T., van Wesemael, B., Demattê, J.A.M., et al., 2014. Sampling optimal calibration sets in soil infrared spectroscopy. Geoderma 226-227, 140–150.

Rawls, W.J., Pachepsky, Y.A., Ritchie, J.C., Sobecki, T.M., Bloodworth, H., 2003. Effect of soil organic carbon on soil water retention. Geoderma 116 (1-2), 61–76.

Rentschler, T., Gries, P., Behrens, T., Bruelheide, H., Kühn, P., et al., 2019. Comparison of catchment scale 3D and 2.5D modelling of soil organic carbon stocks in Jiangxi Province, PR China. PLOS ONE 14 (8), 1–23.

Schmidt, K., Behrens, T., Daumann, J., Ramirez-Lopez, L., Werban, U., et al., 2014. A comparison of calibration sampling schemes at the field scale. Geoderma 232-234, 243–256.

Schmidt, K., Behrens, T., Scholten, T., 2008. Instance selection and classification tree analysis for large spatial datasets in digital soil mapping. Geoderma 146 (1-2), 138–146.

Steyerberg, E.W., Harrell, F.E., 2016. Prediction models need appropriate internal, internal-external, and external validation. Journal of Clinical Epidemiology 69, 245–247.

Tóth, G., Jones, A., Montanarella, L., 2013. The LUCAS topsoil database and derived information on the regional variability of cropland topsoil properties in the European Union. Environmental monitoring and assessment 185 (9), 7409–7425.

Veronesi, F., Corstanje, R., Mayr, T., 2014. Landscape scale estimation of soil carbon stock using 3D modelling. The Science of the total environment 487, 578–586.

Viscarra Rossel, R.A., Chen, C., Grundy, M.J., Searle, R., Clifford, D., et al., 2015. The Australian three-dimensional soil grid: Australia's contribution to the GlobalSoilMap project. Soil Res. 53 (8), 845.

von Hebel, C., van der Kruk, J., Huisman, J.A., Mester, A., Altdorff, D., et al., 2019. Calibration, Conversion, and Quantitative Multi-Layer Inversion of Multi-Coil Rigid-Boom Electromagnetic Induction Data. Sensors 19 (21).

**Table 1**. Intercoil spacings and effective penetration depth for vertical and horizontal coil orientation for the used EMI sensors CMD-Mini-Explorer and CMD-Explorer.

| EMI sensor | intercoil spacing (m) | effective penetration depth (m) | |
|---|---|---|---|
| | | VDP† | HDP† |
| CMD-Mini-Explorer | 0.32 | 0.50 | 0.25 |
| | 0.71 | 1.00 | 0.50 |
| | 1.18 | 1.80 | 0.90 |
| CMD-Explorer | 1.48 | 2.20 | 1.10 |
| | 2.82 | 4.20 | 2.10 |
| | 4.49 | 6.70 | 3.30 |

†VDP: vertically oriented dipole; HDP: horizontally oriented dipole

# Appendix

**Table 2.** Results of the 10-fold cross-validation of the interpolation with ordinary kriging (RMSE units are mS $m^{-1}$ for EMI sensors, % for $^{40}K$, ppm for $^{232}Th$ and $^{238}U$ concentrations and Gy $h^{-1}$ for dose rate). The sensors and sensor setups we used for the wecLHS sampling design are highlighted in bold.

| Sensor setting (coil distance) | effective depth range (m) | Min | Max | $R^2$ | RMSE | nRMSE |
|---|---|---|---|---|---|---|
| CMD-Mini-Explorer VDP (0.32m) | 0.5 | 0.00 | 45.63 | 0.98 | 1.07 | 0.02 |
| CMD-Mini-Explorer VDP (0.71m) | 1.0 | 13.47 | 68.83 | 0.99 | 0.69 | 0.01 |
| CMD-Mini-Explorer VDP (1.18m) | 1.8 | 5.92 | 62.75 | 0.99 | 0.82 | 0.01 |
| CMD-Mini-Explorer HDP (0.32m) | 0.25 | 0.00 | 27,65 | 0.98 | 1.05 | 0.03 |
| CMD-Mini-Explorer HDP (0.71m) | 0.5 | 0.00 | 40.25 | 0.98 | 1.03 | 0.02 |
| CMD-Mini-Explorer HDP (1.18m) | 0.9 | 3.69 | 50.69 | 0.98 | 0.79 | 0.02 |
| CMD-Explorer VDP (1.48m) | 2.2 | 5.89 | 41.95 | 0.97 | 1.65 | 0.04 |
| CMD-Explorer VDP (2.82m) | 4.2 | 7.26 | 47.33 | 0.98 | 2.00 | 0.05 |
| CMD-Explorer VDP (4.49m) | 6.7 | 8.50 | 48.13 | 0.98 | 1.79 | 0.04 |
| CMD-Explorer HDP (1.48m) | 1.1 | 5.31 | 26.10 | 0.97 | 1.91 | 0.08 |
| CMD-Explorer HDP (2.82m) | 2.1 | 4.94 | 34.91 | 0.97 | 1.72 | 0.05 |
| CMD-Explorer HDP (4.49m) | 3.3 | 6.61 | 39.80 | 0.96 | 1.29 | 0.04 |
| $^{40}K$ | - | 0.60 | 1.20 | 0.75 | 0.05 | 0.08 |
| $^{232}Th$ | - | 2.36 | 9.66 | 0.80 | 0.48 | 0.06 |
| $^{238}U$ | - | 0.91 | 3.52 | 0.75 | 0.24 | 0.07 |
| Dose rate | - | 19.24 | 51.87 | 0.92 | 1.50 | 0.05 |

†VDP: vertically oriented dipole; HDP: horizontally oriented dipole

**Table 3.** Summary of the model validation results with the coefficient of determination ($R^2$) and root mean squared error (RMSE in %) of the polynomial and exponential depth functions for SOC and SM.

| | | Function | Min | Median | Mean | Max | SD |
|---|---|---|---|---|---|---|---|
| SOC | $R^2$ | polynomial | 0.93 | 0.98 | 0.98 | 1.00 | 0.02 |
| | | exponential | 0.83 | 0.95 | 0.94 | 1.00 | 0.04 |
| | RMSE | polynomial | 0.02 | 0.15 | 0.14 | 0.29 | 0.08 |
| | | exponential | 0.08 | 0.27 | 0.27 | 0.51 | 0.11 |
| SM | $R^2$ | polynomial | 0.13 | 0.99 | 0.92 | 1.00 | 0.18 |
| | | exponential | 0.00 | 0.88 | 0.77 | 0.99 | 0.26 |
| | RMSE | polynomial | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 |
| | | exponential | 0.00 | 0.01 | 0.01 | 0.02 | 0.01 |

†Min: minimum value; Max: maximum value; SD: standard deviation

**Table 4.** Results of the independent model validation with the explained variation as coefficient of determination ($R^2$), Lin's concordance correlation coefficient (CCC) and root mean squared error (RMSE in %) of the polynomial and exponential soil depth functions for SOC and SM.

| | | Cubist | | | Random forests | | |
|---|---|---|---|---|---|---|---|
| | Function | $R^2$ | CCC | RMSE | $R^2$ | CCC | RMSE |
| SOC | polynomial | 0.86 | 0.91 | 0.56 | 0.89 | 0.92 | 0.54 |
| | exponential | 0.87 | 0.91 | 0.58 | 0.86 | 0.90 | 0.61 |
| SM | polynomial | 0.88 | 0.91 | 0.02 | 0.84 | 0.90 | 0.02 |
| | exponential | 0.87 | 0.93 | 0.02 | 0.86 | 0.92 | 0.02 |

**Table 5.** Results of the independent model validation for each individual sampled depth (as mid-point of the depth increment) with the coefficient of determination ($R^2$), Lin's concordance correlation coefficient (CCC), root mean squared error (RMSE) and the normalised root mean squared error (nRMSE) for SOC and SM. For SOC, Cubist with polynomial functions had the lowest error as well as for SM.

| | | SOC | | | | SM | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Cubist | | Random forests | | Cubist | | Random forests | |
| | Depth | polynomial | exponential | polynomial | exponential | polynomial | exponential | polynomial | exponential |
| $R^2$ | 7.5cm | 0.76 | 0.76 | 0.83 | 0.82 | 0.77 | 0.83 | 0.70 | 0.75 |
| | 22.5cm | 0.64 | 0.80 | 0.76 | 0.77 | 0.84 | 0.92 | 0.87 | 0.89 |
| | 37.5cm | 0.49 | 0.76 | 0.63 | 0.64 | 0.89 | 0.90 | 0.83 | 0.87 |
| | 52.5cm | 0.27 | 0.65 | 0.55 | 0.51 | 0.89 | 0.86 | 0.85 | 0.90 |
| CCC | 7.5cm | 0.84 | 0.83 | 0.86 | 0.88 | 0.85 | 0.83 | 0.85 | 0.83 |
| | 22.5cm | 0.64 | 0.55 | 0.65 | 0.52 | 0.89 | 0.92 | 0.90 | 0.86 |
| | 37.5cm | 0.53 | 0.62 | 0.60 | 0.55 | 0.89 | 0.94 | 0.89 | 0.92 |
| | 52.5cm | 0.47 | 0.72 | 0.65 | 0.62 | 0.92 | 0.93 | 0.90 | 0.94 |
| RMSE (%) | 7.5cm | 0.38 | 0.39 | 0.38 | 0.35 | 0.02 | 0.02 | 0.02 | 0.02 |
| | 22.5cm | 0.58 | 0.77 | 0.61 | 0.81 | 0.01 | 0.02 | 0.01 | 0.02 |
| | 37.5cm | 0.71 | 0.67 | 0.67 | 0.73 | 0.02 | 0.02 | 0.02 | 0.02 |
| | 52.5cm | 0.54 | 0.38 | 0.47 | 0.43 | 0.03 | 0.03 | 0.03 | 0.03 |
| nRMSE | 7.5cm | 0.17 | 0.17 | 0.16 | 0.15 | 0.14 | 0.14 | 0.16 | 0.15 |
| | 22.5cm | 0.23 | 0.31 | 0.25 | 0.33 | 0.15 | 0.15 | 0.14 | 0.21 |
| | 37.5cm | 0.22 | 0.21 | 0.21 | 0.23 | 0.12 | 0.10 | 0.12 | 0.12 |
| | 52.5cm | 0.14 | 0.19 | 0.12 | 0.22 | 0.06 | 0.06 | 0.06 | 0.05 |

## Appendix

**Manuscript 3**

# Improving the spatial prediction of soil organic carbon content in two contrasting climatic regions by stacking machine learning models and rescanning input space

*Remote Sensing 12 1095*

Ruhollah Taghizadeh-Mehrjardi[1,2,*], Karsten Schmidt[1,8,9,*], Alireza Amirian-Chakan[3], Tobias Rentschler[1,8], Mojtaba Zeraatpisheh[4], Fereydoon Sarmadian[5], Roozbeh Valavi[6], Naser Davatgar[7], Thorsten Behrens[1,8,9], and Thomas Scholten[1,8,9]

[1] Department of Geosciences, Soil Science and Geomorphology, University of Tübingen, Tübingen, Germany
[2] Faculty of Agriculture and Natural Resources, Ardakan University, Ardakan, Iran
[3] Department of Soil Science, Lorestan University, Khorramabad, Iran
[4] Key Laboratory of Geospatial Technology for the Middle and Lower Yellow River Regions, College of Environment and Planning, Henan University, Kaifeng, Henan Province, China
[5] Department of Soil Science, College of Agriculture, University of Tehran, Tehran, Iran
[6] The Quantitative & Applied Ecology Group, School of BioSciences, The University of Melbourne, Australia
[7] Agricultural Research, Education and Extension Organization, Karaj, Iran
[8] CRC 1070 RessourceCulture, University of Tübingen, Tübingen, Germany
[9] DFG Cluster of Excellence "Machine Learning", University of Tübingen, Germany

*Corresponding authors:
E-Mail: ruhollah.taghizadeh-mehrjardi@mnf.uni-tuebingen.de; karsten.schmidt@uni-tuebingen.de

Received: February 21, 2020; Accepted: March 26, 2020; Published: March 29,2020

## Abstract

Understanding the spatial distribution of soil organic carbon (SOC) content over different climatic regions will enhance our knowledge of carbon gains and losses due to climatic change. However, little is known about the SOC content in the contrasting arid and sub-humid regions of Iran whose complex SOC-landscape relationships pose a challenge to spatial analysis. Machine learning (ML) models with a digital soil-mapping framework can solve such complex relationships. Current research focusses on ensemble ML models to increase the accuracy of prediction. The usual ensemble method is boosting or weighted averaging. This study proposes a novel ensemble technique: the stacking of multiple ML models through a meta-learning model. In addition, we tested the ensemble through rescanning the input space to maximize the prediction accuracy. We first applied six state-of-the-art ML models (i.e., Cubist, Random Forests (RF), extreme gradient boosting (XGBoost), classical artificial neural network models (ANN), neural network ensemble based on model averaging (AvNNet), and deep learning neural networks (DNN)) to predict and map the spatial distribution of SOC content at six soil depth intervals for both regions. In addition, the stacking of multiple ML models through a meta-learning model with/without rescanning the input space were tested and applied to maximize the prediction accuracy. Out of six ML models, the DNN resulted in the best modelling accuracies, followed by RF, XGBoost, AvNNet, ANN, and Cubist. Importantly, the stacking of models indicated a significant improvement in the prediction of SOC content, especially when combined with rescanning the input space. For instance, the RMSEs for SOC content prediction of the upper 0-5 cm of the soil profiles of the arid site and the sub-humid site by the proposed stacking of the approaches were 17% and 9%, respectively; less than that obtained by the DNN models - the best individual model. This indicates that rescanning the original input space by a meta-learning model can extract more information and improve the SOC content prediction accuracy. Overall, our results suggest that the stacking of diverse sets of models could be used to accurately estimate the spatial distribution of SOC content in different climatic regions.

## Keywords

Digital soil mapping; machine learning models; stacking of models; spatial block cross-validation; deep learning

## Introduction

Soil organic carbon (SOC) is a key function of soils, influencing soil physicochemical properties [1,2], e.g. soil water storage capacity, nutrient holding capacity and infiltration rate. As the world's soils contain more organic carbon than the atmosphere and the biosphere

together, soils are considered to be a crucial pool in the global carbon cycle [3]. Thus, accurate information on the spatial distribution of SOC is vital to estimate and predict greenhouse gas emission and physicochemical functions of soils [4,5]. Such information is most important in arid and semi-arid areas where soils tend to have low levels of organic carbon [6,7] compared to the humid region. These sensitive and fragile ecosystems are less resilient against climate change and, therefore, more vulnerable to desertification [8,9].

Legacy soil maps based on traditional soil mapping approaches are the most common sources for acquiring data and information on soils in Iran [9]. The qualitative nature and coarse scales of the available maps make these maps impractical for quantitative studies and a detailed understanding of the spatial variations of soil properties [10-12]. Furthermore, traditional soil mapping approaches are time-consuming and expensive [13]. Digital soil mapping (DSM) approaches based on the scorpan concept [14] have become a standard approach to generate new soil data to overcome the limitations arising from the legacy soil maps. DSM provides a quantitative-empirical framework for predicting soil properties and classes from spatially referenced covariates using appropriate machine learning (ML) models [5].

Several ML models have successfully linked SOC to environmental covariates to extrapolate SOC to unknown locations [15-29]. Some of the most popular models are multivariate regression, classical artificial neural networks [13], support vector regression [20], regression trees [17,20] and random forests [15,20,30]. Recently, deep neural networks (DNN) based on deep learning approaches were used to solve highly complex soil-landscape problems [31-34]. Padarian et al. [33] and Wadoux et al. [34] predicted and mapped SOC in Chile and France, respectively, using deep learning methods.

One commonly applied technique to improve the predictive capacity and to decrease the variance of the individual ML model is the ensemble model-bagging, boosting and stacking approach [35]. Bagging is a simple and very powerful ensemble method. It generates $m$ new training sets and then $m$ models are fitted to the data sets. Their prediction results are combined by averaging the output or voting. Boosting refers to a group of algorithms that utilize weighted averages to turn weak learners into stronger learners. Other ensemble techniques include model averaging [36,37].

The stacking approaches combine different types of models (lower level) through a meta-learning model (higher level) to maximize the generalization accuracy [38]. Unlike bagging, boosting and averaging methods, stacking ensemble modelling is rarely explored in digital soil mapping. Nevertheless, stacking often performs better than all individual models, especially when combined with rescanning the original input data [39]. For instance, Tajik et al. [40],

Zhou et al. [41] and Chen et al. [42] recently evaluated the efficacy of the ensemble models – by averaging of the model predictions – to predict the spatial variation of soil properties in Iran, China, and France, respectively.

Although different ML models were implemented in order to predict and map SOC [5], their performances are inconsistent in various SOC studies. To the best of our knowledge, there is no study to conduct digital mapping of SOC content using stacking approaches in different climatic conditions. Thus, the authors suggest combining the ML models with the rescanning of the original input training data to explore if it works better than the standard stacking of individual models. Furthermore, so far only a few studies have used deep learning models for DSM, with the notable exceptions of [31,33,34], and a comparison with other models is still needed. Finally, there is a lack of understanding concerning the prediction of SOC content under different climatic regimes in Iran, which has a vast territory and diverse climates. Most studies conducted on SOC content in Iran, only consider a single climatic influence [7,22,30].

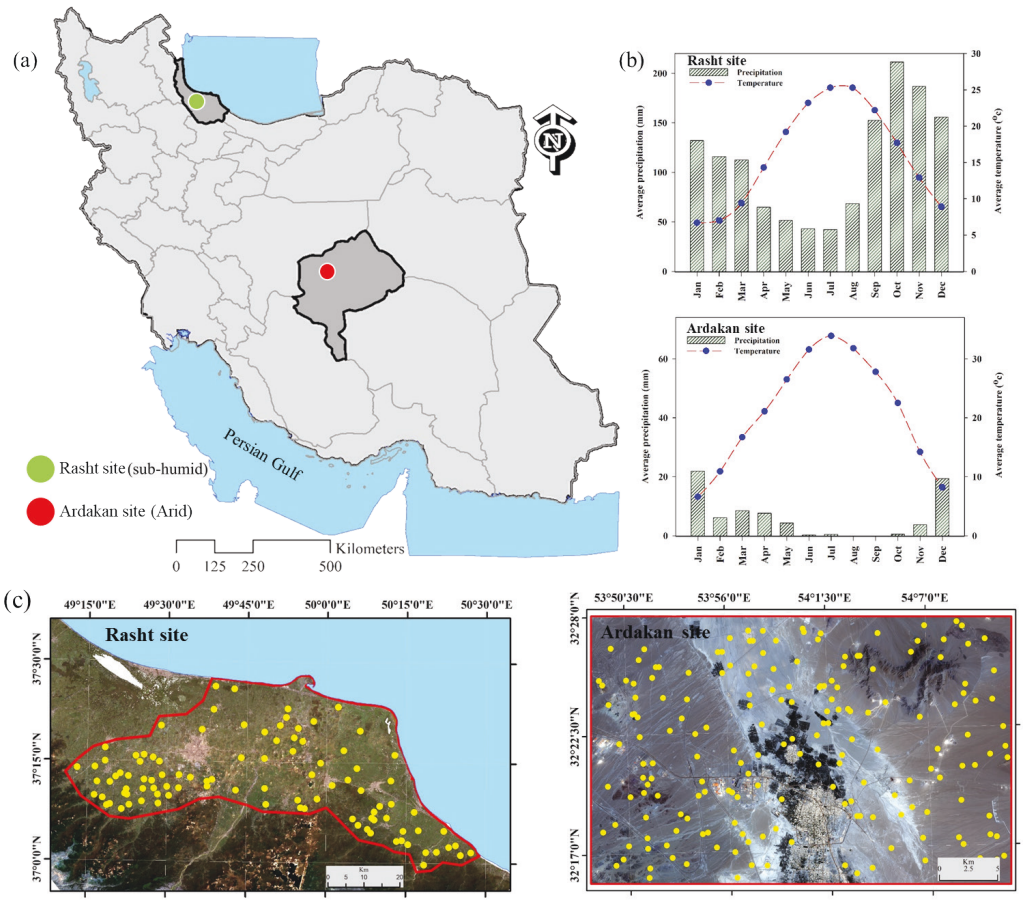


**Figure 1.** Study areas in Iran (a), the corresponding ombrothermic graphs (b) and the geographic coordinates of the extents of the study areas and the spatial distribution of soil observation locations on a false color composite of Sentinel-2 (c).

## Appendix

Hence, the main objective of this study is to evaluate and compare stacking ensemble approaches with six ML models in order to predict and map the spatial distribution of SOC content for two areas with contrasting climate (i.e., arid and sub-humid) in Iran. The models include Cubist, RF, XGBoost, ANN, AvNNet as well as DNN. We further tried to identify the controlling factors of the spatial distribution patterns of SOC content in the contrasting climatic conditions, which has rarely been reported in Iran.

## Materials and Methods

### Study sites

This study was conducted at two sites located in central (Ardakan site) and northern (Rasht site) Iran (Figure 1a). The study sites comprise two diverse climatic regions (Figure 1b), which are arid in Ardakan and sub-humid in Rasht [43]. The general climate conditions of selected sites are presented in Table 1.

**Table 1.** Study sites and data collection details.

| Site | Area (km²) | Soil types | Climate conditions | Precipitation (mm/year) | Elevation (m) | Samples (no.) |
|---|---|---|---|---|---|---|
| Ardakan | 720 | Solonchaks, Gypsisols and Regosols | Arid | 75 | 944–1944 | 154 |
| Rasht | 3000 | Kastanozems, Cambisols and Chernozems | Sub-Humid | 1200 | -26–700 | 99 |

The Ardakan study site is located in the Yazd province in central Iran and covers 720 km². The average annual precipitation, temperature and annual potential evaporation are 75 mm, 18.5 ℃ and 3483 mm, respectively. The soil moisture and temperature regimes are aridic and thermic [43,44]. The elevation ranges from 944 to 1944 m above sea level. The main land use types consist of cropland (pistachio nuts and wheat) and grassland. The major physiographic units from East to West are alluvial fans, coalescing alluvial fans (bajadas), salt plains and gypsiferous hills. The predominant soils in the study area [43,44] are Solonchaks with ~40%, Gypsisols with ~40% and Regosols with ~20% of the area [43,44].

The Rasht study site is located in the Guilan province in northern Iran and covers 3000 km². The climate is sub-humid and the average annual precipitation, temperature and annual potential evaporation are 1200 mm, 15.6 ℃ and 796 mm, respectively. The soil moisture and temperature regimes are udic and thermic [43]. The elevation ranges from -26 m to 700 m above sea level. The main land use types consist of cropland (rice) and forest (oak, beech, lime and elm). Except for the southern parts of the study area, where piedmonts and hills dominate, the topography of the area is mainly flat. Predominant soils of the study area [43] are Kastanozems with ~70%, Cambisols with ~25% and Chernozems with ~5% of the area.

To make the two areas readable, this paper marks the regions as arid site and sub-humid site, respectively, instead of using acronyms.

**Data collection and soil sample analysis**

For the purpose of DSM a well-distributed sample set is needed. We used the conditioned Latin Hypercube Sampling (cLHS), which provides an optimal stratification of the covariate space [45,46], to select representative sample locations based on the covariates [47-52]. We selected a total of 154 and 99 soil profiles for the arid and sub-humid (Figure 1c) sites, respectively. Soil samples were collected from the genetic horizons of each profile down to a depth of 2 m. Air-dried soil samples were ground and sieved (< 0.5 mm), and the SOC content (%) was determined using wet oxidation [53].

Sampling by genetic horizons means that samples do not come from consistent depth intervals in all locations. Therefore, we used an equal-area spline function (ESF) [54] to harmonize SOC data and estimate the vertical variation of SOC content. The ESF was fitted to each profile. Then, the values of SOC content were obtained by the integration of the splines to the defined depth intervals. We estimated the SOC at six depth intervals of 0–5, 5–15, 15–30, 30–60, 60–100 and 100–200 cm in accordance with the standard depths specified by the GlobalSoilMap project [55].

**Covariates used for the development of ML models**

We used a set of 28 covariates (Table 2) as predictors [5,14] representing potential environmental drivers of the spatial and vertical distribution of SOC content. Based on the understanding of the factors affecting the SOC content distribution in the two study areas [7,22,50] and literature reviews [5,56], the covariates were obtained and derived from a digital elevation model (DEM) and remotely sensed satellite data.

The Shuttle Radar Topography Mission (SRTM) DEM with a resolution of 30 × 30 m was used for the terrain analysis [57]. The DEM was preprocessed to fill the sinks and pits before ten terrain attributes were calculated (Table 2) using SAGA GIS [58]. These are elevation, wetness index, catchment area, catchment slope, multi-resolution valley bottom flatness index, valley depth, plane curvature, profile curvature, general curvature and total insolation.

The remote sensing (RS) based covariates were derived and calculated based on the median values of 127 cloud-free Landsat-8 [59] and Sentinel-2 [60] images taken during 2016 under clear and dry weather conditions during the spring/summer season using the Google Earth Engine environment [61]. In general, we used six spectral bands of Landsat-8 (B2, B3, B4, B5, B6 and B7) and ten spectral bands of Sentinel-2 (B2, B3, B4, B5, B6, B7, B8, B8a, B11 and B12), respectively [47,48]. Additionally, we calculated the NDVI (normalized difference vegetation index) using spectral bands of both Landsat-8 and Sentinel-2 [62-64].

## Appendix

All covariates were rescaled using z score standardization and resampled in order to have similar scale and the same cell size of 30 × 30 m.

**Table 2.** Covariates used for the development of ML models.

| No. | Definition | Abr. |
|---|---|---|
| | Terrain-based covariates | |
| 1 | Elevation | Elev |
| 2 | Wetness Index | WI |
| 3 | Catchments area | Ca.Area |
| 4 | Catchment Slope | Ca.Slop |
| 5 | Multi-resolution Valley Bottom Flatness Index | MrVBF |
| 6 | Valley Depth | Vally.D |
| 7 | Plane Curvature | Pl.Cur |
| 8 | Profile Curvature | Pr.Cur |
| 9 | General Curvature | Ge.Cur |
| 10 | Total Insolation | To.In |
| | RS-based covariates | |
| 11 | Blue band of Landsat-8 [0.482 μm] | B2.L |
| 12 | Green band of Landsat-8 [0.561 μm] | B3.L |
| 13 | Red band of Landsat-8 [0.654 μm] | B4.L |
| 14 | Near infrared band of Landsat-8 [0.864 μm] | B5.L |
| 15 | Shortwave IR-1 band of Landsat-8 [1.608 μm] | B6.L |
| 16 | Shortwave IR-2 band of Landsat-8 [2.200 μm] | B7.L |
| 17 | Blue band of Sentinel-2 [0.490 μm] | B2.S |
| 18 | Green band of Sentinel-2 [0.560 μm] | B3.S |
| 19 | Red band of Sentinel-2 [0.665 μm] | B4.S |
| 20 | Vegetation Red Edge of Sentinel-2 [0.705 μm] | B5.S |
| 21 | Vegetation Red Edge of Sentinel-2 [0.740 μm] | B6.S |
| 22 | Vegetation Red Edge of Sentinel-2 [0.783 μm] | B7.S |
| 23 | Near infrared band of Sentinel-2 [0.842 μm] | B8.S |
| 24 | Vegetation Red Edge of Sentinel-2 [0.865 μm] | B8a.S |
| 25 | Shortwave IR-1 band of Sentinel-2 [1.610 μm] | B11.S |
| 26 | Shortwave IR-2 band of Sentinel-2 [2.190 μm] | B12.S |
| 27 | Normalized difference vegetation index (Landsat-8 based) | NDVI.L |
| 28 | Normalized difference vegetation index (Sentinel-2 based) | NDVI.S |

### Covariate selection

In this study, the Boruta algorithm [65] was implemented with the random forest (RF) classifier in the R statistical package [66] to rank the most important covariates for predicting SOC content at six depths. It helps interpret and understand the soil-landscape relationships of the two sites. The algorithm consists of the following steps:

i) the covariate space is extended by adding randomly permuted existing covariates (pC) in order to remove their correlation with SOC content,

ii) a RF prediction using the extended covariate space (i.e. covariates and permuted covariates) is performed to predict SOC content at six standard depths,

iii) the z-score which is an indicator of the importance of all covariates is computed,

iv) the maximum Z-score (MZSA) among the pC's is defined,

v) a hit is assigned to every covariates that scored better than MZSA,

vi) a two-test of equality is performed for undetermined important covariates,

vii) the original covariates are respectively flagged as "unimportant" or "important" if they have significant lower or higher scores than MZSA,

viii) all permuted covariates are removed

ix) repeating the procedure.

In this study, based on Z-score values [67], we grouped the ability of covariates to explain SOC content variability into 4 classes: weakly relevant ($Z<5$), slightly relevant ($5<Z<10$), moderately relevant ($10<Z<15$) and relevant ($Z>15$).

### Stacked generalization

Stacked generalization or simply stacking is an ensemble approach that combines the outcomes of different ML models in a single model to maximize the generalization accuracy [35,37]. Usually, as illustrated in Figure 2, there are two levels in a stacking framework: level 0 and level 1, consisting of several base models and one meta-learning model. Meta-learning models in level 1 use the prediction of the response variables that are estimated by several base models in level 0 in order to generate a final prediction. In other words, the model in level 1 learns with the predictions of the models of level 0.

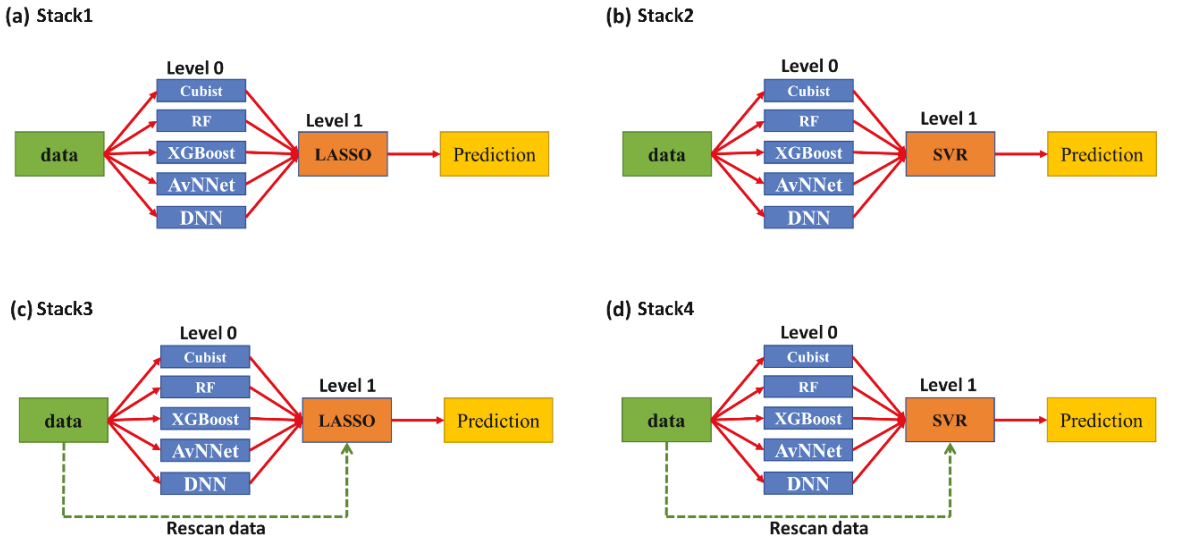


**Figure 2.** General framework of stacking approaches used in this study. (a) Cubist/RF/XGBoost/ANN/AvNNet/DNN + LASSO; Stack1; (b) Cubist/RF/XGBoost/ ANN/AvNNet/DNN + SVR; Stack2; (c) Cubist/RF/XGBoost/ANN/AvNNet/DNN + LASSO + rescan mode; Stack3; (d) Cubist/RF/XGBoost/ANN/AvNNet/DNN + SVR + rescan mode; Stack4

In this study, we used six ML models (Cubist, RF, XGBoost, ANN, AvNNet, and DNN) in level 0. Conventionally the meta-learning model in level 1 is based on a weighted average method or a linear regression model [36,37]. In this study, we applied two new meta-learning models (least absolute shrinkage and selection operator (LASSO); support vector regression (SVR)) in level 1.

Furthermore, we introduced two modes of stacking: the standard mode (Stack1 and Stack2) and the rescan mode (Stack3 and Stack4), as shown in Figure 2 [68]. Special attention should be given to the fact that in the rescan mode, we allow the model in level 1 (LASSO and SVR) to learn again from the original input data in order to extract some missing information. Practically, in the standard mode, we used the predicted SOC contents of individual models (Level 0: such as RF and ANN) as the predictor variables for meta-learning models (Level 1: such as LASSO and SVR). In the rescan mode, we used both the primary covariates (such as NDVI and MrVBF) and predicted SOC contents of individual models (Level 0: such as RF and ANN) as the predictor variables for meta-learning models (Level 1: such as LASSO and SVR). In summary, we tested four stacking approaches. A more detailed account of models used in level 0 and level 1 is given in the following sections.

**The individual ML models in level 0**

Cubist [69] is an extension of the M5 algorithm. It is similar to common regression trees, but the terminal nodes contain linear least square models of covariates used in the previous intermediate node [70] rather than discrete values. Also, there are intermediate linear least squares models at each step of the tree, which are used to adjust the final prediction. Cubist uses "if–then" rules to partition the training data [71]. Whenever the conditions of a rule are satisfied, the associated linear least square model is used to predict the response [72].

RF [73] is an ensemble technique based on the well-known classification and regression tree approach (CART). The ensemble is generated by averaging several trees based on different bootstrap sample sets selected from the training data. Further, only a random subset of covariates is evaluated at each node. RF with a large number of trees is robust against overfitting, noise, as well as non-informative and correlated features. RF has been used in various DSM studies over the past decade [74-76] and for many other environmental problems [77]. Extreme gradient boosting (XGBoost) [78] is also a tree-based ensemble method. However, instead of independent trees and averaging the individual predictions, the XGBoost creates a number of decision trees sequentially. The trees are generated by using the residuals or prediction errors of the previous tree model, thus, the algorithm focuses more on samples with higher uncertainty. Finally, all generated models are added together to calculate the outcome [79].

The most common ANNs also known as multi-layer perceptron (MLP) consist of three layers, i.e. an input layer, a hidden layer and an output layer. Each hidden unit combines all input units of the input layer, where all connections are associated with a weight. Further, an activation function is applied to the sum of weighted unit inputs. The output layer is calculated the same way as the hidden units, but with input from the hidden units. For the MLP with one hidden layer we used the sigmoid function as the activation function in nnet package [80,81]. The network was trained through back-propagation using the Levenberg-marquardt algorithm with 150 iterations [82].

AvNNet is similar to MLP, but multiple neural network models with the same topology are used to predict the response. The models can be different either due to different random number seeds to initialize the network or by fitting the models on bootstrap samples of the original training set (i.e. bagging the neural network). All the resulting models are used for prediction [81]. For regression, the outputs from each network are averaged [83]. The idea behind AvNNet is that we usually train different ANN models for the same problem in order to figure out the best model that produces the best validation statistics. However, instead of choosing the best model, it is possible to combine all models in order to improve the

generalisation power of a single neural network [84]. In this study, we used AvNNet model [80,81] in the Caret package [85]. We note that the tuning parameters used for MLP were kept the same for AvNNet.

Deep Learning Neural Networks (DNN) use the MLP structure, but have more hidden layers and a more hierarchical structure [86]. DNNs with multiple hidden layers have a huge number of hyper-parameters [e.g. optimization algorithm, learning rate, network weight initialization, hidden layers activation function, output activation function, L2 regularization, dropout regularization, and the number of nodes in the hidden layers] [87]. The hyper-parameters potentially allow DNNs to perform better in solving the complex problems compared to the other ML models [88]. Sometimes, however, a lack of control over the learning process of the DNNs may lead to overfitting [32]. One approach, which is also used in this study to avoid or reduce overfitting, is to use a technique called Dropout [89]. Dropout randomly mutes neurons of the hidden layers. This dropout is applied to each of the $_{n}$ training steps, resulting in n different networks that are finally averaged for prediction [90]. For predictions, the ensemble of sparse networks resulting from the dropout process is averaged using the geometric mean of the input weights of the neurons. In this study, for DNNs we used the $H_2O$ package [91] with the rectifier function as a non-linear transformation and the Stochastic Gradient Descent (SGD) as optimization algorithm. Furthermore, in order to save training time, an early stopping was used if no changes in the loss were observed after 150 epochs.

**Meta-learning models in level 1**

The Least absolute shrinkage and selection operator (Lasso) is a regularized linear model. It adds a regularization term as a cost function to a linear model, to reduce its degrees of freedom. To achieve this, the lasso regression performs feature selection by eliminating the weights of the least important predictors. For the Lasso modelling we use the glmnet package [95].

Support vector machines are a kernel method for classification [92] and regression problems [93]. The input data is transformed into a high dimensional feature space with a predefined kernel function. In the high dimensional feature space, a linear regression hyperplane is derived for non-linear relationships. Then, the hyperplane is back-transformed to non-linear space. The kernel used in this study is a radial basis function. The e1071 package [94] was used for radial SVR modelling.

**Optimizing the Hyper-parameters of machine learning models**

We applied a grid-learning method to estimate the best model-parameter by testing different ranges of model parameters listed in Table 3. Importantly, these hyper-parameters are the most likely parameters to have the largest effect on the performance of the ML models. All

other hyper-parameters were set to their defaults [96]. Based on the most relevant parameters, we tuned each model individually and evaluated the prediction performance. Additionally, we combined the grid-learning method with a spatial block cross-validation strategy with the aim to reduce the spatial autocorrelation effect of close neighbors and to choose the optimal model parameter. In this study, we constructed 10 folds for our block cross-validation using R package blockCV [97] in which several spatial blocks can be assigned to a fold (Figure 3). The block-to-fold assignment in this package was done by a repeated random approach that tries to find the most evenly distributed number of observations in each fold. Thus, the observations are separated spatially and in each fold as close as possible to the typical 10-fold cross validation approach.

**Table 3.** Hyper-parameters of ML models tuned in this study.

| ML models | Hyper-parameters | Definition | Defined parameters |
|---|---|---|---|
| Cubist | committees | the number of model trees | 1–100 |
| | neighbors | the number of nearest neighbors | 0–9 |
| XGboost | booster | the type of model | gbtree |
| | max_depth | the depth of tree | 3–10 |
| | min_child_weight | the minimum sum of weights of all observations | 0–5 |
| | colsample_bytree | the number of variables supplied to a tree | 0.5–1 |
| | subsample | the number of samples supplied to a tree | 0.5–1 |
| | eta | learning rate | 0.01–0.5 |
| RF | Mtry | the number of input variables | 1–30 |
| | Ntree | the number of trees | 100–3000 |
| ANN | decay | learning rate | 0.001–0.05 |
| | size | the number of neurons in hidden layer | 1–10 |
| AvNNet | Repeats | the number of MLP with different random number seeds | 3–20 |
| DNN | hidden | the number of hidden layers | 2–10 |
| | size | the number of neurons in hidden layer | 15–200 |
| | network weight initialization | the initialized weight of networks | uniform/he_normal |
| | learning rate | that controls adjusting the weights of the network | 0.001–0.05 |
| | dropout regularization | the amount of the neurons that are randomly dropped | 0.2–0.8 |
| SVM | Kernel type | the kernel function | RBF |
| | C | the penalty parameter | 0.01–100 |
| | | the bandwidth parameter | 0.01–100 |
| Lasso | lambda | the shrinkage parameter | 1–150 |

# Appendix

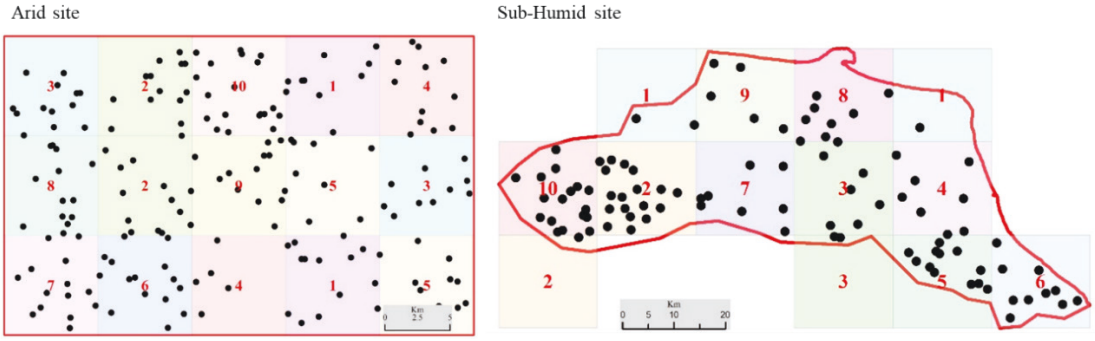


**Figure 3.** Illustration of the spatial blocking strategy in two regions. The numbers in the blocks are fold numbers, showing allocation of blocks to folds.

### Statistical evaluation

In this study, four common performance metrics [98], namely root mean squared error (RMSE), normalized root mean squared error (nRMSE), coefficient of determination ($R^2$) and Ratio of Performance to InterQuartile distance (RPIQ) were used. RMSE indicates the accuracy of the model prediction. nRMSE is without unit and the standardized form of RMSE and well suited for inter-model comparisons. The coefficient of determination ($R^2$) varies between 0 and 1, and indicates the closeness of the observed values to the fitted regression line or the proportion of variance explained by the independent predictors. RPIQ compares the interquartile range to the RMSE [99]. The greater the RPIQ indicates, the better the model's predictive capacity.

$$RMSE = \sqrt{\frac{1}{2}\sum_{i=1}^{n}(P_i - P_o)^2} \tag{1}$$

$$nRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{y_{max} - y_{min}} \tag{2}$$

$$R^2 = \left(\frac{\sum_{i=1}^{n}(O_i - \bar{O})(P_i - \bar{P}))}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2}\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2}}\right)^2 \tag{3}$$

$$RPIQ = \frac{Q_3 - Q_1}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - P_o)^2}} \tag{4}$$

where $P_i$ and $O_i$ are the predicted and observed SOC values at the ith location; n is the number of data points; P and O denote the means for the predicted and observed SOC; $Q_1$ and $Q_3$ are the first and third quartiles, respectively.

## Results and discussion

### Summary statistics of SOC content

The descriptive statistics of the SOC content at six depth intervals across the two study areas are presented in Table 4. For the arid site, the mean SOC content varied from 0.18 to 0.33%, whereas in the sub-humid site it ranged from 1.46 to 4.09% (Table 4). The lower and upper limits of the mean at 95% varied from 0.16 to 0.39 for the arid site, whereas in the sub-humid site it ranged from 1.24 to 4.38%. This indicates a high variability of SOC content across the two sites. The highest variability in SOC content was found at the arid site with a coefficient of variation from 60.39% for the 60 to 100 cm depth to 128.59% for the first depth increment (0-5 cm). Similarly, the sub-humid site showed a high variability of SOC content with a coefficient of variation from 37.15% for the 0 to 5 cm depth to 78.66% for the deepest depth increment (100-200 cm). The arid site, in contrast to the sub-humid site, tended to have higher variability in SOC content at the upmost depth increments.

Furthermore, results revealed that at the sub-humid site, minimum and maximum of SOC content are about ~8 and ~12 times higher than at the arid site. Regardless of other soil forming factors, the sub-humid climate in the study area contributes to the higher SOC content of soils [6]. Whereas, the low SOC content at the arid site is mainly attributed to the scarcity of precipitation (75 mm) as well as a higher mean annual temperature (18.5 ℃), which both contribute to less SOC accumulation [100]. This points out that the climatic conditions in terms of moisture and temperature affect both carbon input into the soil and SOC decomposition [52] and, thus, are key factors in controlling SOC storage in the two study areas.

The results of the mean SOC content comparisons for arid and sub-humid sites are shown in Figure 4. At the sub-humid site, the upper three layers (0-30 cm) are significantly different in terms of mean SOC content, whereas the lower depth intervals show no significant differences in SOC content. This result indicates more variation in the vertical distribution of SOC content at the topsoil compared to the subsoil. The mean SOC contents in arid site show a relatively different trend compared to the sub-humid site. Contrary to the sub-humid site, the upper three depth intervals were not significantly different, while as for sub-humid site, the three subsoil layers were not significantly different in terms of SOC content. This indicated that for the top three layers (0-30 cm), depths intervals had no significant effect on SOC content.

## Appendix

**Table 4.** Descriptive statistics of SOC content at six standard depths in two regions.

| Soil Depth | SOC (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | Min | Max | Mean | Lower | Upper | SD | CV |
| | Arid site | | | | | | |
| 0–5 cm | 0.03 | 2.34 | 0.33 | 0.26 | 0.39 | 0.42 | 128.59 |
| 5–15 cm | 0.04 | 2.21 | 0.31 | 0.25 | 0.37 | 0.39 | 124.56 |
| 15–30 cm | 0.06 | 1.69 | 0.27 | 0.23 | 0.32 | 0.30 | 110.24 |
| 30–60 cm | 0.02 | 1.11 | 0.21 | 0.19 | 0.24 | 0.17 | 77.28 |
| 60–100 cm | 0.01 | 0.75 | 0.18 | 0.16 | 0.19 | 0.11 | 60.39 |
| 100–200 cm | 0.01 | 1.00 | 0.18 | 0.16 | 0.20 | 0.14 | 78.20 |
| | Sub-Humid site | | | | | | |
| 0–5 cm | 1.36 | 9.93 | 4.09 | 3.79 | 4.38 | 1.52 | 37.15 |
| 5–15 cm | 1.28 | 9.51 | 3.68 | 3.41 | 3.95 | 1.39 | 37.89 |
| 15–30 cm | 0.68 | 8.01 | 2.59 | 2.34 | 2.85 | 1.30 | 50.27 |
| 30–60 cm | 0.41 | 5.65 | 1.55 | 1.35 | 1.75 | 1.03 | 66.26 |
| 60–100 cm | 0.07 | 5.65 | 1.46 | 1.24 | 1.69 | 1.15 | 78.21 |
| 100–200 cm | 0.07 | 5.65 | 1.47 | 1.24 | 1.69 | 1.15 | 78.66 |

Min: minimum; Max: maximum; SD: standard deviation; CV: coefficient of variation; Lower and Upper: the lower and upper limits of the mean at 95%.

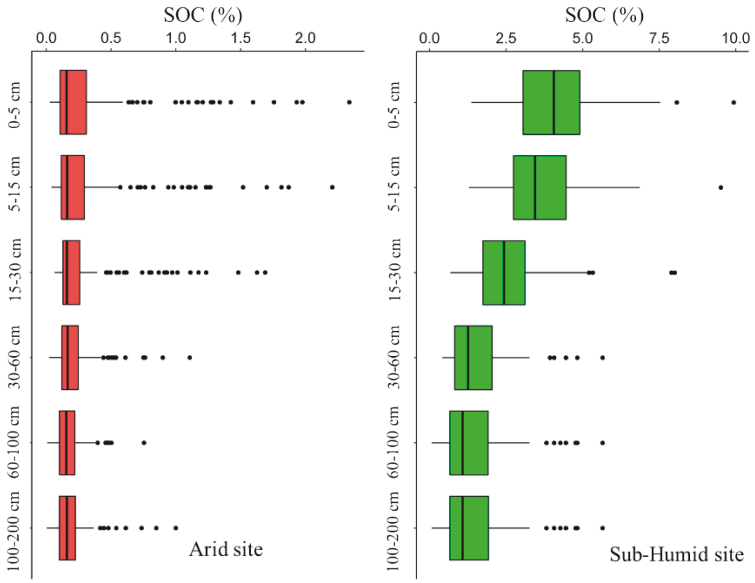


**Figure 4.** The mean SOC content comparisons at six standard depths in two regions. Different letters (a:d) indicate a significant difference at the 0.05 level.

A decreasing trend in SOC content with increasing depth was found on both sites. This is much more evident at the sub-humid site (Figure 5). The SOC content in both, arid and sub-humid areas at the surface layer (0-5 cm) were about 1.8 and 2.8 times higher than the SOC content in the depth of 100 to 200 cm (Table 4). Several studies reported that SOC content in the topsoil was more abundant than in the subsoil [7,22,51,54,101].

### Importance of covariates

The selected covariates for prediction of the SOC concentration at the two sites at all specific depths are presented in Figure 6. The numbers indicate Z-scores and the intensity of colors from light to dark represents the values of Z-scores from low to high, respectively. The covariates used to predict SOC content showed a varying level of importance in the models. Results indicated the covariates in the arid site were weakly to moderately relevant to SOC content. The Z-score varied from 0.40 to 10.60 for valley depth (SOC 100-200 cm) and NIR band of Sentinel-2 (SOC 5-15 cm), respectively. In the sub-humid site it varied from 0.20 to 23.80 for the variables of total insolation (SOC 100-200 cm) and Green band of Sentinel-2 (SOC 0-5 cm), respectively.

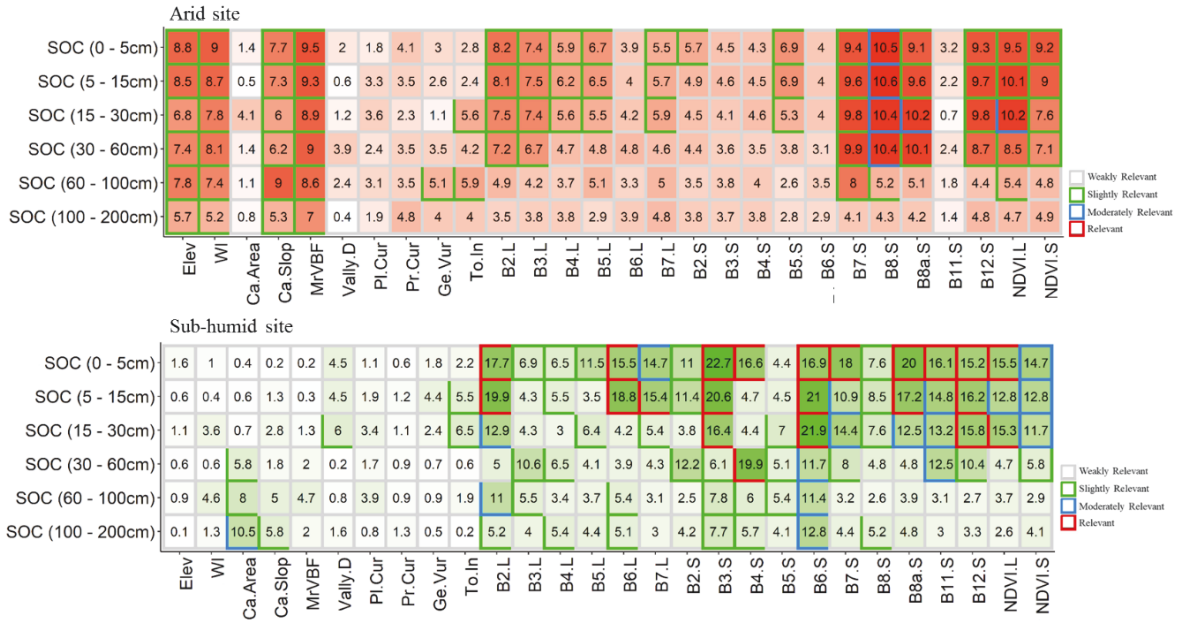


**Figure 5**. Optimal covariate selection using Boruta algorithm for SOC content at six standard depths in two regions.

The plot of selected covariates, Figure 5, shows for the arid site that all covariates were slightly relevant [e.g. Elevation and wetness index] or moderately relevant [e.g. NIR band of Sentinel-2] in predicting SOC content at least at one soil depth interval except for nine covariates [e.g. Catchments area] . Four covariates [e.g. MrVBF] were identified as slightly relevant at all six soil depths (Figure 5). In the sub-humid site, however, only one covariate (vegetation red edge of Sentinel-2) was important at all six depth intervals and six covariates [e.g. Elevation] out of 28 were weakly relevant in predicting SOC at any depth interval (Figure 5). It should be highlighted that all other covariates were classified as slightly relevant [e.g Catchment slope], moderately relevant [e.g. NDVI] and relevant [e.g. Green band of Sentinel-2] variables in predicting SOC content at all six depth intervals.

## Appendix

For the arid site, according to Figure 5, one can further conclude that both terrain- and RS-based covariates play an important role in the prediction of SOC content at six depth intervals, which is in line with the study of Wang et al. [19,102]. Importantly, the RS-based covariates [e.g. NIR band of Sentinel-2] were more important in predicting organic carbon of surface soils in comparison with the terrain-based covariates [e.g. MrVBF]. This can be because most of the area was bare except the center of the area and, thus, the remotely sensed data could represent the spectral behavior of soil at the surface [21,22]. However, terrain-based covariates were more relevant for the SOC content prediction at the depth of 60 to 200 cm, compared to the RS-based covariates. This is particularly true for elevation (Elev), wetness index (WI), catchment slope (Ca.Slop) and multi-resolution valley bottom flatness index (MrVBF). In line with our results, several studies [7,22,54] reported the importance of terrain-based covariates such as WI, MrVBF, slope and elevation for the prediction of SOC at different soil depths.

Whereas for the sub-humid site, the terrain-based covariates are not controlling factors on SOC content variability, with a notable exception of catchments area in the lower depth intervals (30-200 cm). These results were to be expected because the study area in the northern parts of Iran has almost flat terrain [83]. However, the finding revealed the importance of RS-based covariates on SOC content variability in the sub-humid site. For instance, we found NDVI as an important predictor for SOC content at surface layers of soils [47,48]. Furthermore, spectral bands of Sentinel-2 had a substantial influence on the estimation of SOC concentrations at the 0 to 30 cm soil depth in the sub-humid site, as shown in Figure 6. Several studies [47,48,103] also reported a great contribution of Sentinel-2 images to predict SOC contents of soils in the Czech Republic and France.

As discussed above, both terrain- and RS-based covariates can potentially explain the variation of SOC content at the two study sites. Nevertheless, the relative influence of covariates is distinct in arid and sub-humid regions. Our results, as seen in Figure 6, revealed that RS-based covariates could better explain the variation of SOC content in the sub-humid region compared to the arid region. This is expected because vegetation cover by affecting land reflectance in visible and infrared [21], makes RS-based covariates promising explanatory variables to explain SOC contents and variations in the sub-humid regions [22]. However, terrain-based covariates by controlling erosional and depositional processes were more successful in the arid site compared to the sub-humid site to explain SOC content variations. This is also expected because flatness in the sub-humid site minimizes the effect of topography and elevation on the soils [104].

### Performances of the individual ML models

For both sites, the performance of six individual ML models used in level 0 in terms of $R^2$, RMSE and RPIQ at six depth intervals followed as DNN > RF > XGBoost > AvNNet > ANN > Cubist (Tables 5 and 6). Our results indicated the RF models can well predict SOC content in the two study areas. In agreement with our findings, Keskin et al., [67] reported that RF resulted in the lowest RMSE for SOC prediction compared to the other ML models except DNN. Nabiollahi et al., [30] successfully used RF to map SOC stocks at two depth intervals (0-30 and 30-60 cm) using RS- and terrain- based covariates and found it performed fairly good to predict SOC at two soil depths ($R^2$=0.70 and 0.67 respectively). However, Were et al. [20] showed that ANN had lower RMSE and ME values, as well as higher $R^2$ values in predicting SOC in comparison to RF models.

**Table 5**. Performances of the ML models for SOC content at six standard depths in the arid site.

| Models | $R^2$ | RMSE | RPIQ | $R^2$ | RMSE | RPIQ | $R^2$ | RMSE | RPIQ |
|---|---|---|---|---|---|---|---|---|---|
| | 0–5 cm | | | 5–15 cm | | | 15–30 cm | | |
| Cubist | 0.76 | 0.25 | 0.84 | 0.63 | 0.24 | 0.75 | 0.63 | 0.20 | 0.67 |
| XGBoost | 0.79 | 0.20 | 1.12 | 0.71 | 0.19 | 1.02 | 0.69 | 0.17 | 0.85 |
| RF | 0.80 | 0.19 | 1.18 | 0.80 | 0.19 | 1.02 | 0.72 | 0.17 | 0.85 |
| ANN | 0.75 | 0.19 | 1.05 | 0.67 | 0.19 | 0.89 | 0.65 | 0.16 | 0.78 |
| AvNNet | 0.78 | 0.20 | 1.06 | 0.69 | 0.18 | 1.01 | 0.66 | 0.17 | 0.79 |
| DNN | 0.83 | 0.17 | 1.25 | 0.80 | 0.18 | 1.07 | 0.75 | 0.16 | 0.90 |
| Stack1 | 0.83 | 0.17 | 1.25 | 0.78 | 0.18 | 1.07 | 0.74 | 0.15 | 0.92 |
| Stack2 | 0.83 | 0.17 | 1.25 | 0.81 | 0.17 | 1.09 | 0.75 | 0.14 | 0.94 |
| Stack3 | 0.86 | 0.14 | 1.30 | 0.82 | 0.13 | 1.18 | 0.77 | 0.11 | 1.07 |
| Stack4 | 0.90 | 0.14 | 1.37 | 0.85 | 0.13 | 1.20 | 0.78 | 0.10 | 1.11 |
| | 30–60 cm | | | 60–100 cm | | | 100–200 cm | | |
| Cubist | 0.49 | 0.14 | 0.92 | 0.29 | 0.13 | 0.90 | 0.17 | 0.16 | 0.78 |
| XGBoost | 0.56 | 0.14 | 1.00 | 0.33 | 0.13 | 0.99 | 0.26 | 0.16 | 0.84 |
| RF | 0.57 | 0.14 | 1.00 | 0.35 | 0.13 | 0.99 | 0.29 | 0.16 | 0.84 |
| ANN | 0.50 | 0.13 | 0.91 | 0.29 | 0.11 | 0.97 | 0.22 | 0.15 | 0.77 |
| AvNNet | 0.53 | 0.14 | 0.92 | 0.31 | 0.12 | 0.98 | 0.24 | 0.15 | 0.83 |
| DNN | 0.64 | 0.13 | 1.08 | 0.40 | 0.13 | 0.99 | 0.39 | 0.14 | 0.90 |
| Stack1 | 0.63 | 0.11 | 1.13 | 0.41 | 0.12 | 0.99 | 0.40 | 0.13 | 0.94 |
| Stack2 | 0.62 | 0.11 | 1.12 | 0.38 | 0.11 | 1.02 | 0.39 | 0.13 | 0.94 |
| Stack3 | 0.67 | 0.10 | 1.20 | 0.43 | 0.09 | 1.15 | 0.42 | 0.11 | 0.98 |
| Stack4 | 0.72 | 0.09 | 1.29 | 0.46 | 0.08 | 1.19 | 0.44 | 0.10 | 1.06 |

$R^2$: coefficient of determination; RMSE: root mean square error; RPIQ: Ratio of Performance to InterQuartile distance; Stack: refers to the figure 2.

## Appendix

**Table 6.** Performances of the ML models for SOC content at six standard depths in the sub-humid site.

| Models | $R^2$ | RMSE | RPIQ | $R^2$ | RMSE | RPIQ | $R^2$ | RMSE | RPIQ |
|---|---|---|---|---|---|---|---|---|---|
| | 0–5 cm | | | 5–15 cm | | | 15–30 cm | | |
| Cubist | 0.78 | 1.35 | 2.00 | 0.76 | 1.26 | 1.90 | 0.66 | 1.17 | 1.62 |
| XGBoost | 0.78 | 1.28 | 2.08 | 0.76 | 1.23 | 1.92 | 0.66 | 1.10 | 1.69 |
| RF | 0.78 | 1.25 | 2.11 | 0.76 | 1.18 | 1.98 | 0.66 | 1.06 | 1.73 |
| ANN | 0.78 | 1.31 | 2.04 | 0.76 | 1.25 | 1.89 | 0.65 | 1.13 | 1.65 |
| AvNNet | 0.79 | 1.30 | 2.08 | 0.77 | 1.24 | 1.93 | 0.67 | 1.12 | 1.69 |
| DNN | 0.81 | 1.26 | 2.12 | 0.79 | 1.17 | 2.02 | 0.69 | 1.05 | 1.78 |
| Stack1 | 0.83 | 1.21 | 2.16 | 0.82 | 1.17 | 2.05 | 0.73 | 1.06 | 1.78 |
| Stack2 | 0.83 | 1.20 | 2.19 | 0.82 | 1.16 | 2.04 | 0.74 | 1.03 | 1.79 |
| Stack3 | 0.84 | 1.16 | 2.25 | 0.85 | 1.13 | 2.06 | 0.74 | 1.01 | 1.81 |
| Stack4 | 0.87 | 1.15 | 2.29 | 0.86 | 1.12 | 2.10 | 0.78 | 1.01 | 1.83 |
| | 30–60 cm | | | 60–100 cm | | | 100–200 cm | | |
| Cubist | 0.52 | 0.99 | 1.46 | 0.32 | 1.19 | 1.07 | 0.23 | 1.22 | 1.11 |
| XGBoost | 0.61 | 0.95 | 1.49 | 0.36 | 1.12 | 1.12 | 0.27 | 1.15 | 1.16 |
| RF | 0.61 | 0.92 | 1.51 | 0.38 | 1.08 | 1.14 | 0.26 | 1.14 | 1.15 |
| ANN | 0.57 | 0.97 | 1.46 | 0.33 | 1.16 | 1.08 | 0.24 | 1.18 | 1.13 |
| AvNNet | 0.62 | 0.96 | 1.50 | 0.36 | 1.15 | 1.11 | 0.28 | 1.16 | 1.17 |
| DNN | 0.66 | 0.93 | 1.52 | 0.54 | 1.09 | 1.15 | 0.44 | 1.08 | 1.24 |
| Stack1 | 0.72 | 0.91 | 1.57 | 0.55 | 1.06 | 1.20 | 0.47 | 1.04 | 1.29 |
| Stack2 | 0.70 | 0.89 | 1.58 | 0.54 | 1.06 | 1.18 | 0.49 | 1.02 | 1.29 |
| Stack3 | 0.71 | 0.86 | 1.59 | 0.60 | 1.00 | 1.22 | 0.51 | 0.97 | 1.34 |
| Stack4 | 0.74 | 0.85 | 1.61 | 0.60 | 0.97 | 1.27 | 0.54 | 0.97 | 1.36 |

$R^2$: coefficient of determination; RMSE: root mean square error; RPIQ: Ratio of Performance to InterQuartile distance; Stack: refers to the figure 2.

The performance of XGBoost at both sites and six depth intervals closely followed the performance of RF (Tables 5 and 6). In terms of $R^2$, RMSE and RPIQ it outperformed Cubist. In agreement with our findings, Tziachris et al. [104] reported the reasonable accuracy of XGBoost in comparison with RF models to predict SOC in Greece. There are several other examples of DSM experts who applied XGBoost and RF models successfully to predict soil nutrient in Sub-Saharan Africa [105], soil properties in United States [106], soil pH in China [107], soil properties at the global scale [108] and the depth to bedrock at the global scale [109].

Although Cubist resulted in relatively good predictions of SOC content at the two study sites, especially at the surface layers (Tables 5 and 6), they were outperformed by RF and XGBoost. In line with our results, Zeraatpisheh et al. [21] revealed that, in terms of $R^2$ and RMSE, Cubist was outperformed by RF. Despite, the usefulness of Cubist in explaining the relationships between soil properties and covariates and in modeling SOC content, which

has been reported in several studies [93], our findings showed that the model was not very competitive with the other ML models. Nevertheless, we noted that the differences in the ML models' performance at both sites and at all depth intervals were rather small.

For the two areas, the performance of classical ANN at all depth intervals closely followed the performance of AvNNet (Tables 5 and 6). The higher performance of AvNNet, compared to ANN was also reported by Baker and Ellison [84], who evaluated and compared the performance of AvNNet and ANN in order to predict water retention data. They indicated that combining ANNs improves the ability to generalize individual component ANNs. Similarly, Meyer et al. [83] reported the higher performance of AvNNet by comparing it with three ML models —RF, ANN, and SVM— for rainfall area detection and rainfall rate assignment over Germany. Our result, however, is different from that of Taghizadeh-Mehrjardi et al. [7] who found superior performance for ANNs compare to RF for the three-dimensional mapping of SOC content in the western parts of Iran.

In the arid site, the DNN was able to account for 39 to 83% of the total variation of SOC content from the lower depth interval (100 to 200 cm) to the surface layer (0-5 cm), respectively (Table 5). For the sub-humid site (Table 6), the DNN showed the best performance to described SOC patterns. The DNN was able to account 44 to 81% of total SOC content variation at the depths of 100-200 cm and 0-5 cm, respectively. Our results are in line with other studies in the ML literature that reported the capability of the DNN model to reveal and learn the non-linear and complex patterns underlain datasets [32]. In soil science literature, however, the superior performance of DNN in predicting soil properties is only reported in a few studies [32]. For instance, Behrens et al. [31] found the most accurate results for DSM analysis using deep learning, indicating an improvement of 4–7% compared to RF. Additionally to this example, Padarian et al. [33] and Wadoux et al. [34] successfully applied a CNN model (a well-known DNN model) to predict different soil properties (e.g. SOC) from large spectroscopic databases. Similar to our results, they also specified the CNN models performed better than other ML models.

**Performances of the stacking ensemble models**

The performances of two major groups of stacking ensemble models, namely the standard mode (Stack1 and Stack2) and the rescan mode (Stack3 and Stack4), for prediction of SOC content at six depth intervals of arid and sub-humid areas are presented in Tables 5 and 6, followed as Stack4 > Stack3 > Stack2 > Stack1. At the two sites, in terms of $R^2$, RMSE and RPIQ, the Stack4 model followed closely by Stack3 yielded the most satisfactory results, whereas Stack1 showed the worst performance at six depth intervals. Stack4 had the lowest RMSE as well as the highest $R^2$ and RPIQ values; hence, it was the best stacking method in comparison to the other three stacking ensemble methods. For instance, the RMSEs for

SOC content prediction in 0-5 cm of soil profiles of arid site and sub-humid site by Stack4 models were 17 and 5%, respectively, less than that obtained by the Stack1 (Tables 5 and 6). Similarly, Stack4 reduced the RMSEs (23 and 6%) for SOC content prediction in 100-200 cm of soil profiles of the arid site and sub-humid site, compared to the Stack1 models. Generally, stacking ensemble models in rescan mode had higher accuracy than the one in the standard mode [68]. This indicated that meta-learning models (LASSO and SVR) used in level 1 can recapture and extract some missing information from the original datasets.

As an additional visual analysis, we illustrated the performances of the stacking and individual models in Figure 6. The graph shows that, apart from a bit discordance, the stacking ensemble modeling in both modes (standard mode and rescan mode) indicated the higher performance in comparison to the individual models. Here, we compare the performances of the best individual model (DNN) and the best stacking model (Stack4), in brief in terms of $R^2$ and RMSE. What can be clearly seen in Figure 6 is that the Stack4 ensemble models increased $R^2$ values and decreased the RMSE values in comparison to the DNN models. For instance, the RMSEs for SOC content prediction in 0-5 cm of soil profiles of the arid site and sub-humid site by Stack4 models were 17 and 9%, respectively, less than that obtained by the DNN models. Similarly, Stack4 reduced the RMSEs (28 and 10%) for SOC content prediction in 100-200 cm of soil profiles of the arid site and sub-humid site, compared to the DNN models.

Generally, Stack4 ensemble models exhibited the best competence for capturing the spatial variation of SOC content and reducing prediction uncertainty as well. This indicated the stacking ensemble models in level 1 were successful to keep the advantages and to discard the inaccurate aspect of the individual ML models in level 0. This is justified by the fact that the information lost by the models in level 0 is successfully captured by the level 1 models. In fact, the stacking methods used multiple learning algorithms strengths to obtain better predictive performance and make the predictive model more robust than it is from the individual models. Similarly, several studies have revealed that ensemble models exhibited the best performances for predicting soil properties in the DSM community [38,68]. For example, Zhou et al. [39] and Chen et al. [40] recently evaluate the efficacy of the ensemble models to predict the spatial variation of soil properties. Similarly, Tajik et al. [38] found that stack modeling showed better performance to predict SOC content in comparison to the individual models including RF and SVM.

**Performances of ML models in two different climatic regions**

Vertical distribution of $R^2$, nRMSE and RPIQ values to a depth of 200 cm are depicted in Figures 7. Generally, the two sites showed a decreasing trend in $R^2$ and RPIQ values with increasing depth. Otherwise, the percentage of variation in SOC content, which is described by the models, decreased with increasing depth. A reverse trend for nRMSE was also revealed

(Figure 7). Results indicate that models' performance decreased by each depth increment down the soil profile, and confirmed that models had much better prediction efficiency for surface layers than subsurface layers. Increasing uncertainty of SOC content with depth has been reported in numerous studies [24]. For instance, Laub et al. [110] found that the efficiency of the ML models used for SOC prediction in China decreased from about 0.8 in the topsoil to 0.2 at 0.8 to 1 m subsoil depth. A similar pattern of uncertainty variation with depth was reported in several other studies [22]. This decreasing trend in performances could be explained by the fact that most of terrain- and RS-based covariates that are used as predictors of SOC content (listed in Table 2) explain soil surface features [9].

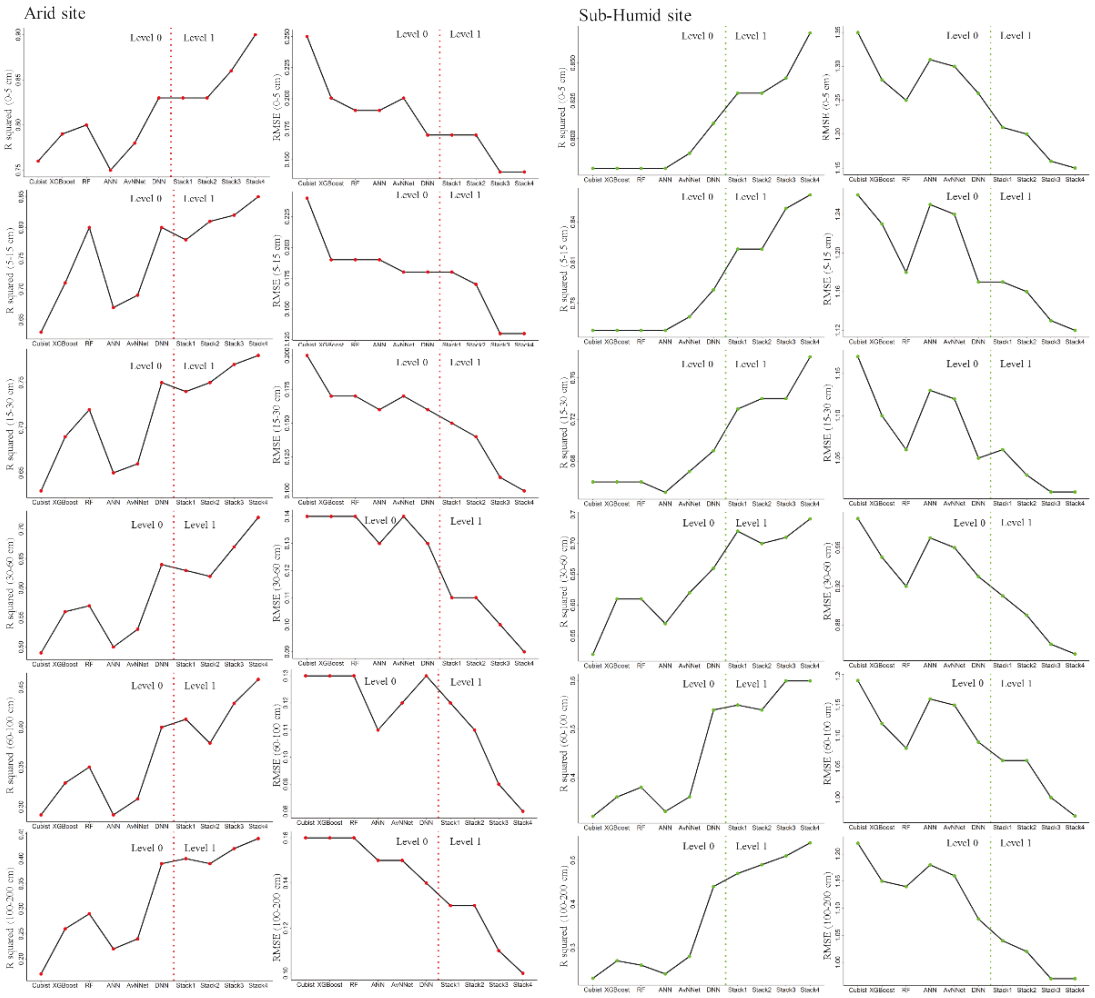


**Figure 6**. $R^2$ and RMSE values of the individual models (Level 0) and stacking models (Level 1) for SOC content at six standard depths in two regions.

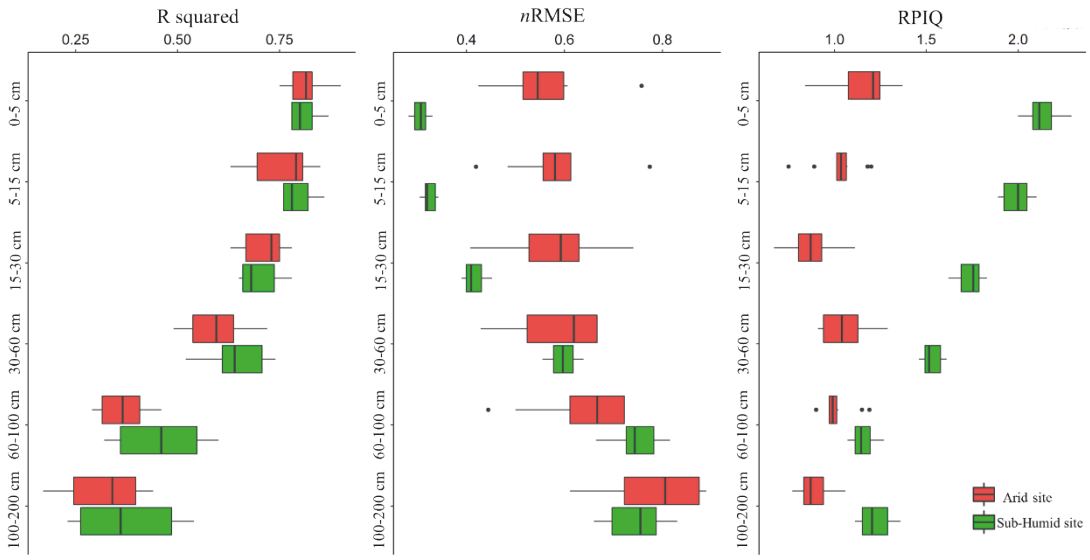


**Figure 7.** Comparison of prediction power of ML models for SOC content at six standard depths in two regions.

According to Figure 7, overall, in terms of $R^2$, nRMSE and RPIQ, all tested ML models (The individual models and stacking models) performed better for the six soil depths at the sub-humid site than for those of the arid site. Here, we compare the performances of the best (Stack4), the worst (Cubist) and AvNNet as a model with intermediate performance, in terms of $R^2$, nRMSE, and RPIQ.

Stack4 resulted in $R^2$ values on average of 0.73 and 0.69 for the sub-humid site and arid site, respectively. $R^2$ values for Cubist were 0.55 and 0.50 and for AvNNet 0.58 and 0.54 for the same areas as for Stack4. Our results, furthermore, indicated the Stack4, Cubist and AvNNet resulted in $R^2$ values for the sub-humid site that were ~9%, ~8% and ~5% higher than the values of arid site. The difference in the performance of ML models is much more evident when we consider nRMSE values, in which nRMSE values obtained by Stack4, Cubist and AvNNet for the arid site were ~25%, ~17% and ~2% more than those values for the sub-humid site (Figure 7). These results indicated that the ML models performed better at six depth intervals in the sub-humid site in comparison to those obtained in the arid site. These results could be partly attributed to large differences between areas in terms of soil forming factors, which results in complex relationships between SOC content and covariates.

The ML models resulted in a decreasing and increasing trend in $R^2$, nRMSE, and RPIQ, respectively, with depth in the two areas (Figure 7). As can be seen, at the top layers (0-60 cm), the arid site tended to have the highest values of nRMSE, but with increasing depth the accuracy of models in terms of nRMSE tended to be almost the same at both sites. This further shows that ML models, such as Stack4 based on the covariates used in this study, cannot capture SOC content variability at the bottom of soil profiles [7,54]. This is consistent with

the results of other researchers [7,17,22,54] who all reported the accuracy of DSM decreased with soil depth. This indicates that the biggest uncertainty is driven by covariate space not by the selection of ML algorithm. It should be added that the typical covariates used in DSM studies are RS- and terrain-based covariates [10], which can only characterize soil properties at the land surface.

**Spatial distribution of SOC**

The spatial distribution of mean, upper and lower limits of SOC contents at the arid site at six interval depths is depicted in Figure 8. A decreasing trend in SOC content down the soil profile was observed. Central parts of the area tended to have the largest amounts of SOC content, which correspond to the cultivated areas mainly under pistachio orchards and wheat. Moreover, the topography of this area is mainly flat and located downslope, which results in more accumulation of fine textured materials and water. In the arid Ardakan site (arid region), because of rainfall scarcity, irrigation is necessary to provide soil moisture for crop production. Thus, irrigated farming and topographic attributes in the central parts promote more vegetation and consequently more organic matter is accumulated in the soil. Lower SOC content in the other parts of the area can be attributed to the higher slope degree, which makes these areas prone to erosion and to higher water discharge. Further, water scarcity in these areas is not compensated by irrigation. In line with our results, Wiesmeier et al. [52] suggested that at the small scales with similar climatic conditions, vegetation, land use and land management have a significant influence on the level of SOC stocks. At the regional scale, climatic effects may be counterbalanced by agricultural practices (e.g., fertilization, irrigation) [49].

The spatial distribution of mean, upper and lower limits of SOC contents at the sub-humid site at six interval depths is depicted in Figure 9. Again, a decreasing trend in SOC content with depth is shown in the sub-humid Rasht site. The map of the spatial distribution of SOC content in the upper layer revealed more SOC accumulation in the northern parts than the other sections. The low slope degrees of the northern parts make these areas favorable for more water accumulation and, thus, result in poorly drained soils. SOC content is more accumulated and less decomposed in poorly drained soils. Mishra et al. [111] reported that high SOC stocks were found in areas characterized by low slope gradient and poorly drained soils. Wiesmeier et al. [52] indicated that areas with low slope degree and concave surface favor water accumulation. Soil moisture, which is largely controlled by terrain attributes, affects the spatial distribution of SOC content [112].

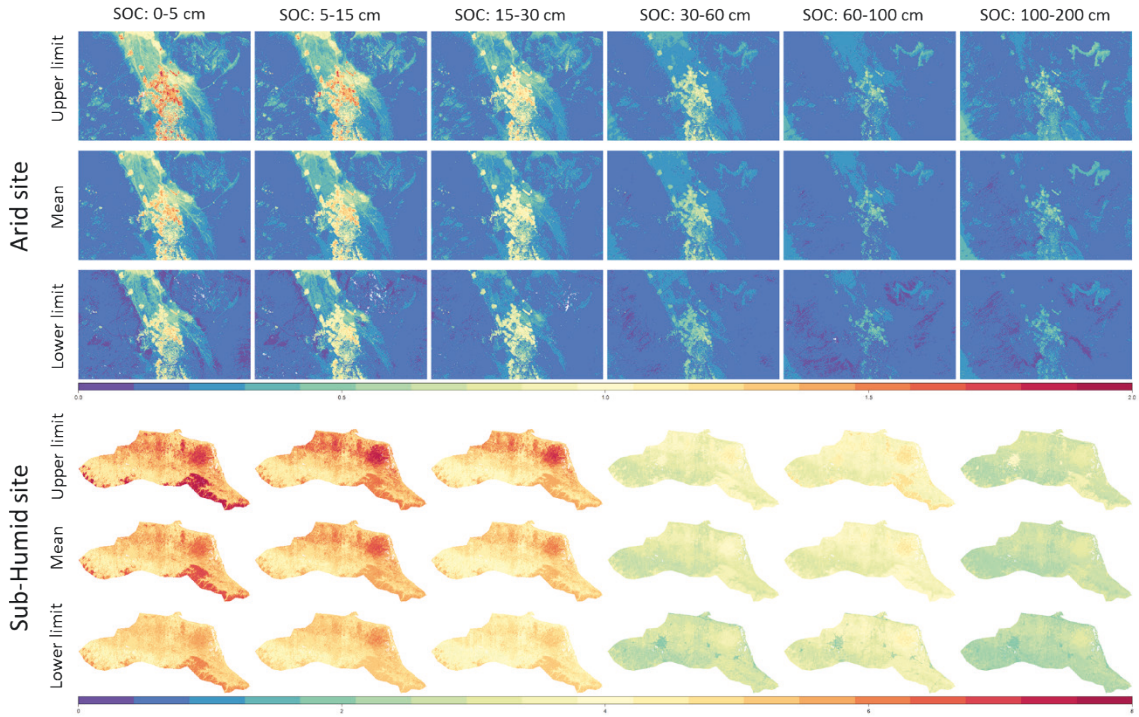


**Figure 8.** The spatial distribution of mean, upper and lower limits of SOC contents at six standard depths in two regions. The upper and lower limits were calculated using Mean ± (1.5 × SD) of the prediction values of SOC contents using the spatial block cross-validation.

## Conclusions

In this study, we introduced stacking ML models in two modes (standard mode and rescan mode) in order to improve the spatial prediction of SOC content at two contrasting climatic regions (arid and sub-humid) of Iran. The main conclusions are:

- Though the differences in the ML models' performance at both sites and at all depth intervals were rather small, DNN was identified as the most suitable individual model.
- The stacking ensemble modeling in both modes (standard mode and rescan mode) indicated the higher performance in comparison to the individual models.
- Although, both terrain- and RS-based covariates were important to explain SOC contents at both sites, their explanatory power was different at both sites and at the soil depth intervals.
- The stacking models are able to explain the effect of contrasting climate on SOC content distribution. Higher content of SOC in the sub-humid site and lower content of SOC in the arid site, however local variation is controlled by moisture, terrain, and land use.

## Author Contributions

Conceptualization—R.T.-M., T.B., T.S. and K.S.; methodology— R.T.-M., T.R., R.V., T.B. and K.S.; software—R.T-M; analysis— R.T.-M. and K.S.; investigation— R.T.-M., A.A-C T.B., T.S. and K.S.; data curation—R.T.-M., F.S. and N.D; writing—original draft preparation—R.T.-M., A.A-C., T.R., M.Z., R.V., F.S. and K.S.; visualization—R.T.-M.

## Funding

This research received no external funding.

## Acknowledgments

R.T-M. has been supported by the Alexander von Humboldt Foundation (grant number: Ref3.4-1164573-IRN-GFHERMES-P). K.S., T.R. and T.S. thank the German Research Foundation (DFG) for supporting this research through the Collaborative Research Center (SFB 1070) 'ResourceCultures' (subprojects Z, S and B02). K.S., T.B. and T.S. are also supported by the DFG Cluster of Excellence "Machine Learning - New Perspectives for Science", EXC 2064/1, project number 390727645. R.V. is supported by an Australian Government Research Training Program Scholarship and a Rowden White Scholarship.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

Scholten, T.; Goebes, P.; Kühn, P.; Seitz, S.; Assmann, T.; Bauhus, J.; Bruelheide, H.; Buscot, F.; Erfmeier, A.; Fischer, M. On the combined effect of soil fertility and topography on tree growth in subtropical forest ecosystems—a study from SE China. Journal of Plant Ecology 2017, 10, 111-127.

Minasny, B.; McBratney, A. Limited effect of organic matter on soil available water capacity. European Journal of Soil Science 2018, 69, 39-47.

Don, A.; Schumacher, J.; Scherer-Lorenzen, M.; Scholten, T.; Schulze, E.-D. Spatial and vertical variation of soil carbon at two grassland sites—implications for measuring soil carbon stocks. Geoderma 2007, 141, 272-282.

Adhikari, K.; Owens, P.R.; Libohova, Z.; Miller, D.M.; Wills, S.A.; Nemecek, J. Assessing soil organic carbon stock of Wisconsin, USA and its fate under future land use and climate change. Science of the Total Environment 2019, 667, 833-845.

Minasny, B.; McBratney, A.; Malone, B.; Wheeler, I. Digital soil mapping of carbon. Adv. Agron 2013, 118, 1-47.

## Appendix

Ajami, M.; Heidari, A.; Khormali, F.; Gorji, M.; Ayoubi, S. Environmental factors controlling soil organic carbon storage in loess soils of a subhumid region, northern Iran. Geoderma 2016, 281, 1-10.

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

Bernhard, N.; Moskwa, L.-M.; Schmidt, K.; Oeser, R.A.; Aburto, F.; Bader, M.Y.; Baumann, K.; von Blanckenburg, F.; Boy, J.; van den Brink, L. Pedogenic and microbial interrelations to regional climate and local topography: new insights from a climate gradient (arid to humid) along the Coastal Cordillera of Chile. Catena 2018, 170, 335-355.

Zeraatpisheh, M.; Jafari, A.; Bodaghabadi, M.B.; Ayoubi, S.; Taghizadeh-Mehrjardi, R.; Toomanian, N.; Kerry, R.; Xu, M. Conventional and digital soil mapping in Iran: Past, present, and future. Catena 2020, 188, 104424.

Ma, Y.; Minasny, B.; Malone, B.P.; McBratney, A.B. Pedology and digital soil mapping (DSM). European journal of soil science 2019, 70, 261-235.

McBratney, A.B.; Stockmann, U.; Angers, D.A.; Minasny, B.; Field, D.J. Challenges for soil organic carbon research. In Soil carbon, Springer: 2014; pp. 3-16.

Minasny, B.; McBratney, A.B.; Lark, R.M. Digital soil mapping technologies for countries with sparse data infrastructures. In Digital soil mapping with limited data, Springer: 2008; pp. 15-30.

Behrens, T.; Scholten, T. Digital soil mapping in Germany—a review. Journal of Plant Nutrition and Soil Science 2006, 169, 434-443.

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

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

Moreno, R.; Irigoyen, A.I.; Monterubbianesi, M.G.; Studdert, G.A. Application of artificial neural networks to estimate soil organic carbon in a high-organic-matter Mollisol. Spanish Journal of Soil Science (SJSS) 2017, 7.

Rentschler, T.; Gries, P.; Behrens, T.; Bruelheide, H.; Kühn, P.; Seitz, S.; Shi, X.; Trogisch, S.; Scholten, T.; Schmidt, K. Comparison of catchment scale 3D and 2.5 D modelling of soil organic carbon stocks in Jiangxi Province, PR China. PloS one 2019, 14.

Stumpf, F.; Keller, A.; Schmidt, K.; Mayr, A.; Gubler, A.; Schaepman, M. Spatio-temporal land use dynamics and soil organic carbon in Swiss agroecosystems. Agriculture, ecosystems & environment 2018, 258, 129-142.

Wang, B.; Waters, C.; Orgill, S.; Cowie, A.; Clark, A.; Li Liu, D.; Simpson, M.; McGowen, I.; Sides, T. Estimating soil organic carbon stocks using different modelling techniques in the semi-arid rangelands of eastern Australia. Ecological indicators 2018, 88, 425-438.

Were, K.; Bui, D.T.; Dick, Ø.B.; Singh, B.R. 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 2015, 52, 394-403.

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

Amirian-Chakan, A.; Taghizadeh-Mehrjardi, R.; Kerry, R.; Kumar, S.; Khordehbin, S.; Khanghah, S.Y. Spatial 3D distribution of soil organic carbon under different land use types. Environmental monitoring and assessment 2017, 189, 131.

Owusu, S.; Yigini, Y.; Olmedo, G.F.; Omuto, C.T. Spatial prediction of soil organic carbon stocks in Ghana using legacy data. Geoderma 2020, 360, 114008.

Adhikari, K.; Hartemink, A.E.; Minasny, B.; Kheir, R.B.; Greve, M.B.; Greve, M.H. Digital mapping of soil organic carbon contents and stocks in Denmark. PloS one 2014, 9, e105519.

Hinge, G.; Surampalli, R.Y.; Goyal, M.K. Prediction of soil organic carbon stock using digital mapping approach in humid India. Environmental earth sciences 2018, 77, 172.

Dharumarajan, S.; Kalaiselvi, B.; Suputhra, A.; Lalitha, M.; Hegde, R.; Singh, S.; Lagacherie, P. Digital soil mapping of key GlobalSoilMap properties in Northern Karnataka Plateau. Geoderma Regional 2020, 20, e00250.

Zhao, Z.; Yang, Q.; Sun, D.; Ding, X.; Meng, F.-R. Extended model prediction of high-resolution soil organic matter over a large area using limited number of field samples. Computers and Electronics in Agriculture 2020, 169, 105172.

Guo, L.; Fu, P.; Shi, T.; Chen, Y.; Zhang, H.; Meng, R.; Wang, S. Mapping field-scale soil organic carbon with unmanned aircraft system-acquired time series multispectral images. Soil and Tillage Research 2020, 196, 104477.

Yang, L.; He, X.; Shen, F.; Zhou, C.; Zhu, A.-X.; Gao, B.; Chen, Z.; Li, M. Improving prediction of soil organic carbon content in croplands using phenological parameters extracted from NDVI time series data. Soil and Tillage Research 2020, 196, 104465.

Nabiollahi, K.; Eskandari, S.; Taghizadeh-Mehrjardi, R.; Kerry, R.; Triantafalis, J. Assessing soil organic carbon stocks under land-use change scenarios using random forest models. Carbon Management 2019, 10, 63-77.

Behrens, T.; Schmidt, K.; MacMillan, R.A.; Rossel, R.A.V. Multi-scale digital soil mapping with deep learning. Scientific reports 2018, 8.

Ng, W.; Minasny, B.; Malone, B.P.; McBratney, A.B. Pedology and digital soil mapping (DSM). European journal of soil science 2019, 70, 261-235.

Padarian, J.; Minasny, B.; Montazerolghaem, M.; Padarian, J.; Ferguson, R.; Bailey, S.; McBratney, A.B. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. Geoderma 2019, 352, 251-267.

Wadoux, A.M.-C.; Padarian, J.; Minasny, B. Multi-source data integration for soil mapping using deep learning. Soil 2019, 5, 107-119.

Breiman, L. Stacked regressions. Machine learning 1996, 24, 49-64.

Caubet, M.; Dobarco, M.R.; Arrouays, D.; Minasny, B.; Saby, N.P. Merging country, continental and global predictions of soil texture: Lessons from ensemble modelling in france. Geoderma 2019, 337, 99-110.

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

Sagi, O.; Rokach, L. Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2018, 8, e1249.

Ribeiro, M.H.D.M.; dos Santos Coelho, L. Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. Applied Soft Computing 2020, 86, 105837.

Tajik, S.; Ayoubi, S.; Zeraatpisheh, M. Digital mapping of soil organic carbon using ensemble learning model in Mollisols of Hyrcanian forests, northern Iran. Geoderma Regional 2020, e00256.

Zhou, Y.; Xue, J.; Chen, S.; Zhou, Y.; Liang, Z.; Wang, N.; Shi, Z. Fine-Resolution Mapping of Soil Total Nitrogen across China Based on Weighted Model Averaging. Remote Sensing 2020, 12, 85.

Chen, S.; Mulder, V.L.; Heuvelink, G.B.; Poggio, L.; Caubet, M.; Dobarco, M.R.; Walter, C.; Arrouays, D. Model averaging for mapping topsoil organic carbon in France. Geoderma 2020, 366, 114237.

Smith, D. Soil survey staff: keys to soil taxonomy. Washington: Natural Resources Conservation Service 2014.

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

Minasny, B.; McBratney, A.B. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Computers & geosciences 2006, 32, 1378-1388.

Stumpf, F.; Schmidt, K.; Behrens, T.; Schönbrodt-Stitt, S.; Buzzo, G.; Dumperth, C.; Wadoux, A.; Xiang, W.; Scholten, T. Incorporating limited field operability and legacy soil samples in a hypercube sampling design for digital soil mapping. Journal of Plant Nutrition and Soil Science 2016, 179, 499-509.

Gholizadeh, A.; Zizala, D.; Saberioon, M.; Boruvka, L. Soil organic carbon content monitoring and mapping using airborne and Sentinel-2 spectral imaging. In Proceedings of Sixth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2018); p. 1077319.

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

Goidts, E.; Wesemael, B.V.; Van Oost, K. Driving forces of soil organic carbon evolution at the landscape and regional scale using data from a stratified soil monitoring. Global change biology 2009, 15, 2981-3000.

Roozitalab, M.H.; Toomanian, N.; Dehkordi, V.R.G.; Khormali, F. Major Soils, Properties, and Classification. In The Soils of Iran, Springer: 2018; pp. 93-147.

Taghizadeh-Mehrjardi, R.; Neupane, R.; Sood, K.; Kumar, S. Artificial bee colony feature selection algorithm combined with machine learning algorithms to predict vertical and lateral distribution of soil organic matter in South Dakota, USA. Carbon Management 2017, 8, 277-291.

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

Nelson, D.W.; Sommers, L.E. Total carbon, organic carbon, and organic matter. Methods of soil analysis: Part 3 Chemical methods 1996, 5, 961-1010.

Malone, B.P.; McBratney, A.; Minasny, B.; Laslett, G. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 2009, 154, 138-152.

Arrouays, D.; Grundy, M.G.; Hartemink, A.E.; Hempel, J.W.; Heuvelink, G.B.; Hong, S.Y.; Lagacherie, P.; Lelyk, G.; McBratney, A.B.; McKenzie, N.J. GlobalSoilMap: Toward a fine-resolution global grid of soil properties. In Advances in agronomy, Elsevier: 2014; Vol. 125, pp. 93-134.

Xiong, X.; Grunwald, S.; Myers, D.B.; Ross, C.W.; Harris, W.G.; Comerford, N.B. Interaction effects of climate and land use/land cover change on soil organic carbon sequestration. Science of the Total Environment 2014, 493, 974-982.

Farr, T.G.; Rosen, P.A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L. The shuttle radar topography mission. Reviews of geophysics 2007, 45.

Conrad, O.; Olaya, V. SAGA-GIS module library documentation (v2. 2.3). Module Valley Depth 2012.

Wulder, M.A.; White, J.C.; Loveland, T.R.; Woodcock, C.E.; Belward, A.S.; Cohen, W.B.; Fosnight, E.A.; Shaw, J.; Masek, J.G.; Roy, D.P. The global Landsat archive: Status, consolidation, and direction. Remote Sensing of Environment 2016, 185, 271-283.

Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote sensing of Environment 2012, 120, 25-36.

Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment 2017, 202, 18-27.

Fongaro, C.; Demattê, J.; Rizzo, R.; Lucas Safanelli, J.; Mendes, W.; Dotto, A.; Vicente, L.; Franceschini, M.; Ustin, S. Improvement of clay and sand quantification based on a novel approach with a focus on multispectral satellite images. Remote Sensing 2018, 10, 1555.

Poggio, L.; Lassauce, A.; Gimona, A. Modelling the extent of northern peat soil and its uncertainty with Sentinel: Scotland as example of highly cloudy region. Geoderma 2019, 346, 63-74.

Huete, A.R. A soil-adjusted vegetation index (SAVI). Remote sensing of environment 1988, 25, 295-309.

Kursa, M.B.; Rudnicki, W.R. Feature selection with the Boruta package. J Stat Softw 2010, 36, 1-13.

Liaw, A.; Wiener, M. Classification and regression by randomForest. R news 2002, 2, 18-22.

Keskin, H.; Grunwald, S.; Harris, W.G. Digital mapping of soil carbon fractions with machine learning. Geoderma 2019, 339, 40-58.

Wang, R. Significantly improving the prediction of molecular atomization energies by an ensemble of machine learning algorithms and rescanning input space: A stacked generalization approach. The Journal of Physical Chemistry C 2018, 122, 8868-8873.

Quinlan, J.R. Learning with continuous classes. In Proceedings of 5th Australian joint conference on artificial intelligence; pp. 343-348.

Kuhn, M.; Johnson, K. Applied predictive modeling; Springer: 2013; Vol. 26.

Minasny, B.; McBratney, A.B. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. Chemometrics and intelligent laboratory systems 2008, 94, 72-79.

Kuhn, M.; Weston, S.; Keefer, C.; Kuhn, M.M. Package 'Cubist'. 2018.

Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. Classification and regression trees. Wadsworth Int. Group 1984, 37, 237-251.

Behrens, T.; Schmidt, K.; Zhu, A.-X.; Scholten, T. The ConMap approach for terrain-based digital soil mapping. European Journal of Soil Science 2010, 61, 133-143.

Hounkpatin, K.O.; Schmidt, K.; Stumpf, F.; Forkuor, G.; Behrens, T.; Scholten, T.; Amelung, W.; Welp, G. Predicting reference soil groups using legacy data: A data pruning and Random Forest approach for tropical environment (Dano catchment, Burkina Faso). Scientific reports 2018, 8, 9959.

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

Peters, J.; De Baets, B.; Verhoest, N.E.; Samson, R.; Degroeve, S.; De Becker, P.; Huybrechts, W. Random forests as a tool for ecohydrological distribution modelling. ecological modelling 2007, 207, 304-318.

Elith, J.; Leathwick, J.R.; Hastie, T. A working guide to boosted regression trees. Journal of Animal Ecology 2008, 77, 802-813.

Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y. Xgboost: extreme gradient boosting. R package version 0.4-2 2015, 1-4.

Ripley, B.; Venables, W.; Ripley, M. Package 'nnet'. R package version. 2016.

Ripley, B.D. Pattern recognition and neural networks; Cambridge university press: 2007.

Hagan, M.T.; Menhaj, M.B. Training feedforward networks with the Marquardt Algorithm. IEEE transactions on Neural Networks 1994, 5, 989-993.

Meyer, H.; Kühnlein, M.; Appelhans, T.; Nauss, T. Comparison of four machine learning algorithms for their applicability in satellite-based optical rainfall retrievals. Atmospheric research 2016, 169, 424-433.

Baker, L.; Ellison, D. Optimisation of pedotransfer functions using an artificial neural network ensemble method. Geoderma 2008, 144, 212-224.

Kuhn, M.; Wing, J.; Weston, S. Package 'caret'. Classification and regression training 2015.

Hinton, G.E.; Osindero, S.; Teh, Y.-W. A fast learning algorithm for deep belief nets. Neural computation 2006, 18, 1527-1554.

Shen, C. A transdisciplinary review of deep learning research and its relevance for water resources scientists. Water Resources Research 2018, 54, 8558-8593.

Suthaharan, S. Big data analytics: Machine learning and Bayesian learning perspectives—What is done? What is not? Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2019, 9, e1283.

Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research 2014, 15, 1929-1958.

Zhai, J.; Zang, L.; Zhou, Z. Ensemble dropout extreme learning machine via fuzzy integral for data classification. Neurocomputing 2018, 275, 1043-1052.

Candel, A.; Parmar, V.; LeDell, E.; Arora, A. Deep learning with H2O. H2O. ai Inc 2016.

Cortes, C.; Vapnik, V. Support-vector networks. Machine learning 1995, 20, 273-297.

Drucker, H.; Burges, C.J.; Kaufman, L.; Smola, A.J.; Vapnik, V. Support vector regression machines. In Proceedings of Advances in neural information processing systems; pp. 155-161.

Dimitriadou, E.; Hornik, K.; Leisch, F.; Meyer, D.; Weingessel, A.; Leisch, M.F. The e1071 package. Misc Functions of Department of Statistics (e1071), TU Wien 2006.

Friedman, J.; Hastie, T.; Tibshirani, R. glmnet: Lasso and elastic-net regularized generalized linear models. R package version 2009, 1.

Bischl, B.; Lang, M.; Kotthoff, L.; Schiffner, J.; Richter, J.; Studerus, E.; Casalicchio, G.; Jones, Z.M. mlr: Machine Learning in R. The Journal of Machine Learning Research 2016, 17, 5938-5942.

Valavi, R.; Elith, J.; Lahoz-Monfort, J.J.; Guillera-Arroita, G. blockCV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. Methods in Ecology and Evolution 2019, 10, 225–232.

Khaledian, Y.; Miller, B.A. Selecting appropriate machine learning methods for digital soil mapping. Applied Mathematical Modelling 2019.

Nawar, S.; Mouazen, A.M. Predictive performance of mobile vis-near infrared spectroscopy for key soil properties at different geographical scales by using spiking and data mining techniques. Catena 2017, 151, 118-129.

Jafari, M.; Chahouki, M.Z.; Tavili, A.; Azarnivand, H.; Amiri, G.Z. Effective environmental factors in the distribution of vegetation types in Poshtkouh rangelands of Yazd Province (Iran). Journal of Arid Environments 2004, 56, 627-641.

Goebes, P.; Schmidt, K.; Seitz, S.; Both, S.; Bruelheide, H.; Erfmeier, A.; Scholten, T.; Kühn, P. The strength of soil-plant interactions under forest is related to a Critical Soil Depth. Scientific reports 2019, 9, 8635.

Wang, B.; Waters, C.; Orgill, S.; Gray, J.; Cowie, A.; Clark, A.; Li Liu, D. High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern Australia. Science of The Total Environment 2018, 630, 367-378.

Vaudour, E.; Gomez, C.; Fouad, Y.; Lagacherie, P. Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems. Remote sensing of environment 2019, 223, 21-33.

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

Hengl, T.; Leenaars, J.G.; Shepherd, K.D.; Walsh, M.G.; Heuvelink, G.B.; Mamo, T.; Tilahun, H.; Berkhout, E.; Cooper, M.; Fegraus, E. Soil nutrient maps of Sub-Saharan Africa: assessment of soil nutrient content at 250 m spatial resolution using machine learning. Nutrient Cycling in Agroecosystems 2017, 109, 77-102.

Ramcharan, A.; Hengl, T.; Nauman, T.; Brungard, C.; Waltman, S.; Wills, S.; Thompson, J. Soil property and class maps of the conterminous United States at 100-meter spatial resolution. Soil Science Society of America Journal 2018, 82, 186-201.

Chen, S.; Liang, Z.; Webster, R.; Zhang, G.; Zhou, Y.; Teng, H.; Hu, B.; Arrouays, D.; Shi, Z. A high-resolution map of soil pH in China made by hybrid modelling of sparse soil data and environmental covariates and its implications for pollution. Science of the Total Environment 2019, 655, 273-283.

Hengl, T.; de Jesus, J.M.; Heuvelink, G.B.; Gonzalez, M.R.; Kilibarda, M.; Blagotić, A.; Shangguan, W.; Wright, M.N.; Geng, X.; Bauer-Marschallinger, B. SoilGrids250m: Global gridded soil information based on machine learning. PLoS one 2017, 12, e0169748.

Shangguan, W.; Hengl, T.; de Jesus, J.M.; Yuan, H.; Dai, Y. Mapping the global depth to bedrock for land surface modeling. Journal of Advances in Modeling Earth Systems 2017, 9, 65-88.

Laub, M.; Blagodatsky, S.; Lang, R.; Yang, X.; Cadisch, G. A mixed model for landscape soil organic carbon prediction across continuous profile depth in the mountainous subtropics. Geoderma 2018, 330, 177-192.

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

Vasques, G.; Grunwald, S.; Comerford, N.; Sickman, J. Regional modelling of soil carbon at multiple depths within a subtropical watershed. Geoderma 2010, 156, 326-336.

## Appendix

**Manuscript 4**

# 3D soil quality mapping and interpretability of scales in Andalusia, Spain

*(in preparation for Scientific Reports)*

Tobias Rentschler[1,2,*], Thorsten Behrens[2], Thomas Scholten[1,2], Sandra Teuber[1,2] and Karsten Schmidt[1,2,3]

[1] SFB 1070 ResourceCultures, University of Tübingen, D-72074 Tübingen, Germany
[2] Department of Geosciences, Soil Science and Geomorphology, University of Tübingen, D-72074 Tübingen, Germany
[3] eScience-Center of the Eberhard Karls Universität Tübingen, University of Tübingen, D-72074 Tübingen, Germany

*Corresponding author:
E-Mail: t.rentschler@uni-tuebingen.de

## Abstract

Soils and soil functions are recognized as a key resource for human well-being throughout time. In an agricultural and forestry perspective, soil functions contribute to food and timber production. Other soil functions are related to freshwater security and energy provisioning. In general, the capacity of a soil to function within specific boundaries is summarised as soil quality. Knowledge about the spatial distribution of soil quality is crucial for sustainable land use and the protection of soils and their functions. This spatial knowledge can be obtained with accurate and efficient machine-learning-based soil mapping approaches, which allow the estimation of the soil quality at distinct locations. However, the vertical distribution of soil properties is usually neglected when assessing soil quality at distinct locations. To overcome such limitations, the depth function of soil properties needs to be incorporated in the modelling. This is not only important to get a better estimation of the overall soil quality throughout the rooting zone, but also to identify factors that limit plant growth, such as strong acidity or alkalinity, and the water holding capacity. Thus, the objective of this study was to model and map the soil quality indicators pH, soil organic carbon, sand, silt and clay content as a volumetric entity. The study area is located in southern Spain in the Province of Seville at the Guadalquivir river. It covers 1,000 km² of farmland, citrus and olive plantations, pastures and wood pasture (Dehesa) in the Sierra Morena mountain range, at the Guadalquivir flood plain and tertiary terraces. Soil samples were taken at 130 soil profiles in five depths (or less at shallow soils). The profiles were randomly stratified depending on slope position and land cover. We used a subset of 99 samples from representative soil profiles to assess the overall 513 samples with FT-IR spectroscopy and machine learning methods to model equal-area spline, polynomial and exponential depth functions for each soil quality indicator at each of the 130 profiles. These depth functions were modelled and predicted spatially with a comprehensive set of environmental covariates from remote sensing data, multi-scale terrain analysis and geological maps. By solving the spatially predicted depth functions with a vertical resolution of 5 cm, we obtained a volumetric, i.e. three-dimensional, map of pH, soil organic carbon content and soil texture. Preliminary results are promising for volumetric soil mapping and the estimation of soil quality and limiting factors in three-dimensional space.

## Introduction

Soils are important for agriculture, timber, biofuels[1]. Growing world population requires changes in food production towards sustainability through policies and management practices[1,2]. According to Godfray et al.[1] one part of this is closing the so-called yield gap, which is the difference between realised productivity and the potentially best productivity.

## Appendix

Spatial knowledge of soil conditions can support management practices to increase the realised productivity towards the potential productivity of soils which depends, amongst others, on water, nutrients, and knowledge to use and preserve the soil[1]. This knowledge can be achieved with sophisticated soil models for spatial prediction of soil properties based on machine learning.

Soil quality is used to assess the status and use potential of soil for specific types of land use, e.g. agriculture, plantations, pastures or forest. To estimate or calculate soil quality different soil quality indicators have been identified. Soil quality indicators are physical, chemical or biological soil properties. Widely used soil quality indicators are water storage or content ($\theta_{FC}$), soil organic carbon (SOC), pH and cation exchange capacity (CEC), amongst others[3]. For relatively simple soil quality indices these indicators are rated[4], summed up and divided by the number of indicators[5]. This soil quality rating can be modelled spatially with machine learning techniques such as random forests[6].

However, soil is a three-dimensional continuum that requires sophisticated 3D models. In fact, climate, land cover and terrain have influence of varying strength on topsoil and subsoil properties[7,8]. But knowledge about the relevance of independent variables for soil models at different depths, e.g. topsoil, subsurface soil and subsoil is sparce.

## Results

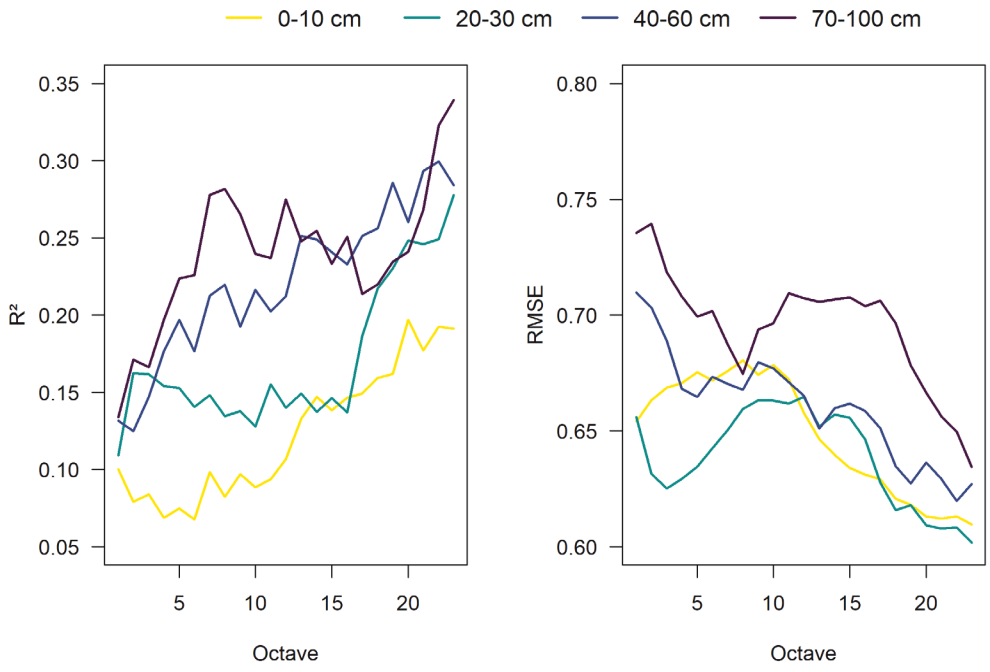


**Figure 1.** Results of the repeated 10-fold cross-validation of the additive model training with sequentially adding scales (octaves).

## Discussion

## Methods

### Study area

The study area of 1,000 km$^2$ is located in the middle Guadalquivir basin, Andalusia, Spain, 50 km NE of Seville. The geology separates the area in three main parts: the Sierra Morena mountain range in the North with Paleozoic granite, gneiss and slate; the Guadalquivir flood plain with Pleistocene marl, calcarenite and coarse sand and Holocene sands and loams; and tertiary terraces in the South[9,10]. The study area is a heterogenous agricultural landscape with arable land, citrus and olive plantations, pastures and Dehesa, an agrosylvopastoral system. Occurring soil types according to USDA Soil Taxonomy are Alfisols, Entisols, Inceptisols and Vertisols[11].

### Soil data and environmental covariates

Soil samples were taken at 130 stratified random locations in October 2018. The strata are combinations of four geomorphic positions (geomorphons; namely footslope, slope and shoulder)[12] and for the CORINE Land Cover level 2 classes "arable land", "permanent crops", "pastures", "forest" and "shrub and/or herbaceous vegetation associations"[13], which are most common in the area. The point density of the sampled locations is proportional to the stratum area with a minimum of 3 samples for the smallest stratum. At each location up to fife samples bulked from 3 replicates were taken depending on the soil thickness. The sampled increments were 0-10, 10-20, 20-30, 40-60 and 70-100 cm.

For lab analysis the samples were dried at 40 ℃ for 24 h, root fragments were removed, sieved (<2 mm) and ground. The soil spectra of all 506 samples were measured with a Tensor II (Bruker Optics, Ettlingen, Germany) for NIR (833–2500 nm) and a GladiATR (Pike Technlogies, Madison, WI, USA) for MIR (2500–25.000 nm). A subset of 97 samples representative for the strata was analysed for soil organic carbon (SOC) and total Nitrogen (Ntot) content (%) with a Vario EL III (Elementar, Hanau, Germany), for pH with a pH meter in KCl solution ($pH_{KCl}$) and for grain size fractions clay (<2 µm), silt (2-50 µm) and sand (50-2000 µm) with a SediGraph III (Micromeritics, Norcross, GA, USA). The 97 samples were used as dependent variables to train PLS models[14] (see below) and make predictions to the remaining samples according to Hobley and Prater[15]. The root mean squared errors of these models were 0.5 % SOC, 0.02 %Ntot, 0.4 $pH_{KCl}$, 4 % clay, 5 % silt and 5 % sand content.

## Appendix

Water content at field capacity ($\theta_{FC}$ in $cm^3 cm^{-3}$) was calculated with a pedo-transfer function developed for Europe according to Tóth et al.[16] and the cation exchange capacity (CEC in cmol $kg^{-1}$) was calculated after an equation by Khaledian et al.[17] developed for spain. Errors (RMSE) of these pedo-transfer functions were reported to be 0.06 $cm^3$ $cm^{-3}$ and 0.73 cmol $kg^{-1}$, respectively.

**Random forests**

Random forests were developed by Breiman[18] as an ensemble of classification and regression trees[19]. Binary splits are used for a single decision tree to homogenize the predictor variables according to the dependent variable, thus minimizing the node impurity. Random forests use a bootstrap approach, where random predictor variables are chosen at each split of a tree. The final regression model results from averaging all decision tree outputs[18]. Random forests are robust against overfitting and interpretable[18].

**Modelling**

## Data Availability

Terrain data is published under the CC-BY 4.0 license by Centro Nacional de information Geográfica (CNIG) of the Spanish government; last accessed March, 31st 2020 http://centrodedescargas.cnig.es/CentroDescargas/locale?request_locale=en#. Landsat images were accessed and pre-reprocessed via the Google Earth Engine [20,21]. The soil data is published at PANGAEA.

## References

Godfray, H. C. J. et al. Food Security: The Challenge of Feeding 9 Billion People. Science 327, 812–818; 10.1126/science.1185383 (2010).

Bouma, J. Soil science contributions towards Sustainable Development Goals and their implementation: linking soil functions with ecosystem services. J. Plant Nutr. Soil Sci. 177, 111–120; 10.1002/jpln.201300646 (2014).

Bünemann, E. K. et al. Soil quality – A critical review. Soil Biology and Biochemistry 120, 105–125; 10.1016/j.soilbio.2018.01.030 (2018).

Hazelton, P. & Murphy, B. Interpreting soil test results. What do all the numbers mean? (Csiro Publishing, Clayton South Vic, 2016).

Pulido, M., Schnabel, S., Contador, J. F. L., Lozano-Parra, J. & Gómez-Gutiérrez, Á. Selecting indicators for assessing soil quality and degradation in rangelands of Extremadura (SW Spain). Ecological Indicators 74, 49–61; 10.1016/j.ecolind.2016.11.016 (2017).

Grimm, R., Behrens, T., Märker, M. & Elsenbeer, H. Soil organic carbon concentrations and stocks on Barro Colorado Island — Digital soil mapping using Random Forests analysis. Geoderma 146, 102–113; 10.1016/j.geoderma.2008.05.008 (2008).

Jobbágy, E. G. & Jackson, R. B. The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecological Applications 10, 423–436; 10.1890/1051-0761(2000)010[0423:TVDOSO]2.0.CO;2 (2000).

Rentschler, T. et al. Comparison of catchment scale 3D and 2.5D modelling of soil organic carbon stocks in Jiangxi Province, PR China. PloS one 14, e0220881; 10.1371/journal.pone.0220881 (2019).

Civis, J. et al. in Geológica de España, edited by J. A. Vera (Madrid, 2004), pp. 543–550.

Wolf, D. & Faust, D. Western Mediterranean environmental changes: Evidences from fluvial archives. Quaternary Science Reviews 122, 30–50; 10.1016/j.quascirev.2015.04.016 (2015).

Gómez-Miguel, V. Mapa de Suelos de España (Centro Nacional de Información Geográfica (CNIG), Madrid, 2005).

Jasiewicz, J. & Stepinski, T. F. Geomorphons — a pattern recognition approach to classification and mapping of landforms. Geomorphology 182, 147–156; 10.1016/j.geomorph.2012.11.005 (2013).

Büttner, G. & Kosztra, B. CLC2018Technical Guidelines. Available at https://land.copernicus.eu/user-corner/technical-library/clc2018technicalguidelines_final.pdf (2010).

Breiman, L. Random Forests. Machine Learning 45, 5–32; 10.1023/A:1010933404324 (2001).

Hobley, E. U. & Prater, I. Estimating soil texture from vis-NIR spectra. Eur J Soil Sci 70, 83–95; 10.1111/ejss.12733 (2019).

Tóth, B. et al. New generation of hydraulic pedotransfer functions for Europe. Eur J Soil Sci 66, 226–238; 10.1111/ejss.12192 (2015).

Khaledian, Y. et al. Modeling soil cation exchange capacity in multiple countries. Catena 158, 194–200; 10.1016/j.catena.2017.07.002 (2017).

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

Breiman, L., Friedman, J. H., Stone, C. J. & Olshen, R. A. Classification and Regression Trees (Chapman and Hall, New York, NY, 1984).

Gorelick, N. et al. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment 202, 18–27; 10.1016/j.rse.2017.06.031 (2017).

Bindereif, L. et al. Analysis and mapping of spatio-temporal land use dynamics in Andalusia, Spain using the Google Earth Engine cloud computing platform and the Landsat archive. Geophysical Research Abstracts 21, 8168 (2019).

## Appendix


### Acknowledgements

We thank the German Research Foundation (DFG) for supporting this research through the Collaborative Research Centre SFB 1070 ResourceCultures (subprojects A02, S and Z). Further, we are indebted to Laura Bindereif, Matthias Czechowski, Philipp Gries, Susan Mentzer, Renate Nitschke and Carolin Walper for their assistance with field work, lab work and data analysis.


### Author contributions

T.R. and K.S. conceived and designed the study, lead and performed the data analyses, interpretations and writing, T.B. contributed to data analysis, interpretations and writing. T.S. contributed to project administration, interpretation and writing.

### Competing Interests

The authors declare no competing interests.

**Manuscript 5**

# Spatial monitoring of grassland management using multi-temporal satellite imagery

*Ecological Indicators 113, 106201*

Felix Stumpf[1,2,*], Manuel K. Schneider[1], Armin Keller[1], Andreas Mayr[3], Tobias Rentschler[4], Reto G. Meuli[1], Michael Schaepman[2], Frank Liebisch[1,5]

[1] Agroscope, Reckenholzstrasse 191, 8046 Zurich, Switzerland.
[2] Department of Geography, Remote Sensing Laboratories (RSL), University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland.
[3] Institute of Geography, University of Innsbruck, Innrain 52, 6020 Innsbruck, Austria.
[4] Department of Geosciences, Chair of Soil Science and Geomorphology, University of Tübingen, Rümelinstraße 19-23, 72070 Tübingen, Germany.
[5] Department of Environmental Systems Science, Institute of Agricultural Sciences, ETH Zurich, Universitätstrasse 2, 8092 Zurich, Switzerland.

*Corresponding author:
E-mail: felix.stumpf@agroscope.admin.ch

Received: January 8, 2019; Accepted: February 6, 2020; Published: February 17, 2020

# Appendix


## Abstract

Spatial monitoring of grassland management is crucial for ecosystem assessment and the establishment of sustainable agriculture. Switzerland is covered by large areas of small structured grassland parcels differing in management practices and use intensities, making the mapping of grassland management challenging. We present a monitoring tool to map grassland management, distinguishing between mowing- and grazing practice, and between different use intensities for Swiss agroecosystems. By analyzing pixelwise spectral time series of 2015, derived from satellite imagery of the Landsat archive, we estimated the number of management events and biomass productivity. Both estimates were used to map classes of dominant management practices and use intensities following a stepwise clustering approach. The grassland management (GM) classes were evaluated relative to established spectral and topographical patterns of grassland use intensity, and in terms of spatial conformity with available regional land use data. The GM classes were also analyzed with respect to management related vegetation plot data on species diversity, as well as on indicator values for nutrient supply and management tolerance. The stepwise clustering gave three use intensity classes for each dominant management practice of grazing (pasture) and mowing (meadow). Use intensity was higher for meadows than pastures with a distinct intensity gradient for each grassland practice. The GM classes reproduced established spectral and topographical patterns of grassland use intensity, indicated by increased standard deviations (SD) of spectral time series profiles (e.g. mean SD of 0.048 for pastures and 0.054 for meadows) and lower slopes (e.g. mean slopes of 10° for pastures and 7° for meadows). The averaged spatial conformity of the GM classes with a cantonal land use map was 82% for meadows and 97% for pastures. The GM classes spatially matched with land use patterns of three subregions, e.g. with an areal proportion of 73% pasture classes for a subregion dominated by grazing. Moreover, the GM classes reproduced established vegetation patterns of grassland use intensity along the GM intensity gradient, showing a mean decrease in species richness (33%), as well as a mean increase in indicator values for nutrient supply (5%), grazing tolerance (4%), and mowing tolerance (6%).



## Keywords

Grassland management, Spatial monitoring, Spectral time series, Management practice, Mowing, Grazing, Use intensity

## Introduction

More than a third of the agricultural area in Europe is covered by grasslands, widely being subject to land use intensification (O`Mara, 2012; Smit et al., 2008). Grassland agroecosystems, as grazed pastures or mown meadows, provide most of the fodder required by ruminant livestock to meet the growing demand for animal-derived food (Allen et al., 2011; Oenema et al., 2014; Orr et al., 2016). Grasslands also provide key ecosystem services related to soil quality, hydrological balance, and climate change (Askari and Holden, 2014; Isselstein and Kayser, 2014; Soussana and Lemaire, 2014). Moreover, temperate grasslands in Europe are rich in biodiversity, which is recognized as the foundation for ecosystem functioning (Cardinale et al., 2012; Kleijn et al., 2011; Wilson et al., 2012). For example, ~18% of the endemic vascular plant species in Europe are bound to grassland habitats (Habel et al., 2013).

Grassland management (GM) largely determines the capacity of the agroecosystem to provide ecosystem services (Power, 2010; Rodríguez et al., 2006). GM mainly involves practices such as grazing and mowing in various combinations and use intensities (Giménez et al., 2017; Peeters et al., 2014). The term 'use intensity' often refers to the frequency of management practices and alongside, to the amounts of mineral or organic fertilizer inputs, which determines biomass productivity (Hudewenz et al., 2012; Rose and Leuschner, 2012). Use intensity depends on the environmental setting related to terrain conditions, growth period, water supply, soil nutrient levels, and plant species compositions (Kizeková et al., 2018; Mottet et al., 2006; Tasser and Tappeiner, 2002), and often adversely affects ecosystem services (Foley et al., 2011; Porqueddu et al. 2016; Soussana and Lemaire, 2014). Rose et al. (2012) found that higher mowing intensity decreased groundwater recharge. Zhou et al. (2017) demonstrated that higher grazing intensity significantly increased the loss of carbon and nitrogen in the soil. Moreover, higher use intensity is associated with biodiversity declines, often accompanied by landscape homogenization (Allan et al., 2014; Foley et al., 2011; Gossner et al., 2016).

The large extents of agricultural grasslands, combined with the adverse effects of land use intensification on ecosystem services and biodiversity, outline the need for sustainable GM systems (Huyghe et al., 2014; Simons and Weisser, 2017; Tälle et al., 2016). This is in line with European agricultural policies, such as the Common Agricultural Policy of the European Union (CAP) and the Swiss Agricultural Policy (SAP; EC, 2018; FOAG, 2018). A key measure of these policies comprises payments linked to management practices and use intensities that promote ecosystem services and the conservation of biodiversity (Henle et al., 2008; Kleijn and Sutherland, 2003). However, ecological benefits in terms of reduced greenhouse gas emissions and biodiversity conservation seem to be limited in Switzerland (Kleijn et al.,

2006; Knop et al., 2006; Leifeld and Fuhrer, 2005). Moreover, it has recently been claimed that CAP measures, intended to preserve ecosystem services and halt biodiversity declines in the European Union, are inefficient (Kleijn et al., 2011; Pe'er et al., 2014; Pywell et al., 2012).

Current efforts to establish sustainable grassland management systems include the development of monitoring tools to derive detailed spatial and temporal information on management practices and use intensities (Nagendra et al., 2013; Zaks and Kucharik, 2011). Spectral time series from space-borne imagery allow progressing phenological stages to be detected, which in turn allows vegetation trends and GM to be mapped (Ali et al., 2016; Kennedy et al., 2014; Svoray et al., 2013). However, spatial and temporal resolutions of spectral imagery are often inadequate to capture small-structured landscapes with diverse GM (Giménez et al., 2017; Li et al., 2017; Zhu et al. 2012). Moreover, the spatio-temporal availability of spectral data is often constrained by sparse data records and high monetary expenses (Asam et al., 2015; Kolecka et al., 2018; Sakowska et al., 2016).

With this study we contribute to the establishment of a sustainable GM system for Switzerland at a national scale. The objective is the development of a monitoring tool for annually mapping the management practice (mowing vs. grazing) and use intensity of small-structured grassland systems. Exemplarily, the tool is applied and evaluated for the year 2015.

The methodological framework is based on a spatial estimation of the management frequency and biomass productivity using pixelwise time series from multi-temporal satellite imagery. Both variables are used for a stepwise classification and clustering approach to derive GM classes, which describe the use intensity for areas dominated by grazing and mowing, respectively. The GM classes are evaluated by a statistical comparison with established spectral and topographical patterns of grassland use intensity. The GM classes are also evaluated by analyzing their spatial conformity with regional land use data, available for the entire canton of Berne and for three subregions with characteristic GM. In addition, the GM classes were analyzed with respect to management related vegetation plot data on species diversity, as well as bioindicators for nutrient supply and management tolerance.

## Materials and Methods

### Study area

Switzerland covers an area of 41,000 km$^2$. The altitude varies between 196 m and 4,634 m (a.s.l.). Total annual precipitation is ~500-2000 mm and the monthly mean temperature ranges from 1℃ in January to 17℃ in July (MeteoSwiss, 2018a, b). The Swiss agroecosystem covers 10,000 km$^2$ and is small-structured with a mean farm size of about 20 ha. Approximately 70% of the agricultural land is managed grassland, of which 12% is temporary grassland in rotation with other crops that is typically managed intensively. The majority of the

managed grasslands are permanent pastures and meadows (SFSO, 2018; Stumpf et al., 2018). Most intensive grassland farming in Switzerland occurs in a band of moderate climate and suitable topography, running from southwest to the northeast of the northern half of the country (Leifeld et al., 2005; Price et al., 2015). With increasing elevations towards the alpine foothills in the south and the Jura uplands in the north, extensive grassland farming becomes dominant (SFSO, 2018). The study area demarcates the agricultural grassland area < 1,400 m (a.s.l.) in 2015, which covered ~5,400 km² (Figure 1; Stumpf et al., 2018).

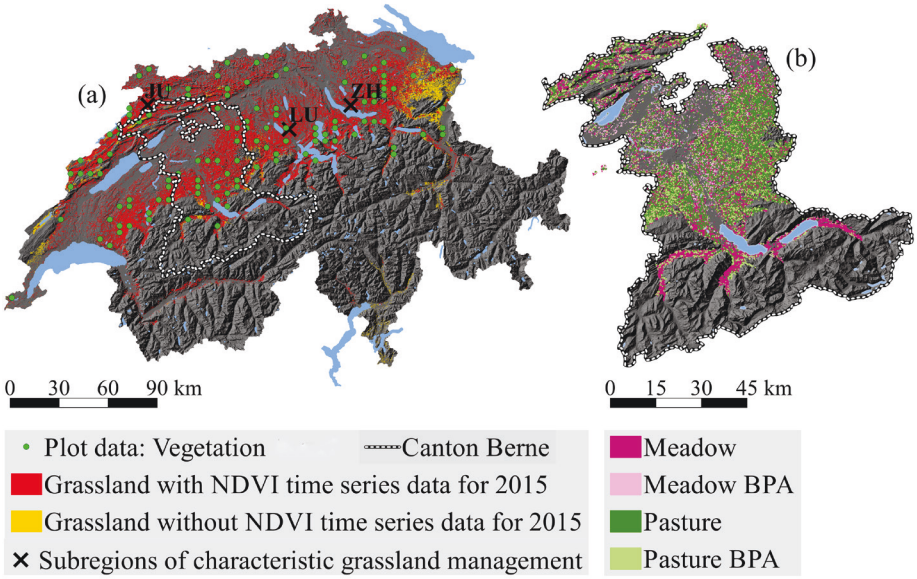


**Fig. 1.** Territory of Switzerland (a) with managed grasslands < 1,400 m a.s.l. (red + yellow) and the territory of the canton of Berne (b) with official data on grassland use ('NDVI': Normalized Differenced Vegetation Index; 'BPA': Biodiversity Promotion Area, see Section 2.2.3).

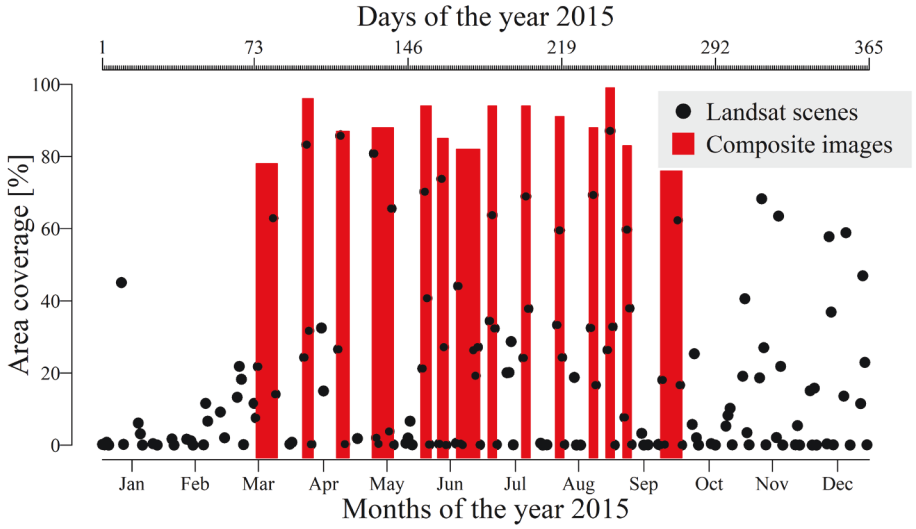


**Fig. 2.** Available Landsat scenes (sensors: ETM+ and OLI) for the Swiss grassland area in 2015 (black dots), aggregated to a time series of composite images (red bars), and the areal coverage of the composite images relative to the total grassland area.

## Appendix

### Data base

#### *Spectral time series data*

Spectral time series for the study area were acquired from all available scenes of the optical satellite sensors Landsat ETM+ and Landsat OLI in 2015.

Both sensors record a blue, a green, a red (RED), a near-infrared band (NIR), as well as two shortwave-infrared bands at a spatial resolution of 30 m x 30 m. Each scene was corrected to surface reflectance according to the Landsat Ecosystem Disturbance Adaptive Processing System (Schmidt et al., 2013). All pixels covered by clouds, cloud shadows, water, and snow were masked according to the CFmask algorithm (Zhu et al., 2015). Corrupted pixels, defined by at least one 'NoData' score in the spectral bands, were also removed (Stumpf et al., 2018). The Normalized Difference Vegetation Index (NDVI; Eq. 1; Rouse et al., 1974) was calculated for each scene, serving as a proxy for plant biomass (Liu et al., 2017; Schweiger et al., 2015; Weber et al., 2018).

$$NDVI = \frac{NIR - Red}{NIR + Red} \qquad \text{(Eq. 1)}$$

The NDVI data were aggregated to a stack of median composite images, which establish pixelwise and spatially harmonized time series. The time series aggregation was optimized to cover as much of the growth period and study area as possible, and to provide an accurate temporal match between the composite periods and vegetation status. Accordingly, the most adequate time series stack was selected from a pool of iteratively aggregated time series stacks following the optimization criteria i) temporal extent and resolution, ii) balanced time step intervals and composite periods, and iii) spatial coverage. Moreover, time series with more than three missing values or with at least two consecutive missing values were excluded from the analysis. Missing values for the remaining time series were imputed using cubic spline interpolation (Wolberg, 1999). The final NDVI time series ran from 16 March to 24 September of 2015 in 14 time steps, with composite periods between 3 to 9 days, time step intervals between 5 and 13 days, and a grassland area coverage of 89% (Figures 1 and 2). The scenes were selected and processed using the computing platform Google Earth Engine API (Gorelick et al., 2017). Data aggregation for the final NDVI time series and imputation of missing values was processed using the R-packages "zoo" and "raster" (Hijmans, 2017; Zeileis et al., 2018).

### *Terrain data*

We acquired a digital elevation model (Elev) and derived the slope (Slope) with a spatial resolution of 25 m x 25 m. The elevation model is based on interpolating digitized terrain elements and contains mean errors of 1.5-3 m (SwissTopo, 2018). The slope was calculated according to the algorithm by Horn (1981), which applies 3 x 3 pixel window and is particularly suitable for rough surfaces. The terrain grids were resampled based on bilinear interpolation to a raster cell size of 30 m x 30 m to fit the spatial resolution and extent of the spectral time series data. Terrain data were processed using the R-packages "raster" and "base" (Hijmans et al., 2016; R Core Team, 2016; Table 1).

**Table 1**. Summary statistics for the terrain variables elevation and slope (Min: Minimum; Q1: 25%-quartile; Q3: 75%-quartile; Max: Maximum; SD: Standard deviation).

| | Min | Q1 | Median | Mean | Q3 | Max | SD |
|---|---|---|---|---|---|---|---|
| Elevation (Elev) [m] | 388 | 556 | 693 | 712 | 840 | 1400 | 191 |
| Slope[°] | 0 | 4.1 | 8.2 | 9.6 | 13.6 | 33.5 | 6.9 |

### *Cantonal grassland use data*

We obtained georeferenced grassland use data, which determine the payment of agricultural subsidies for the canton of Berne for 2015 (Figure 1; LANAT, 2018). The data include two classes of conventional grassland use ('Meadow' and 'Pasture'; see Figure 1) and two classes of biodiversity promotion areas ('Meadow BPA' and 'Pasture BPA'; see Figure 1). These land use classes are defined in terms of management practices related to grazing and mowing, as well as related to the application of farmyard manure and plant protection agents. The Meadow class represents land subject to at least one mowing event without further restrictions. The Meadow BPA class contains land subject to at least one mowing event after 15 June, 1 July, or 15 July, depending on elevation and associated growth period. While grazing is permitted after 1 September, the application of farmyard manure and plant protection agents is prohibited. The Pasture class is confined to grazing as dominant management practice without further restrictions. The Pasture BPA class represents land dominated by grazing, while the application of manure and plant protection agents is prohibited. All land use units < 1 ha were excluded for the analysis to ensure a match with the spatial resolution of the spectral and terrain data (Table 2; AGRIDEA, 2018, SFA, 2018).

## Appendix

**Table 2.** Summary area statistics for the grassland use units of the canton of Berne according to grassland management classes in 2015 (Min: Minimum; Q1: 25%-quartile; Q3: 75%-quartile; Max: Maximum; SD: Standard deviation; BPA: Biodiversity promotion area).

| | Min | Q1 | Median | Mean | Q3 | Max | SD | Total |
|---|---|---|---|---|---|---|---|---|
| Meadow [ha] | 1 | 1.3 | 1.8 | 2.3 | 2.7 | 33.1 | 1.7 | 40,702 |
| Meadow BPA [ha] | 1 | 1.1 | 1.4 | 1.6 | 1.8 | 21.9 | 1 | 3,933 |
| Pasture [ha] | 1 | 1.3 | 1.7 | 2.4 | 2.6 | 40.5 | 2.1 | 20,659 |
| Pasture BPA [ha] | 1 | 1.3 | 1.7 | 2.4 | 2.6 | 22.5 | 1.9 | 5,723 |

### *Subregional grassland use data*

We also compiled semi-quantitative management information for three contrasting grassland-dominated subregions in the cantons of Lucerne, Zurich, and Jura based on the Swiss National Farm Census and other published data. Each subregion was randomly selected within the cantons of interest, covering a total area of 2.3 km x 2.3 km of which at least 50% are managed grasslands. The Swiss National Farm Census is the main instrument for administering direct payment schemes from the Swiss Federal Agency for Agriculture and includes annual data on livestock units (LSU) at farm level (AGIS, 2015; FOAG, 2018; Gärtner et al., 2013). In the Swiss context, LSU indicates use intensity of animal farming systems, since it is related to management practices such as grazing and mowing frequencies and organic fertilizer inputs (FOAG, 2018; Giménez et al., 2017). The livestock density index, defined as the ratio of livestock units to grassland area, was calculated using the livestock data for the farms in each subregion and served as indicator for use intensity. The mean mowing frequency and dominant management practice for each subregion were determined from published data. The grassland of the subregion in Lucerne (LU) is predominantly managed by intensive mowing and high application rates of farmyard manure (Liebisch, 2011). Grassland management of the subregion in Zurich (ZU) is characterized by mowing and grazing in varying intensities, and by different manure application rates (Giménez et al., 2017). Grasslands in the subregion of Jura (JU) are predominantly managed by grazing in low to moderate intensities (Table 3; Figure 1; Chételat et al., 2013; Masé, 2005).

**Table 3.** Grassland management characterization of subregions in the cantons Lucerne (LU), Zurich (ZU), and Jura (JU) in terms of the dominant practice, mowing frequency in 2014, and the livestock density index in 2015 (LSU: livestock unit).

| | Dominant practice [-] | Mowing frequency [-] | Livestock density index [LSU*ha$^{-1}$] |
|---|---|---|---|
| LU | Mowing | > 5 | 1.87 |
| ZU | Mowing/Grazing | 2 - 5 | 1.36 |
| JU | Grazing | - | 0.94 |

### *Vegetation data*

We acquired vegetation plot data on species occurrence from the Swiss Biodiversity Monitoring Network (BDM), which spans a 6 km x 4 km sampling grid across Switzerland (BDM, 2014, 2017). Data were obtained at circular plots of 10 $m^2$ around the grid intersections (±5 cm) between 20 July and 20 September in 2012 to 2015. We identified a total of 120 sample plots that intersected spatially with the study area (Figure 1).

The vegetation data is based on recording each identifiable vascular plant species within the sample plot. The data set includes species richness (SR) and the Landolt N indicator, which is similar to Ellenberg's nutrient indicator but adapted for the Swiss biogeography (Ellenberg, 1974; Landolt, 2010). Landolt N values are an ordinally scaled quantification of the species` nutrient preferences, ranging from 1 to 5 with a higher value indicating an increased nutrient supply (Diekmann, 2003; Landolt, 2010; Pauler et al., 2019). The plotwise unweighted mean of the Landolt values (NLandolt) indicates the site specific nutrient availability or soil fertility (Bartelheimer and Poschlod, 2016; Duprè et al., 2009; Klaus et al., 2012). Based on the species record, we also calculated the unweighted mean indicator values for grassland utilization according to Briemle et al. (2002), indicating tolerance for mowing (MTBrie.) and grazing (GTBrie.). Values of MTBrie. and GTBrie. range between 1 and 9, corresponding to the range from least to most tolerant for mowing and grazing, respectively (Briemle et al., 2002; Moog et al., 2002).

**Table 4**. Summary statistics for the vegetation data (SR, NLandolt, MTBrie., GTBrie.: see Section 2.2.5; Min: Minimum; Q1: 25%-quartile; Q3: 75%-quartile; Max: Maximum; SD: Standard deviation).

| | Min | Q1 | Median | Mean | Q3 | Max | SD |
|---|---|---|---|---|---|---|---|
| SR [-] | 12 | 21 | 26 | 26 | 30 | 44 | 7 |
| $N_{Landolt}$ | 3.1 | 3.5 | 3.7 | 3.6 | 3.8 | 4 | 0.2 |
| $GT_{Brie.}$ | 4.5 | 5.3 | 5.7 | 5.7 | 6.1 | 6.8 | 0.5 |
| $MT_{Brie.}$ | 5.6 | 6.8 | 7.1 | 7.0 | 7.3 | 7.7 | 0.4 |

## Spatial monitoring of grassland management

### *Mapping management practice and use intensity*

Grassland use intensity was spatially quantified from the pixelwise NDVI time series (see Section 2.2.1), described by the management frequency (Mfreq) and biomass productivity (BP). NDVI changes of a specific time series were used as local proxy for fluctuations in biomass (Flynn et al., 2008; Gao et al., 2013; Jiang et al., 2014). A loss of biomass greater than a threshold value q was deemed to be land use induced, and therefore identified as a management event (Kennedy et al., 2010; Sulla-Menashe, 2014). Based on the probability

density function of all NDVI changes across the time series, q was specified for the probability p = 0.01 using the quantile-function of the R-package "stats" (Hyndman and Fan, 1996; R Core Team, 2016). Subsequently, Mfreq was defined as the count of management events for each time series, while BP was defined as the accumulated absolute NDVI changes in the management events (Eq. 2 and 3).

$$M_{freq} = \sum_{i=1}^{K} [(NDVI_{i+1} - NDVI_i) > q] \quad \text{(Eq. 2)}$$

$$BP = \sum_{i=1}^{K} |(NDVI_{i+1} - NDVI_i)[(NDVI_{i+1} - NDVI_i) > q]| \quad \text{(Eq. 3)}$$

where K is the time series (i = 1,2,…,13), NDVI is the NDVI value (Eq. 1), and q is the threshold to ensure human interference for a biomass loss (q = 0.02). The variables Mfreq and BP were calculated using the R-package "base" and "stats" (R Core Team, 2016).

We mapped GM classes as a combination of dominant management practice (grazing vs. mowing) and use intensity, applying a stepwise classification and clustering approach based on Mfreq and BP. First, the study area was sub-divided into three initial grassland classes using the categorization of use intensity for Swiss grassland, Mfreq <= 2, 3 <= Mfreq <= 4, and Mfreq >= 5 (PRIF, 2017). Second, unsupervised Kmeans clustering for each initial grassland class was performed using BP to identify areas dominated by grazing and mowing. Grazing areas are characterized by low BP values and gradually changing NDVI profiles, while mowing areas show higher BP values and abrupt changes in the NDVI profile (Dusseux et al., 2014a, b; Fang et al., 2015; Taugourdeau et al., 2013; Weber et al., 2018). Therefore, the six GM classes were intended to describe gradual changes in use intensity for areas dominated by grazing ("Pasturelow", "Pasturemoderate", "Pasturehigh"), and for areas dominated by mowing ("Meadowlow", "Meadowmoderate", "Meadowhigh"). The Kmeans clustering was performed using the Hartigan-Wong implementation of the R-package "stats" (Hartigan and Wong, 1979; R Core Team, 2016).

### *Evaluation of grassland management classes*

The adequacy of the estimated GM classes was assessed using different approaches and independent datasets.

First, we evaluated the GM classes by considering the statistical conformity with established change patterns of NDVI profiles for grassland. Meadows are predominantly characterized by bimodal NDVI profiles and increased BP values compared to pastures with mainly unimodal profiles. High intensity grasslands have bimodal NDVI profiles with increased variability, Mfreq, and BP compared to lower intensity grasslands with flattened or unimodal NDVI profiles (Dusseux et al., 2014b; Estel et al., 2015; Taugourdeau et al., 2013; Weber et al., 2018).

We compared the GM classes based on the mean NDVI profile changes for each class, and based on the absolute NDVI profile changes from a randomly selected location within each class. Descriptive statistics, Mfreq, and BP were used as comparative measures.

Second, we evaluated the GM classes in relation to topographical conditions, which largely determine use intensity of grasslands. Increasing elevation is related to decreasing mean temperature, and therefore to a shorter growth period, which involves a decrease in use intensity (Lamarque et al., 2011; Weber et al., 2018; Zeeman et al., 2010). Similarly, increasing slope is related to decreasing use intensity, because of constraints on access for mowing machinery and livestock to a lesser extent (Peter et al., 2008; Tasser and Tappeiner, 2002; Weber et al., 2018). We analyzed these relationships between GM and topography by comparing the GM classes based on the classwise distributions of elevation and slope (see Section 2.2.2; Table 1).

Third, we evaluated the GM classes using the georeferenced grassland use data for the canton of Berne in 2015 (see Section 2.2.3; Table 2). The conventional cantonal grassland use classes ("Meadow", "Pasture") are characterized by higher variability and use intensity in terms of Mfreq and fertilizer inputs compared to the biodiversity promotion classes ("Meadow BPA", "Pasture BPA"; AGRIDEA, 2018; SFA, 2018). We assessed the representativeness of the GM classes according to the cantonal grassland classes by identifying the areal proportions of the GM classes within each cantonal grassland class.

Fourth, we evaluated the GM classes based on semi-quantitative GM information from the three subregions LU, ZH, and JU (see Section 2.2.4; Table 3). Use intensity decreased and the dominance of grazing increased in the order LU, ZH, JU. We assessed how well the GM classes match GM in the three subregions. Thus, we determined the areal proportion of each GM class within each subregion.

Fifth, we evaluated the GM classes by a comparison to established management related change patterns of SR, NLandolt, MTBrie., and GTBrie. (see Section 2.2.5; Table 4). In this context, SR decreases with higher grassland use intensity (Allan et al., 2015; Gossner et al., 2016; Socher et al. 2013). Moreover, NLandolt, MTBrie., and GTBrie. increases with higher use intensity (Blüthgen et al., 2012; Peter et al., 2008; Weber et al., 2018). We investigated the relationships between GM and grassland vegetation by comparing the GM classes based on the classwise distributions of SR, NLandolt, MTBrie, GTBrie.

Classwise outliers of the respective evaluation data, defined by a distance of two standard deviations from the mean, were removed for the analyses. Tukey`s HSD tests were used to analyze the difference of classwise distributions with respect to BP and topography. Statistical

calculations were performed using the R-package “base” and “stats”, while the R-packages “raster” and “rgdal” were applied for geospatial processing (Hijmans et al., 2016; Bivand et al., 2016; R Core Team, 2016).

## Results

### *Grassland management classes and NDVI patterns*

The NDVI profiles of the GM classes show statistical conformity with established change patterns for grassland (Figure 3a, d; see Section 2.2.1 and 2.3.2). The mean NDVI profile changes for all GM classes show bimodal temporal distributions, while use intensities are higher for meadows than for pastures with an increase across the classwise intensity graduation (“low”, “moderate”, “high”). The standard deviations (SDs) of the pasture classes increase in the order “Pasturelow” (SD = 0.046), Pasturemoderate (SD = 0.049), Pasturehigh (SD = 0.50). The profiles of the meadow classes are consistently more variable, but SDs increase in the same order, Meadowlow (SD = 0.056), Meadowmoderate (SD = 0.062), “Meadowhigh” (SD = 0.065). The median NDVI of the classwise profiles (Mmed) varies in the same way as the variability increases, with Mmed being 0.04 for Pasturelow, 0.056 for Pasturemoderate, 0.067 for Pasturehigh, 0.078 for Meadowlow, 0.103 for Meadowmoderate, and 0.113 for Meadowhigh. For all the GM classes, management was more frequent in May, July, and less pronounced in September (Figure 3a; Appendix A-B). The classwise profiles of the absolute changes in NDVI for the randomly selected sites show similar but less distinct patterns compared with the mean NDVI profile changes (Figure 3b-c; Appendix B-C). The median BP of the GM classes indicates higher use intensity for meadows than for pastures and an increase from low to high intensity grades. The median BP values are 0.13 to 0.253 and 0.366 for Pasturelow, Pasturemoderate, and Pasturehigh, respectively, and 0.34, 0.501, and 0.619 for Meadowlow, Meadowmoderate, and Meadowhigh (Figure 3d; Appendix C).

### Grassland management classes and topography

The topography of the GM classes corresponds to established patterns of grassland use intensity (Figure 4a; Table 1; see Section 2.2.2 and 2.3.2). Slopes are lower for meadows than pastures, and decrease as use intensity increases. The median slopes are 12° for Pasturelow, 10° for Pasturemoderate, 9° for Pasturehigh, 8° for Meadowlow, 7° for Meadowmoderate, and 6° for Meadowhigh. Elevations decrease as the use intensity increases and are slightly lower for meadows than pastures (Figure 4b). The median elevations are 748 m for Pasturelow, 720 m for Pasturemoderate, 667 m for Pasturehigh, 712 m for Meadowlow, 687 m for Meadowmoderate, and 649 m for Meadowhigh (Appendix D).

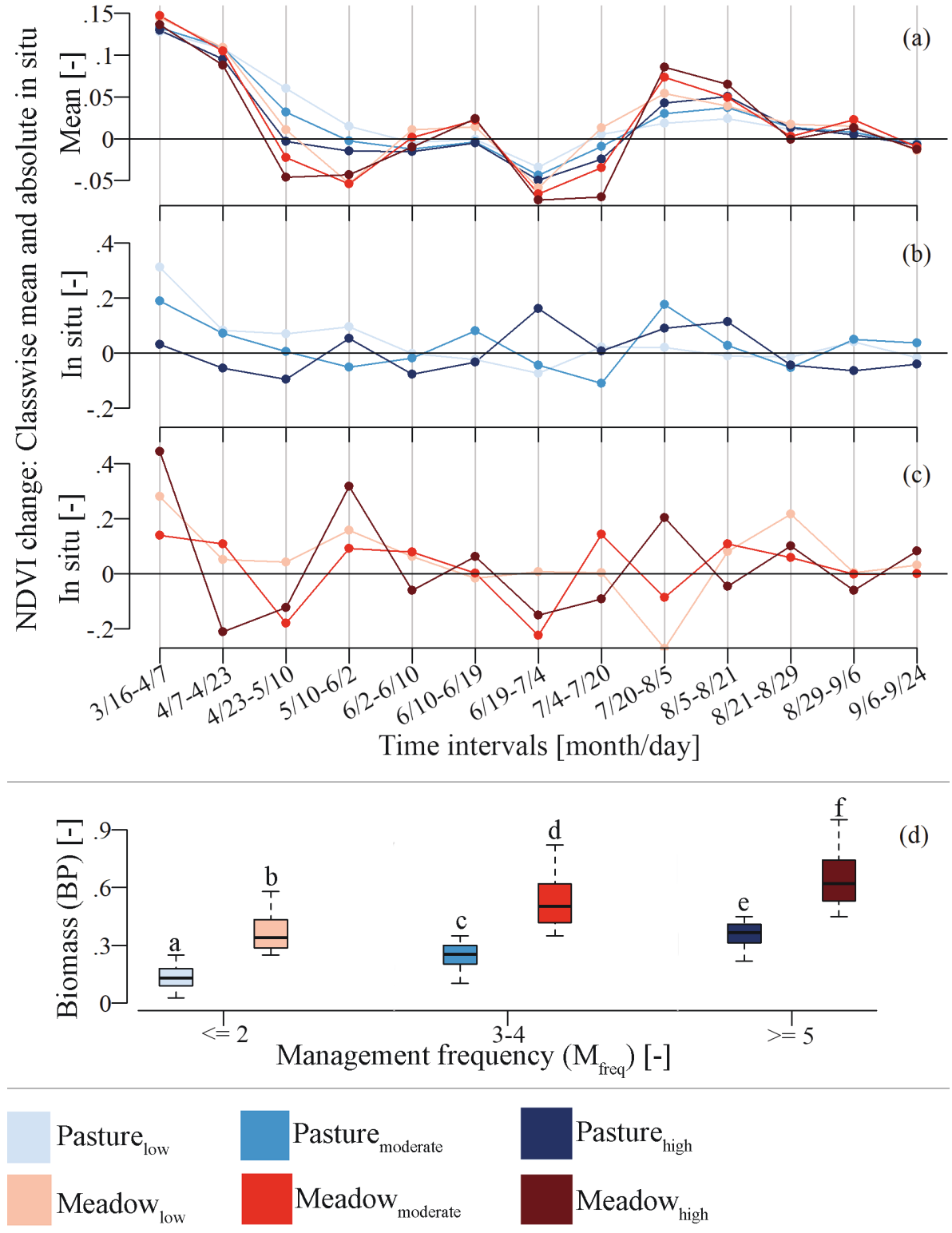


**Fig. 3.** The grassland management classes described by temporal profiles of mean NDVI changes (a), by temporal profiles of NDVI changes from randomly selected locations (in situ; b-c), and related to the underlying biomass productivity (BP) and management frequency (Mfreq, d). Significant differences between the classwise distributions (Tukey HSD, $P < 0.05$) are indicated by different letters above the boxplots.

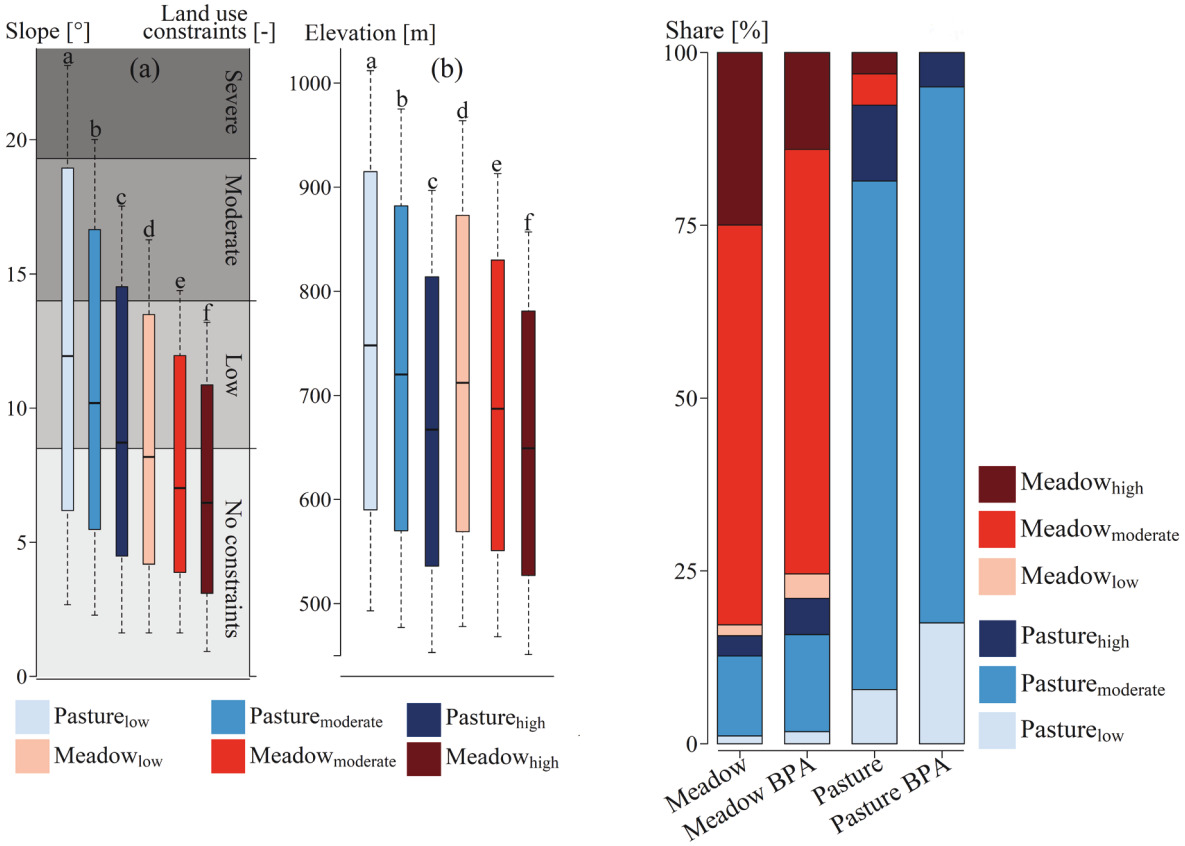


**Fig. 4.** (Left panel) Grassland management classes compared to slope as proxy for land use constraints (a), and to elevation as proxy for the length of the growth period (b). Significant differences between the classwise distributions (Tukey HSD, $P < 0.05$) are indicated by different letters above the boxplots.

**Fig. 5.** (Right panel) The areal proportions of the grassland management classes within the cantonal land use classes of Berne (Meadow, Meadow BPA, Pasture, Pasture BPA; BPA: Biodiversity promotion area).

### Grassland management classes and cantonal grassland use

The GM classes spatially match the main patterns of the grassland use data for the canton of Berne (Figure 5; Table 2; see Section 2.2.3 and 2.3.2). For example, 85% and 79% of the cantonal meadow classes Meadow and Meadow BPA, respectively, are covered by the GM meadow classes (Meadowlow, Meadowmoderate, Meadowhigh). The high intensity GM meadow class (Meadowhigh) covers 11% less of the biodiversity promoting meadow class (Meadow BPA) than the conventional meadow class (Meadow). The cantonal pasture classes Pasture and Pasture BPA are covered by 93% and 100%, respectively, by the GM pasture classes (Pasturelow, Pasturemoderate, Pasturehigh). The GM class Pasturelow covers 10% more of the biodiversity promoting pasture class (Pasture BPA) than the conventional pasture class (Pasture; Appendix E).

### Grassland management classes and subregional grassland use

The GM classes capture the main GM patterns of the subregions LU, ZH, and JU (Figure 6; Table 3; see Section 2.2.4 and 2.3.2). With an increase of mowing frequency and livestock density in the order JU, ZH, LU, the areal proportions of GM meadow classes increase, while the areal proportions of GM pasture classes show a contrary trend. For example, JU is dominated by low intensity grazing (livestock density index: 0.94 LSU*ha$^{-1}$), which is reproduced by the GM pasture classes with an areal proportion of 73% and prevailing low to moderate intensities (Pasturelow: 13%, Pasturemoderate: 49%, Pasturehigh: 11%). The mixture of mowing and grazing at different intensities in ZH (mowing frequency: 2-5; livestock density index: 1.36 LSU*ha$^{-1}$) is reproduced by the GM classes with high areal proportions of moderate to high intensity GM pasture and meadow classes (Pasturemoderate: 31%, Pasturehigh: 13%, Meadowmoderate: 29%, Meadowhigh: 15%). The LU subregion is dominated by high intensity mowing (mowing frequency: >5; livestock density index: 1.87 LSU*ha$^{-1}$), which is reproduced by GM meadow classes of moderate to high use intensity with an areal proportion of 74% (Meadowmoderate: 37%, Meadowhigh: 37%; Appendix A).

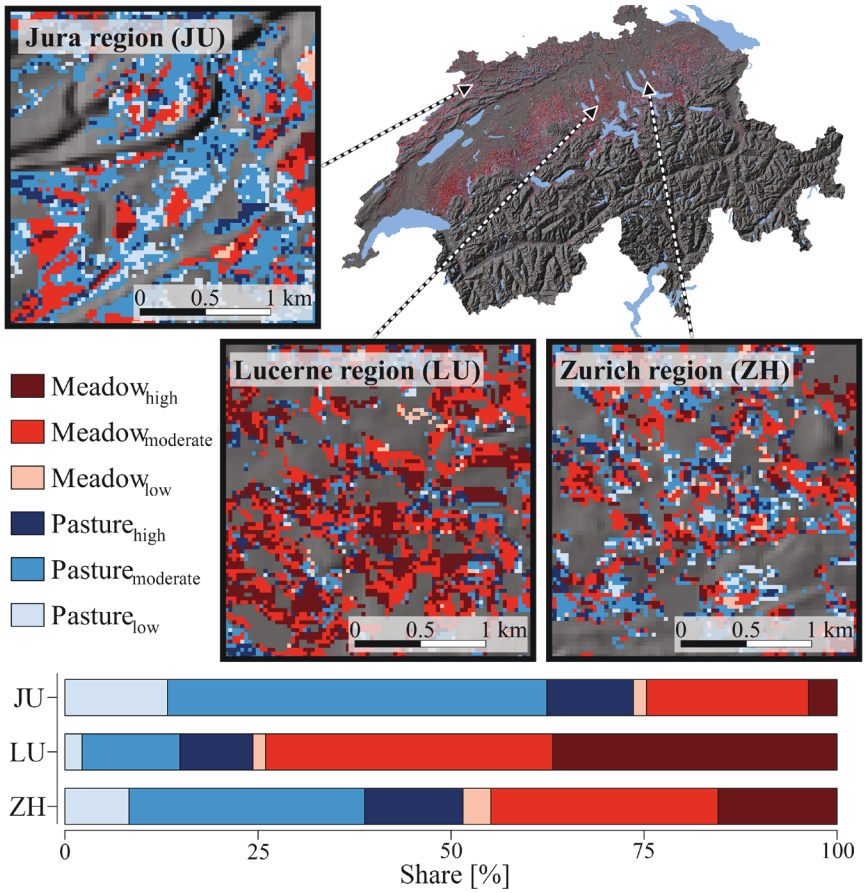


**Fig. 6.** Mapped grassland management classes and their areal proportions within three subregions of characteristic grassland management in Switzerland (JU: Jura, LU: Lucerne, ZH: Zurich).

# Appendix

## Grassland management classes and vegetation patterns

Changes of plotdata-SR, -NLandolt, -GTBrie., and -MTBrie. across the GM classes correspond to established patterns related to grassland use intensity (Figure 7; Table 4; see Section 2.2.5 and 2.3.2). SR is higher for pastures than meadows, and decreases as use intensity increases. The classwise median SR values are 33 for Pasturelow, 28 for Pasturemoderate, 26 for Pasturehigh, 27 for Meadowlow, 23 for Meadowmoderate, and 21 for Meadowhigh. NLandolt is lower for pastures than meadows, and increases with higher use intensity, confirming the well-known negative correlation with SR (Appendix F, G). The classwise increase of NLandolt is more pronounced for meadows. The median NLandolt values are 3.53 for Pasturelow, 3.53 for Pasturemoderate, 3.56 for Pasturehigh, 3.61 for Meadowlow, 3.67 for Meadowmoderate, and 3.74 for Meadowhigh. Similar trends are observed for GTBrie. and MTBrie.. The classwise median GTBrie. values are 5.34 for Pasturelow, 5.7 for Pasturemoderate, and 5.86 for Pasturehigh, while the median MTBrie. values range from 6.79 for Meadowlow, 7.14 for Meadowmoderate to 7.15 for Meadowhigh (Appendix F, G).

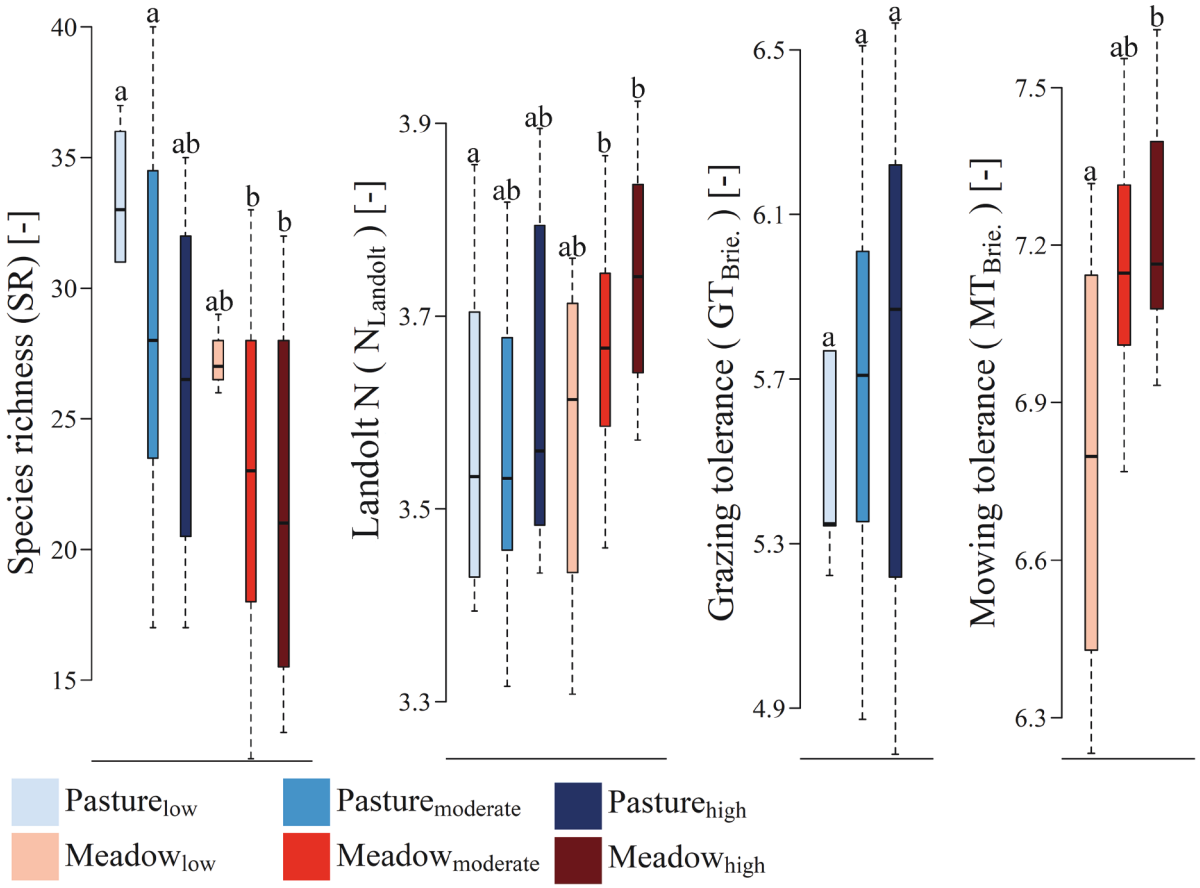


**Fig. 7.** Grassland management classes compared to species richness (SR), to Landolt N (NLandolt), to grazing tolerance (GTBrie.), and to mowing tolerance (MTBrie.). Significant differences between the classwise distributions (Tukey HSD, $P < 0.05$) are indicated by different letters above the boxplots.

## Discussion

### Mapping grassland management

Mapping GM from multi-temporal satellite imagery is often confined to a spatial representation of use intensity, disregarding the heterogeneity of GM practices (Green et al., 2016; Kolecka et al., 2018; Li et al., 2017). However, GM practices alternate spatially and often involve combinations of grazing and mowing at different intensities (Asam et al., 2015; Jeangros and Thomet, 2004). Integrating GM practices and use intensity is therefore highly relevant when monitoring spatial patterns of GM (Erb et al., 2013; Kuemmerle et al., 2013). In this study, we mapped the dominance of mowing and grazing, combined with associated use intensities exemplarily for Swiss grasslands in 2015. Pixelwise spectral time series were used to estimate Mfreq and BP in order to subsequently define GM according to six classes using a stepwise classification and clustering approach (Figure 3; see Section 2.3.1). A similar method was previously applied for the Swiss case by Giménez et al. (2017), who mapped mowing frequencies and grazing intensities for an area of 67 km² based on RapidEye data. However, their approach was limited to spectral time series of only five time steps and a fragmentary coverage of the growth period. The pixel resolution of 5 m x 5 m allowed to capture small structured land use patterns, but the commercial nature and a relatively low spatial coverage of RapidEye data qualifies particularly for mapping small areas. Other studies used MODIS imagery (Moderate Resolution Imaging Spectroradiometer), which are available at a daily basis and a spatial resolution of 250 m x 250 m since 2002, and therefore mainly qualify to map large areas with homogenous spatio-temporal landscape structures (Estel et al., 2018; Green et al., 2016; Li et al., 2017). Monitoring exercises targeting current and future GM benefit from the use of Sentinel imagery, which are characterized by a revisit of 3 to 10 days and a spatial resolution of 10 m x 10 m, available since 2017 (Claverie et al., 2018; Griffiths et al., 2019; Kolecka et al., 2018). We used spectral data from the Landsat archive, available since 1984 and allowing to annually cover large areas with dense time series for timely capturing land use activities at national scale and for most of the growth period. Our Landsat time series for 2015 covers the growth period in 14 time steps and the grassland area by 89% at a spatial resolution of 30 m x 30 m (Figure 1, 2).

GM classes were determined by identifying management related patterns in the temporal profiles of spectral time series (Ali et al., 2016). The profiles describe dynamics of biophysical vegetation properties based on NDVI as indicator for plant biomass (Dusseux et al., 2014b; Psomas et al., 2011). The strength of the NDVI-biomass relationship varies spatially, determined by local site characteristics with respect to plant species composition, vegetation growth stage, topography, soil exposure, and dense biomass (Garroutte et al., 2016; Metzger

et al., 2016; Porter et al., 2014). However, NDVI has been approved to be a stable and robust biomass proxy in various landscapes (Ali et al., 2016; Nestola et al., 2016; Zhang et al., 2003). In the present study, we used the NDVI because it is resistant to spectral noise, caused by topography and cloud shadows, particularly for rugged terrain (Huete et al., 2002). Moreover, remote sensing based NDVI profiles have been approved to be a reliable data source to describe grassland management in the Swiss context. Kolecka et al. (2018) successfully mapped mowing frequencies in the northern Swiss canton Aargau (1403 km$^2$) based on NDVI profiles from Sentinel imagery. Weber et al. (2018) showed the adequacy of Landsat NDVI profiles to discriminate grassland management practice and use intensity at 3000 sites distributed across Switzerland and covering ~240 km$^2$.

**Validity of the grassland management classes**

The GM classes were evaluated with respect to typical NDVI profiles for GM, topographical site conditions related to GM, cantonal grassland use data, and semi-quantitative grassland use data from three subregions. Moreover, the GM classes were evaluated using established management related change patterns of species richness (SR), and vegetation based indicator values for nutrient supply (NLandolt) and management tolerance (MTBrie., GTBrie.; see Section 2.3.2).

Temporal NDVI profiles of the proposed GM classes are statistically distinct, while showing higher variabilities, BPs, and Mfreq, as well as enhanced bimodal distributions for meadows than for pastures and from low to high use intensity. Thus, the GM classes match typical management related NDVI patterns for grassland systems (Figure 3; Appendix B-C; Dusseux et al., 2014a, b; Fang et al., 2015; Taugourdeau et al., 2013; Weber et al., 2018).

Elevations and slopes of the GM classes are lower for meadows than pastures and decrease as the use intensity increases. These patterns indicate the validity of the GM classes, because with increasing elevation BP is incrementally limited due to a lower mean temperature and shorter growth period (Figure 4b; Schermer et al., 2016; Weber et al., 2018; Zeeman et al., 2010). In addition, increasing slopes hamper the use of machinery for GM, which is associated with decreased use intensity (Figure 4a; Appendix D; Lamarque et al., 2011; Peter et al., 2008; Tasser and Tappeiner, 2002). The large overlapping ranges in the elevation and slopes of the GM classes for the entire Swiss grassland < 1,400 m (a.s.l.) could be attributed to typical spatial variability in BP, caused by numerous interrelated factors, such as topography, climate, soil quality, and species composition (Porqueddu et al., 2016).

The GM classes are consistent in terms of spatial and contextual detail with georeferenced data on grassland use for the canton of Berne. The conformity is very high for pastures, but lower for meadows, which could be explained by grazing events that are possible according to the class definition of the cantonal land use data (Figure 5; Appendix E; AGRIDEA, 2018,

SFA, 2018). This discrepancy in class definitions also explains the relatively small deviations between the GM classes and the cantonal data comparing conventional and BPA meadows and pastures, respectively.

Regional differences in GM of the subregions in the cantons Lucerne (LU), Zurich (ZH), and Jura (JU) are realistically reflected in the GM classes. For instance, LU with a long growth period, high livestock densities, and mowing frequencies is largely characterized by moderate to high intensity GM meadow classes. In contrast, ZH and JU with shorter growth periods, lower livestock densities, but predominantly managed by grazing are characterized by larger proportions of low to moderate GM pasture classes (Figure 6; Appendix A; Chételat et al., 2013; Giménez et al., 2017; Masé, 2005).

Our results reveal trends of decreasing SR, as well as increasing NLandolt, MTBrie., and GTBrie. along the increase of use intensity. These trends reproduce established change patterns of species diversity, nutrient availability, and management tolerance on agricultural grasslands (Blüthgen et al., 2012; Peter et al., 2008; Weber et al., 2018). Some of the difference between pasture and meadow GM classes may also stem from the geographical distribution of the two classes, e.g. more pastures at steeper slopes and higher elevations. The large overlapping ranges in species based variables (SR, NLandolt, MTBrie., GTBrie.) of the GM classes could be associated to relatively low SR ranges for pastures of the used data, as well as to the limited and imbalanced number observations n across the GM classes (Figure 7; Appendix F-G; SRPastures: 17-44; SRMeadows: 12-39; nPastures: 6-29; nMeadows: 4-38). Previous studies found SR ranges of 24-64 for pastures and 10-36 for meadows with different grazing and mowing frequencies (Kleijn and Müller-Schärer, 2006; Liebisch et al., 2013, Pauler et al., 2019). However, observed differences between meadows and pastures, and vegetation trends along the GM intensity gradient, approve the GM classification to be robust. The generally large variability of SR observed in the GM classes is likely an effect of additional factors such as nutrient management or regional differences in species abundance among others.

## Conclusion

The developed spatial monitoring tool for detecting grassland management systems discriminates mowing and grazing practices, as well as associated use intensities for Swiss grasslands. The tool is based on high resolution spectral time series derived from freely available satellite images with adequate spatial resolution. This data source allows small structured agroecosystems and frequent management activities to be captured at the national scale and on an annual basis for future and retrospective studies. We applied the grassland monitoring tool exemplarily for the growth period of the year 2015. The plausibility of the

tool was approved by an evaluation scheme that includes regional land use data, as well as established management related patterns of NDVI phenology, topography and vegetation. The land use information provided by the grassland monitoring tool may help to establish a balance between agricultural production and the maintenance of ecosystem functioning at the landscape scale.

## Acknowledgements

This study contributes to the development of a spatially distributed Land management Model for the Swiss Plateau that links agricultural management to soil services and policies (Project Number: 406840-161896). The project is funded by the Swiss National Science Foundation within the framework of the National Research Program "Sustainable Use of Soil as a Resource" (NRP 68; www.nrp68.ch). The contribution of Michael Schaepman is supported by the University of Zurich Research Priority Program on Global Change and Biodiversity (URPP GCB).

## References

AGIS, 2017. Agricultural Information System AGIS (Swiss Federal Office for Agriculture FOAG). https://www.blw.admin.ch/blw/de/home/politik/datenmanagement/agate/agis.html. (Accessed October 2018).

AGRIDEA, 2018. Centre for Agricultural Advisory and Environment. http://www.agridea.ch/en/publications/publication/environment-nature-and-landscape/legal-and-administrative-aspects/promotion-de-la-biodiversite-dans-lexploitation-agricole/.(Accessed July 2018).

Ali, I., Cawkwell, F., Dwyer, E., Barrett, B., Green, S., 2016. Satellite remote sensing of grasslands: from observation to management. Journal of Plant Ecology 9, 649-671.

Allan, E., Bossdorf, O., Dormann, C.F., Prati, D., Gossner, M.M., Tscharntke, T., Blüthgen, N., Bellach, M., Birkhofer, K., Boch, S., 2014. Interannual variation in land-use intensity enhances grassland multidiversity. Proceedings of the National Academy of Sciences 111, 308-313.

Allan, E., Manning, P., Alt, F., Binkenstein, J., Blaser, S., Blüthgen, N., Böhm, S., Grassein, F., Hölzel, N., Klaus, V.H., 2015. Land use intensification alters ecosystem multifunctionality via loss of biodiversity and changes to functional composition. Ecology Letters 18, 834-843.

Allen, V.G., Batello, C., Berretta, E., Hodgson, J., Kothmann, M., Li, X., McIvor, J., Milne, J., Morris, C., Peeters, A., 2011. An international terminology for grazing lands and grazing animals. Grass and Forage Science 66, 2-28.

Asam, S., Klein, D., Dech, S., 2015. Estimation of grassland use intensities based on high spatial resolution LAI time series. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 40, 285-291.

Askari, M.S., Holden, N.M., 2014. Indices for quantitative evaluation of soil quality under grassland management. Geoderma 230, 131-142.

Bartelheimer, M., Poschlod, P., 2016. Functional characterizations of Ellenberg indicator values–a review on ecophysiological determinants. Functional Ecology 30, 506-516.

BDM, 2014: Swiss Biodiversity Monitoring – Methods and Indicators. https://www.bafu.admin.ch/bafu/en/home/topics/biodiversity/publications-studies/publications/biodiversity-monitoring.html (Accessed October 2019).

BDM, 2017. Biodiversity Monitoring Switzerland. http://www.biodiversitymonitoring.ch/fileadmin/user_upload/documents/daten/anleitungen/1440_Anleitung_Z9-Pflanzen_v13.pdf. (Accessed July 2018).

Bivand, R., Keitt, T., Rowlingson, B., 2016. rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 1.1-9, 'rgdal' package.

Blüthgen, N., Dormann, C.F., Prati, D., Klaus, V.H., Kleinebecker, T., Hölzel, N., Alt, F., Boch, S., Gockel, S., Hemp, A., 2012. A quantitative index of land-use intensity in grasslands: integrating mowing, grazing and fertilization. Basic and Applied Ecology 13, 207-220.

Briemle, G., Nitsche, S., Nitsche, L., 2002. Nutzungswertzahlen für Gefäßfplanzen des Grünlandes. Bundesamt für Naturschutz, Schriftenreihe für Vegetationskunde 38, 203-225.

Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., 2012. Biodiversity loss and its impact on humanity. Nature 486, 59-67.

Chételat, J., Kalbermatten, M., Lannas, K.S., Spiegelberger, T., Wettstein, J.-B., Gillet, F., Peringer, A., Buttler, A., 2013a. A contextual analysis of land-use and vegetation changes in two wooded pastures in the Swiss Jura Mountains. Ecology and Society 18, 39.

Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.-C., Skakun, S.V., Justice, C., 2018. The Harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sensing of Environment 219, 145-161.

Diekmann, M., 2003. Species indicator values as an important tool in applied plant ecology–a review. Basic and Applied Ecology 4, 493-506.

Duprè, C., Stevens, C.J., Ranke, T., Bleeker, A., Peppler-Lisbach, C., Gowing, D.J., Dise, N.B., Dorland, E., Bobbink, R., Diekmann, M., 2010. Changes in species richness and composition in European acidic grasslands over the past 70 years: the contribution of cumulative atmospheric nitrogen deposition. Global Change Biology 16, 344-357.

Dusseux, P., Corpetti, T., Hubert-Moy, L., Corgne, S.J.R.S., 2014a. Combined use of multi-temporal optical and radar satellite images for grassland monitoring. Remote Sensing 6, 6163-6182.

Dusseux, P., Vertès, F., Corpetti, T., Corgne, S., Hubert-Moy, L.J.E.m.a.a., 2014b. Agricultural practices in grasslands detected by spatial remote sensing. Environmental Monitoring and Assessment 186, 8249-8265.

EC, 2018. The common agricultural policy. https://ec.europa.eu/info/food-farming-fisheries/key-policies/common-agricultural-policy_en. (Accessed October 2018).

Ellenberg, H., 1974. Indicator values of vascular plants in central Europe, Scripta Geobotanica, Vol. 9, Goltze Publishing, Göttingen, Germany.

Erb, K.-H., Haberl, H., Jepsen, M.R., Kuemmerle, T., Lindner, M., Müller, D., Verburg, P.H., Reenberg, A., 2013. A conceptual framework for analysing and measuring land-use intensity. Current Opinion in Environmental Sustainability 5, 464-470.

Estel, S., Kuemmerle, T., Alcántara, C., Levers, C., Prishchepov, A., Hostert, P., 2015. Mapping farmland abandonment and recultivation across Europe using MODIS NDVI time series. Remote Sensing of Environment 163, 312-325.

Estel, S., Mader, S., Levers, C., Verburg P.H., Baumann, M., 2018. Combining satellite data and agricultural statistics to map grassland management intensity in Europe. Environmental Research Letters 13, 074020.

Fang, J., Bai, Y., Wu, J., 2015. Towards a better understanding of landscape patterns and ecosystem processes of the Mongolian Plateau. Landsacpe Ecology 30, 1573-1578.

Flynn, E.S., Dougherty, C.T., Wendroth, O.J.A.J., 2008. Assessment of pasture biomass with the normalized difference vegetation index from active ground-based sensors. Agronomy Journal 100, 114-121.

FOAG, 2018. Swiss agricultural policy. https://www.cbd.int/financial/pes/swiss-pesagriculturalpolicy.pdf. (Accessed October 2018).

Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., Mueller, N.D., O'Connell, C., Ray, D.K., West, P.C., 2011. Solutions for a cultivated planet. Nature 478, 337-342.

Gao, T., Xu, B., Yang, X., Jin, Y., Ma, H., Li, J., Yu, H., 2013. Using MODIS time series data to estimate aboveground biomass and its spatio-temporal variation in Inner Mongolia's grassland between 2001 and 2011.International Journal of Remote Sensing 34, 7796-7810.

Garroutte, E.L., Hansen, A.J., Lawrence, R.L., 2016. Using NDVI and EVI to map spatiotemporal variation in the biomass and quality of forage for migratory elk in the Greater Yellowstone ecosystem. Remote Sensing 8, 404.

Gärtner, D., Keller, A., Schulin, R., 2013. A simple regional downscaling approach for spatially distributing land use types for agricultural land. Agricultural Systems 120, 10-19.

Giménez, M.G., de Jong, R., Della Peruta, R., Keller, A., Schaepman, M.E., 2017. Determination of grassland use intensity based on multi-temporal remote sensing data and ecological indicators. Remote Sensing of Environment 198, 126-139.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment 202, 18-27.

Gossner, M.M., Lewinsohn, T.M., Kahl, T., Grassein, F., Boch, S., Prati, D., Birkhofer, K., Renner, S.C., Sikorski, J., Wubet, T., 2016. Land-use intensification causes multitrophic homogenization of grassland communities. Nature 540, 266-269.

Green, S., Cawkwell, F., Dwyer, E., 2016. Cattle stocking rates estimated in temperate intensive grasslands with a spring growth model derived from MODIS NDVI time-series. International Journal of Applied Earth Observation and Geoinformation 52, 166-174.

Griffiths, P., Nendel, C., Pickert, J., Hostert, P., 2019. Towards national-scale characterization of grassland use intensity from integrated Sentinel-2 and landsat time series. Remote Sensing of Environment 238,111124.

Habel, J.C., Dengler, J., Janišová, M., Török, P., Wellstein, C., Wiezik, M., 2013. European grassland ecosystems: threatened hotspots of biodiversity. Biodiversity and Conservation 22, 2131-2138.

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics) 28, 100-108.

Henle, K., Alard, D., Clitherow, J., Cobb, P., Firbank, L., Kull, T., McCracken, D., Moritz, R.F., Niemelä, J., Rebane, M., 2008. Identifying and managing the conflicts between agriculture and biodiversity conservation in Europe–A review. Agriculture, Ecosystems & Environment 124, 60-71.

Hijmans, R.J., van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J.A., Lamigueiro, O.P., Bevan, A., Racine, E.B., Shortridge, A., 2017. Package 'raster'. R Foundation for Statistical Computing.

Holden, C.E., Woodcock, C.E., 2016. An analysis of Landsat 7 and Landsat 8 underflight data and the implications for time series investigations. Remote Sensing of Environment 185, 16-36.

Horn, B.K., 1981. Hill shading and the reflectance map. Proceedings of the IEEE 69, 14-47.

Hudewenz, A., Klein, A.-M., Scherber, C., Stanke, L., Tscharntke, T., Vogel, A., Weigelt, A., Weisser, W.W., Ebeling, A., 2012. Herbivore and pollinator responses to grassland management intensity along experimental changes in plant species richness. Biological Conservation 150, 42-52.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment 83, 195-213.

Huyghe, C., Vliegher, A.d., Goliński, P., 2014. European grasslands overview: temperate region. EGF at 50: The future of European grasslands. Proceedings of the 25th General Meeting of the European Grassland Federation, Aberystwyth, Wales, 7-11 September 2014. IBERS, Aberystwyth University, pp. 29-40.

Hyndman R.J., Fan, Y., 1996. Sample quantiles in statistical packages. The American Statistician 50, 361-365.

Isselstein, J., Kayser, M., 2014. Functions of grassland and their potential in delivering ecosystem services. The Future of European Grasslands, 199-214.

Jeangros, B., Thomet, P., 2004. Multi-functionality of grassland systems in Switzerland. Grassland Science in Europe 9, 11-23.

Jiang, Y., Tao, J., Huang, Y., Zhu, J., Tian, L., Zhang, Y., 2014. The spatial pattern of grassland aboveground biomass on Xizang Plateau and its climatic controls. Journal of Plant Ecology 8, 30-40.

Kennedy, R.E., Andréfouët, S., Cohen, W.B., Gómez, C., Griffiths, P., Hais, M., Healey, S.P., Helmer, E.H., Hostert, P., Lyons, M.B., 2014. Bringing an ecological view of change to Landsat-based remote sensing. Frontiers in Ecology and the Environment 12, 339-346.

Kennedy, R.E., Yang, Z., Cohen, W.B., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—Temporal segmentation algorithms. Remote Sensing of Environment 114, 2897-2910.

Kizeková, M., Hopkins, A., Kanianska, R., Makovníková, J., Pollák, Š., Pálka, B., 2018. Changes in the area of permanent grassland and its implications for the provision of bioenergy: Slovakia as a case study. Grass and Forage Science 73, 218-232.

Klaus, V.H., Kleinebecker, T., Boch, S., Müller, J., Socher, S.A., Prati, D., Fischer, M., Hölzel, N., 2012. NIRS meets Ellenberg's indicator values: Prediction of moisture and nitrogen values of agricultural grassland vegetation by means of near-infrared spectral characteristics. Ecological Indicators 14, 82-86.

Kleijn, D., Müller-Schärer, H., 2006. The relation between unpalatable species, nutrients and plant species richness in Swiss montane pastures. Biodiversity and Conservation 15, 3971-3982.

Kleijn, D., Baquero, R., Clough, Y., Diaz, M., De Esteban, J., Fernández, F., Gabriel, D., Herzog, F., Holzschuh, A., Jöhl, R., 2006. Mixed biodiversity benefits of agri-environment schemes in five European countries. Ecology Letters 9, 243-254.

Kleijn, D., Rundlöf, M., Scheper, J., Smith, H.G., Tscharntke, T., 2011. Does conservation on farmland contribute to halting the biodiversity decline? Trends in Ecology & Evolution 26, 474-481.

Kleijn, D., Sutherland, W.J., 2003. How effective are European agri-environment schemes in conserving and promoting biodiversity? Journal of Applied Ecology 40, 947-969.

Knop, E., Kleijn, D., Herzog, F., Schmid, B., 2006. Effectiveness of the Swiss agri-environment scheme in promoting biodiversity. Journal of Applied Ecology 43, 120-127.

Kolecka, N., Ginzler, C., Pazur, R., Price, B., Verburg, P., 2018. Regional Scale Mapping of Grassland Mowing Frequency with Sentinel-2 Time Series. Remote Sensing 10, 1221.

Kuemmerle, T., Erb, K., Meyfroidt, P., Müller, D., Verburg, P.H., Estel, S., Haberl, H., Hostert, P., Jepsen, M.R., Kastner, T., 2013. Challenges and opportunities in mapping land use intensity globally. Current Opinion in Environmental Sustainability 5, 484-493.

Lamarque, P., Tappeiner, U., Turner, C., Steinbacher, M., Bardgett, R.D., Szukics, U., Schermer, M., Lavorel, S.J.R.e.c., 2011. Stakeholder perceptions of grassland ecosystem services in relation to knowledge on soil fertility and biodiversity. Regional Environmental Change 11, 791-804.

LANAT, 2018. Cantonal Office for Agriculture and Envrionment, Berne. http://www.vol.be.ch/vol/de/index/landwirtschaft/landwirtschaft.html.(Accessed July 2018).

Landolt, E., Bäumler, B., Erhardt, A., Hegg, O., Klötzli, F., Lämmler, W., Nobis, M., Rudmann-Maurere, K., Schweingruber, F.H., Theurillat, J.P., Urmi, E., Vust, M., Wohlgemuth, T., 2010. Flora indicativa – ecological indicator values and biological attributes of the flora of Switzerland and the Alps (2nd ed.). Haupt Publishing, Berne, Switzerland.

Leifeld, J., Bassin, S., Fuhrer, J., 2005. Carbon stocks in Swiss agricultural soils predicted by land-use, soil characteristics, and altitude. Agriculture, Ecosystems & Environment 105, 255-266.

Leifeld, J., Fuhrer, J., 2005. Greenhouse gas emissions from Swiss agriculture since 1990: implications for environmental policies to mitigate global warming. Environmental Science & Policy 8, 410-417.

Li, F., Bai, Y., Wan, H., Zheng, J., Luo, J., Zhao, D., Liu, P., 2017. Quantifying Grazing Intensity in China Using High Temporal Resolution MODIS Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10, 515-523.

Liebisch, F., 2011. Plant and soil indicators to assess the phosphorus nutrition status of agricultural grasslands. ETH Zurich.

Liebisch, F., Bünemann, E.K., Huguenin-Eilie, O., Jeangros, B., Frossard, E., Oberson, A., 2013. Plant phosphorus nutrition indicators evaluated in agricultural grasslands managed at different intensities. European Journal of Agronomy 44, 67-77.

Liu, S., Cheng, F., Dong, S., Zhao, H., Hou, X., Wu, X., 2017. Spatiotemporal dynamics of grassland aboveground biomass on the Qinghai-Tibet Plateau based on validated MODIS NDVI. Scientific Reports 7, 4182.

Masé, G., 2005. The management of dry grassland in Switzerland. A Swiss federal program and its local practical application. Biotechnologie, Agronomie, Société et Environnement 9, 133-138.

MeteoSwiss, 2018a. Documentation of MeteoSwiss grid-data products. Daily mean, minimum and maximum temperature: TabsD, TminD, TmaxD. https://www.meteoswiss.admin.ch/content/dam/meteoswiss/de/service-und-publikationen/produkt/raeumliche-daten-temperatur/doc/ProdDoc_TabsD.pdf. (Accessed July 2018).

MeteoSwiss, 2018b. Documentation of MeteoSwiss grid-data products. Daily precipitation(-final analysis): RhiresD. https://www.meteoswiss.admin.ch/content/dam/meteoswiss/de/service-und-publikationen/produkt/raeumliche-daten-niederschlag/doc/ProdDoc_RhiresD.pdf. (Accessed July 2018).

Metzger, C.M.H., Heinichen, J., Eickenscheidt, T., Drosler, M., 2016. Impact of land-use intensity on the relationship between vegetation indices, photosynthesis and biomass of intensively and extensively managed grassland fens. Grass and Forage Science 72, 50-63.

Moog, D., Poschlod, P., Kahmen, S., Schreiber, K.F., 2002. Comparison of species composition between different grassland management treatments after 25 years. Applied Vegetation Science 5, 99-106.

Mottet, A., Ladet, S., Coqué, N., Gibon, A., 2006. Agricultural land-use change and its drivers in mountain landscapes: A case study in the Pyrenees. Agriculture, Ecosystems & Environment 114, 296-310.

Nagendra, H., Lucas, R., Honrado, J.P., Jongman, R.H., Tarantino, C., Adamo, M., Mairota, P., 2013. Remote sensing for conservation monitoring: Assessing protected areas, habitat extent, habitat condition, species diversity, and threats. Ecological Indicators 33, 45-59.

Nestola, E., Calfapietra, C., Emmerton, C.A., Wong, C., Thayer, D.R., Gamon, J.A., 2016. Monitoring grassland seasonal carbon dynamics, by integrating MODIS NDVI, proximal optical sampling, and eddy covariance measurements. Remote Sensing 8, 260.

Oenema, O., de Klein, C., Alfaro, M., 2014. Intensification of grassland and forage use: driving forces and constraints. Crop and Pasture Science 65, 524-537.

Orr, R., Murray, P., Eyles, C., Blackwell, M., Cardenas, L., Collins, A., Dungait, J., Goulding, K., Griffith, B., Gurr, S., 2016. The North Wyke Farm Platform: effect of temperate grassland farming systems on soil moisture contents, runoff and associated water quality dynamics. European Journal of Soil Science 67, 374-385.

Pauler, C.M., Isselstein, J., Braunbeck, T., Schneider, M.K., 2019. Influence of highland and production-oriented cattle breeds on pasture vegetation: A pairwise assessment across broad environmental gradients. Agriculture, Ecosystem & Environment 284, 106585.

Pe'er, G., Dicks, L., Visconti, P., Arlettaz, R., Báldi, A., Benton, T., Collins, S., Dieterich, M., Gregory, R., Hartig, F., 2014. EU agricultural reform fails on biodiversity. Science 344, 1090-1092.

Peeters, A., Beaufoy, G., Canals, R., De Vliegher, A., Huyghe, C., Isselstein, J., Jones, J., Kessler, W., Kirilovsky, D., Van Den Pol-Van Dasselaar, A., 2014. Grassland term definitions and classifications adapted to the diversity of European grassland-based systems. 25th EGF General Meeting on "EGF at 50: The Future of European Grasslands, pp. 743-750.

Peter, M., Edwards, P.J., Jeanneret, P., Kampmann, D., Lüscher, A., 2008. Changes over three decades in the floristic composition of fertile permanent grasslands in the Swiss Alps. Agriculture, Ecosystem & Environment 125, 204-212.

Porqueddu, C., Ates, S., Louhaichi, M., Kyriazopoulos, A., Moreno, G., del Pozo, A., Ovalle, C., Ewing, M., Nichols, P., 2016. Grasslands in 'Old World'and 'New World'Mediterranean-climate zones: past trends, current status and future research priorities. Grass and Forage Science 71, 1-35.

Porter, T.F., Chen, C., long, J.A., Lawrence, R.L., Sowell, B.F., 2014. Estimating biomass on CRP pastureland: A comparison of remote sensing techniques. Biomass Bioenergy 66, 268-274.

Power, A.G., 2010. Ecosystem services and agriculture: tradeoffs and synergies. Philosophical Transactions of the Royal Society of London B: Biological Sciences 365, 2959-2971.

Price, B., Kienast, F., Seidl, I., Ginzler, C., Verburg, P.H., Bolliger, J., 2015. Future landscapes of Switzerland: Risk areas for urbanisation and land abandonment. Applied Geography 57, 32-41.

PRIF, 2017. Principles of fertilization of agricultural crops in Switzerland (PRIF 2017), Agroscope, Switzerland. https://www.agroscope.admin.ch/agroscope/en/home/topics/plant-production/field-crops/Pflanzenernaehrung/grud.html. (Accessed August 2018).

Psomas, A., Kneubühler, M., Huber, S., Itten, K., Zimmermann, N., 2011. Hyperspectral remote sensing for estimating aboveground biomass and for exploring species richness patterns of grassland habitats. International Journal of Remote Sensing 32, 9007-9031.

Pywell, R.F., Heard, M.S., Bradbury, R.B., Hinsley, S., Nowakowski, M., Walker, K.J., Bullock, J.M., 2012. Wildlife-friendly farming benefits rare birds, bees and plants. Biology Letters 8, 772-775.

R, Team, C., 2016. R: A language and environment for statistical computing. pp. http://www.R-project.org/; accessed: March 2016.

Rodríguez, J.P., Beard Jr, T.D., Bennett, E.M., Cumming, G.S., Cork, S.J., Agard, J., Dobson, A.P., Peterson, G.D., 2006. Trade-offs across space, time, and ecosystem services. Ecology and Society 11, 1-28.

Rose, L., Coners, H., Leuschner, C., 2012. Effects of fertilization and cutting frequency on the water balance of a temperate grassland. Ecohydrology 5, 64-72.

Rose, L., Leuschner, C., 2012. The diversity–productivity relationship in a permanent temperate grassland: negative diversity effect, dominant influence of management regime. Plant Ecology & Diversity 5, 265-274.

Rouse Jr, J.W., Haas, R., Schell, J., Deering, D., 1974. Monitoring vegetation systems in the Great Plains with ERTS. University of Texas, Paper A20.

Sakowska, K., Juszczak, R., Gianelle, D., 2016. Remote sensing of grassland biophysical parameters in the context of the Sentinel-2 satellite mission. Journal of Sensors 2016, 4612809.

Schermer, M., Darnhofer, I., Daugstad, K., Gabillet, M., Lavorel, S., Steinbacher, M., 2016. Institutional impacts on the resilience of mountain grasslands: an analysis based on three European case studies. Land Use Policy 52, 382-391.

Schmidt, G., Jenkerson, C., Masek, J., Vermote, E., Gao, F., 2013. Landsat ecosystem disturbance adaptive processing system (LEDAPS) algorithm description. US Geological Survey.

Schweiger, A.K., Risch, A.C., Damm, A., Kneubühler, M., Haller, R., Schaepman, M.E., Schütz, M., 2015. Using imaging spectroscopy to predict above-ground plant biomass in alpine grasslands grazed by large ungulates. Journal of Vegetation Science 26, 175-190.

SFA, 2018. Swiss Federal Administration, administrative order for agricultural terminology (910.91). https://www.admin.ch/opc/de/classified-compilation/19983381/201801010000/910.91.pdf. (Accessed July 2018).

SFSO, 2018. Swiss Federal Statistical Office: land use in Switzerland. Results of the Swiss land use statitics. https://www.bfs.admin.ch/bfs/en/home.html. (Accessed July 2018).

Simons, N.K., Weisser, W.W., 2017. Agricultural intensification without biodiversity loss is possible in grassland landscapes. Nature Ecology & Evolution 1, 1136.

Smit, H., Metzger, M., Ewert, F., 2008. Spatial distribution of grassland productivity and land use in Europe. Agricultural Systems 98, 208-219.

Socher, S.A., Prati, D., Boch, S., Müller, J., Baumbach, H., Gockel, S., Hemp, A., Schöning, I., Wells, K., Buscot, F., 2013. Interacting effects of fertilization, mowing and grazing on plant species diversity of 1500 grasslands in Germany differ between regions. Basic and Applied Ecology 14, 126-136.

Soussana, J.-F., Lemaire, G., 2014. Coupling carbon and nitrogen cycles for environmentally sustainable intensification of grasslands and crop-livestock systems. Agriculture, Ecosystems & Environment 190, 9-17.

Stumpf, F., Keller, A., Schmidt, K., Mayr, A., Gubler, A., Schaepman, M., 2018. Spatio-temporal land use dynamics and soil organic carbon in Swiss agroecosystems. Agriculture, Ecosystems & Environment 258, 129-142.

Sulla-Menashe, D., Kennedy, R.E., Yang, Z., Braaten, J., Krankina, O.N., Friedl, M.A.J.R.S.o.E., 2014. Detecting forest disturbance in the Pacific Northwest from MODIS time series using temporal segmentation. Remote Sensing of Environment 151, 114-123.

Svoray, T., Perevolotsky, A., Atkinson, P.M., 2013. Ecological sustainability in rangelands: the contribution of remote sensing. International Journal of Remote Sensing 34, 6216-6242.

SwissTopo, 2018. Federal Office of Topography. https://www.swisstopo.admin.ch/en/home.html. (Accessed July 2018).

Tasser, E., Tappeiner, U., 2002. Impact of land use changes on mountain vegetation. Applied Vegetation Science 5, 173-184.

Taugourdeau, S., Huguenin-Elie, O., Plantureux, S., Michel, N., Dumont, B., 2013. A Grass Use Intensity index to be used across regions and grassland managements. The role of grasslands in a green future: threats and perspectives in less favoured areas. Proceedings of the 17th Symposium of the European Grassland Federation, Akureyri, Iceland, 23-26 June 2013. Agricultural University of Iceland, pp. 397-399.

Tälle, M., Deák, B., Poschlod, P., Valkó, O., Westerberg, L., Milberg, P., 2016. Grazing vs. mowing: A meta-analysis of biodiversity benefits for grassland management. Agriculture, Ecosystems & Environment 222, 200-212.

Weber, D., Schaepman-Strub, G., Ecker, K., 2018. Predicting habitat quality of protected dry grasslands using Landsat NDVI phenology. Ecological Indicators 91, 447-460.

Wilson, J.B., Peet, R.K., Dengler, J., Pärtel, M., 2012. Plant species richness: the world records. Journal of Vegetation Science 23, 796-802.

Wolberg, G., Alfy, I., 1999. Monotonic cubic spline interpolation. 1999 Proceedings Computer Graphics International, pp. 188-195.

Zaks, D.P., Kucharik, C.J., 2011. Data and monitoring needs for a more ecological agriculture. Environmental Research Letters 6, 014017.

Zeeman, M.J., Hiller, R., Gilgen, A.K., Michna, P., Plüss, P., Buchmann, N., Eugster, W., 2010. Management and climate impacts on net CO2 fluxes and carbon budgets of three grasslands along an elevational gradient in Switzerland. Agricultural and Forest Meteorology 150, 519-530.

Zeileis, A., Grothendieck, G., Ryan, J.A., Andrews, F., Zeileis, M.A., 2018. Package 'zoo'.

Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C., Gao, F., Reed, B.C., Huete, A., 2003. Monitoring vegetation phenology using MODIS. Remote Sensing of Environment 84, 471-475.

Zhou, G., Zhou, X., He, Y., Shao, J., Hu, Z., Liu, R., Zhou, H., Hosseinibai, S., 2017. Grazing intensity significantly affects belowground carbon and nitrogen cycling in grassland ecosystems: a meta-analysis. Global Change Biology 23, 1167-1179.

Zhu, Z., Wang, S., Woodcock, C.E., 2015. Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. Remote Sensing of Environment 159, 269-277.

Zhu, Z., Woodcock, C.E., Olofsson, P., 2012. Continuous monitoring of forest disturbance using all available Landsat imagery. Remote Sensing of Environment 122, 75-91.

# Appendix

## Scientific publications and conference contributions

### Peer-reviewed publications

**Rentschler, T.**, Werban, U., Ahner, M., Behrens, T., Gries, P., Scholten, T., Teuber, S., Schmidt, K. (2020): 3D mapping of soil organic carbon content and soil moisture with multiple geophysical sensors and machine learning. Vadose Zone Journal 19, 1. doi: 10.1002/vzj2.20062

Taghizadeh-Mehrjardi, R., Schmidt, K., Amirian-Chakan, A., **Rentschler, T.**, Zeraatpisheh, M., Sarmadian, F., Valavi, R., Davatgar, N., Behrens, T., Scholten, T. (2020): Improving the Spatial Prediction of Soil Organic Carbon Content in Two Contrasting Climatic Regions by Stacking Machine Learning Models and Rescanning Covariate Space. Remote Sensing 12 1095. doi: 10.3390/rs12071095

Stumpf, F., Schneider, M.K., Keller, A., Mayr, A., **Rentschler, T.**, Meuli, R.G., Schaepman, M., Liebisch, F. (2020): Spatial monitoring of grassland management using multi-temporal satellite imagery. Ecological Indicators 113, 106201. doi: 10.1016/j.ecolind.2020.106201

**Rentschler, T.**, Gries, P., Behrens, T., Bruelheide, H., Kühn, P., Seitz, S., Shi, X., Trogisch, S., Scholten, T., Schmidt, K., 2019. Comparison of catchment scale 3D and 2.5D modelling of soil organic carbon stocks in Jiangxi Province, PR China. PloS one 14 (8), e0220881. doi: 10.1371/journal.pone.0220881

Schmaltz, E.M., Rosner, H.-J., **Rentschler, T.**, Märker, M. (2017): Assessment of groundwater response and soil moisture fluctuations in the Mugello Basin (central Italy). Geography, Environment, Sustainability (GES Journal) 10/2, 15-27. doi: 10.24057/2071-9388-2017-10-2-15-27

### Conference contributions

**Rentschler, T.**, Bartelheim, M., Díaz-Zorita Bonilla, M., Gries, P., Scholten, T., Schmidt, K. (2020): Volumetric soil quality modelling with machine learning in a diverse agricultural landscape in Andalusia, Spain. EGU General Assembly 2020, Vienna. [poster]

Bindereif, L., **Rentschler, T.**, Batelheim, M., Díaz-Zorita Bonilla, M., Gries, P., Scholten, T., Schmidt, K. (2020): Synthetic sampling for spatio-temporal land cover mapping with machine learning and the Google Earth Engine in Andalusia, Spain. EGU General Assembly 2020, Vienna. [poster]

Scholten, T., **Rentschler, T.**, Taghizadeh-Mehrjardi, R., Schmidt, K. (2019): Prediction of Drought Risk and Soil Quality using FTIR and ML - Examples from Prehistoric Andalusia, Spain. Interdisciplinary and International Workshop 2019, Linares, Spain. [talk]

**Rentschler, T.**, Bartelheim, M., Díaz-Zorita Bonilla, M., Scholten, T., Schmidt, K. (2019): Convolutional neural networks und Nahinfrarotspektroskopie für die Prognose von Bodenqualitätsindikatoren. Jahrestagung der Deutschen Bodenkundlichen Gesellschaft 2019, Kommission V /Digital Soil Mapping, 25. –2019, Bern, Schweiz. [talk]

**Rentschler, T.**, Ahner, M., Behrens, T., Gries, P., Kühn, P., Werban, U., Scholten, T., Schmidt, K.: Three-dimensional mapping of soil organic carbon and soil water content with proximal soil sensing data. EGU General Assembly 2019, Vienna. [PICO]

Bindereif, L., **Rentschler, T.**, Bartelheim, M., Díaz-Zorita Bonilla, M., Gries, P., Scholten, T., Schmidt, K.: Analysis and mapping of spatio-temporal land use dynamics in Andalusia, Spain using the Google Earth Engine cloud computing platform and the Landsat archive. EGU General Assembly 2019, Vienna. [poster]

**Rentschler, T.**, Ahner, M., Behrens, T., Gries, P., Kühn, P., Werban, U., Scholten, T., Schmidt, K.: Dreidimensionale Kartierung des organischen Bodenkohlenstoffs und des Bodenwassergehalts mit geophysikalischen Methoden. 7th Digital Soil Mapping Workshop, Deutsche Bodenkundliche Gesellschaft (DBG), AG DSM 2019, Tübingen. [talk]

**Rentschler, T.**, Gries, P., Behrens, T., Bruelheide, H., Kühn, P., Seitz, S., Trogisch, S., Scholten, T., Schmidt, K.: Comparison of catchment scale 3D and 2.5D modelling of soil organic carbon stocks in Jiangxi Province, PR China. 2nd ISMC Conference: New Perspectives on Soil Models 2018, Wageningen, Netherlands. [poster]

Stumpf, F., Keller, A., Mayr, A., **Rentschler, T.**, Schneider, M., Meuli, R., Schaepman, M., Liebisch, F.: Spatial monitoring of grassland management and plant diversity in Switzerland. SUSALPS Conference 2018, Garmisch-Partenkirchen, Germany. [talk]

Taghizadeh-Mehrjardi, R., Schmidt, K., Eftekhari K., **Rentschler, T.**, Scholten, T.: Updating the categorical soil map of Iran using limited soil legacy data. 21st World Congress of Soil Science 2018, Rio de Janeiro, Brazil. [poster]

**Rentschler, T.**, Schmidt, K.: Learning QGIS - Getting started with geospatial analysis and mapping. In Poseidons Reich XXIII, Jahrestagung der DEGUWA e.V. 2018 [workshop]

## Appendix

**Rentschler, T.**, Kühn, P., Scholten, T., Schmidt, K.: Dreidimensionale Kartierung von SOC basierend auf maschinellem Lernen in Jiangxi, VR China. Deutscher Kongress für Geographie 2017, Tübingen, Germany. [talk]

**Rentschler, T.**, Kühn, P., Behrens, T., Scholten, T., Schmidt, K.: Dreidimensionale Modellierung von organischem Kohlenstoff im Boden basierend auf multi-skaliger Reliefanalyse und Methoden des Data Minings in Jiangxi, VR China. Jahrestagung der Deutschen Bodenkundlichen Gesellschaft (DBG) 2017, Göttingen, Germany. [talk]

**Rentschler, T.**, Kühn, P., Scholten, T., Schmidt, K.: Three-dimensional mapping of soil organic carbon (SOC) based on multi-scale digital terrain analysis and data mining in Jiangxi Province, PR China. Pedometrics 2017, Wageningen, Netherlands. [poster]

**Acknowledgements**

First, I want to thank Dr. Karsten Schmidt for offering this fascinating and challenging research topic. I am very grateful for his support, trust, discussions, help and patience through the past three years. Further, I thank Prof. Dr. Thomas Scholten for the opportunity to become part of his working group and the SFB 1070 ResourceCultures. He supported me in many aspects, which allowed me to find my way through field work, conferences, teaching, and this thesis.

I am grateful to Prof. Dr. Martin Bartelheim to become part of the SFB 1070 and for the fieldwork in Andalusia. I further thank Dr. Hans-Joachim Rosner and Dr. Thorsten Behrens, who brought me to geoinformatics and machine learning. And I thank Prof. Dr. Volker Hochschild for kindly refereeing this thesis.

It has been a great experience to be part of the Soil Science and Geomorphology group and the SFB 1070. Many thanks go to all SFB and soil science colleagues, especially to Laura Bindereif, Annika Condit, Jonas Froehlich, Corinna Gall, Philipp Gries, Benjamin Höpfer, Arne Meier, Maike Melles, Nadja Mozdzen, Valerie Palmowski, Kerstin Rau, Sascha Scherer as well as Dr. Marta Díaz-Zorita Bonilla, Dr. Tobias Schade, Dr. Steffen Seitz, and Dr. Sandra Teuber for their support, help and the good times with ice-cream, bouldering, cycling, and Wednesday Stammtisch.

Special thanks go to Dr. Peter Kühn, Sabine Flaiz, Rita Mögenburg and the student assistants of the Laboratory of Soil Science and Geoecology, University of Tübingen, Dr. Susan Mentzer and Matthias Czechowski from the working group Geoarchaeology, University of Tübingen, and Renate Nitschke from the lab of the Chair of Silviculture, University of Freiburg. And I thank the student assistants Marvin Schmidt and Carolin Walper, who helped me with data analysis.

Last but not least, I thank my family for their constant and never-ending support during my studies.