Spatial modelling and mapping of stable soil organic carbon, sequestration potentials and deficits in agricultural and grassland top soils of Lower Austria

MASTER THESIS

to obtain the academic joint degree
„Master of Science"

Submitted by
FRANZ JOSEF KILIAN MAYR

Vienna, 08 December 2020

Supervisor:
Univ.Prof. Dipl.-Ing. Dr.nat.techn. Walter W. Wenzel
BOKU, Vienna, Austria

Co-Supervisor:
Prof. Ing. Pavel Tlustos, CSc.
ČZU, Prague, Czech Republic

University of Natural Resources and Life Science

Vienna, Austria

Department of Forest- and Soil Sciences

Institute of Soil Research

University of Natural Resources
and Life Sciences, Vienna

Czech University of Life Sciences, Prague

Faculty of Agrobiology, Food and Natural Resources

Department of Agroenvironmental Chemistry and Plant Nutrition

CZECH
UNIVERSITY
OF LIFE SCIENCES PRAGUE

Affirmation

I certify that the Master Thesis was written by me, not using sources and tools other than quoted and without use of any other illegitimate support.

Furthermore, I confirm that I have not submitted this master thesis elsewhere, either nationally or internationally in any form.

Vienna, 08 December 2020   Franz Josef Kilian Mayr   signature

Acknowledgement

I would like to sincerely thank my supervisor Univ.Prof. Dipl.-Ing. Dr.nat.techn. Walter Wenzel for the continuous support, insightful comments, immense knowledge, patience and guidance throughout the time of writing this thesis.

I also want to thank the Bundesforschungszentrum für Wald and the Institute of Metrology of BOKU, especially Dipl.-Ing. Günther Aust and Assoc. Prof. Dr. Herbert Formayer, for providing important data for my work

Further I want to thank the Institute of Soil research at BOKU for providing for me a working space and Computational Equipment.

I am also grateful for Jakob Schmidt, Robindro von Gierke, Samuel Jeske and Tobias Blessing for the proofreading of this thesis.

Last, I am thankful for my family, especially my parents, for supporting me and my studies and being so patient with me.

Abstract

Keywords: *Digital Soil Mapping, Soil Organic Carbon, Soil, Land Use, Stable Soil Organic Carbon, Carbon Deficit, Machine Learning, Random Forest, Soil Management*

Climate change and mitigation measures are important topics in science and public attention these days. Carbon storage in soil has been claimed as one potential measure of mitigation. In addition, soil carbon is linked to many important soil functions. For example, it positively influences soil fertility and cushions the negative effect of climate change on soil attributes. Especially in agricultural context soil organic carbon (SOC) is a good indicator for soil health. Different land management techniques can be used in advisory level, to have a positive impact in improving SOC levels over time. For targeted action, detailed saturation or deficit maps for consultancy are needed. Due to the increase in availability of high-resolution environmental data and increase of computing power, these maps can be created cost efficiently with several Digital Soil Mapping (DSM) methods and with the help of related spatial environmental layers. Stable soil organic carbon (SSOC), soil organic carbon saturation potential (Csat) and soil organic carbon deficits (Cdef) are important key figures to identify areas that need improvement through agricultural management. To track the situation for the province of Lower Austria for arable and grassland soils to a depth of 20 cm, SSOC, Csat and Cdef were calculated for the sampling points. Subsequently, three different DSM methods were compared for prediction accuracy: Stepwise Regression Kriging, Random Forest and Support Vector Machine Learning. In addition, different sets of covariates were tested. Kriging performed best regarding accuracy but was too time intensive. Random Forest achieved the second-best results, but in a reasonable time. It was used as the method for the final prediction of SSOC, Csat and Cdef, in combination with all available environmental covariates' layers including data from the Austrian soil map. The results were then prepared via Q-GIS into maps which can be used for policy making and agricultural advisory activities. In particular, these maps can be helpful for identifying areas, where soil health can be improved through change in management. This would improve soil fertility, increase water retention and hinder soil erosion. The point data which is used for prediction was gathered 1990/91. Therefore, the final prediction maps will only give a baseline map for further research and later comparison.

Zusammenfassung

Schlüsselwörter: Digitale Bodenkartierung, Organischer Kohlenstoff im Boden, Boden, Landnutzung, Stabiler organischer Kohlenstoff im Boden, Kohlenstoffdefizit, Maschinelles Lernen, Random Forest, Bodenmanagement

Der Klimawandel und Klimaschutzmaßnahmen sind heutzutage ein wichtiges Thema in den Nachrichten und wissenschaftlichen Veröffentlichungen. Dabei ist die Kohlenstoffspeicherung im Boden eine mögliche Gegenmaßnahme. Darüber hinaus ist der Bodenkohlenstoff mit vielen wichtigen Bodenfunktionen verknüpft. Zum Beispiel beeinflusst er die Bodenfruchtbarkeit positiv und federt die negativen Auswirkungen des Klimawandels auf die Bodeneigenschaften ab. Insbesondere im landwirtschaftlichen Kontext ist organischer Kohlenstoff im Boden (SOC) ein guter Indikator für die Bodengesundheit. Verschiedene Landmanagement-Techniken können in der Beratung eingesetzt werden, um SOC-Werte im Laufe der Zeit positiv zu verbessern. Dazu sind detaillierte Karten für die Beratung erforderlich. Aufgrund der zunehmenden Verfügbarkeit von hochauflösenden Umweltdaten und der Zunahme der Rechenleistung können diese Karten kosteneffizient mit verschiedenen Methoden der digitalen Bodenkartierung (DSM) und mit der Hilfe verschiedener räumlicher Umweltlayern erstellt werden. Stabiler organischer Kohlenstoff im Boden (SSOC), Sättigungspotential des organischen Kohlenstoffs im Boden (Csat) und Defizite des organischen Kohlenstoffs im Boden (Cdef) sind wichtige Kenngrößen zur Identifizierung von Gebieten, die durch landwirtschaftliches Management verbessert werden müssen. Um die Situation für das Land Niederösterreich für Acker- und Grünlandböden bis zu einer Tiefe von 20 cm zu erfassen, wurden SSOC, Csat und Cdef für die Probenahmepunkte berechnet. Im Weiteren wurden drei verschiedene DSM-Methoden zur Vorhersage getestet. Es handelt sich dabei um Stepwise Regression Kriging, Random Forest und Support Vector Machine. Darüber hinaus wurden auch verschiedene Sets von Kovariaten getestet. Kriging schnitt hinsichtlich der Genauigkeit am besten ab, war aber zu zeitintensiv. Random Forest hatte die zweitbesten Ergebnisse, bei geringerer Rechendauer. Er wurde als Methode für die endgültige Vorhersage und in Kombination mit allen verfügbaren Umweltlayern einschließlich der Daten aus der österreichischen Bodenkarte verwendet. Dabei wurden SSOC, Csat und Cdef vorhergesagt. Die Ergebnisse wurden dann in Q-GIS zu Karten aufbereitet, die für politische Entscheidungen landwirtschaftliche Beratungstätigkeiten verwendet werden können. Insbesondere können diese Karte bei der Identifizierung von Gebieten hilfreich sein, in denen die Bodengesundheit durch Managementveränderungen verbessert werden kann. Dies würde die Bodenfruchtbarkeit verbessern, die Wasserrückhaltung erhöhen und die Bodenerosion verhindern. Die Punktdaten, die für die Vorhersage verwendet werden, wurden 1990/91 gesammelt. Daher sind die endgültigen Vorhersagekarten nur eine Basiskarte für weitere Untersuchungen und spätere Vergleiche. Zusätzlich sind sie erste Referenzkarten für Beratungstätigkeiten bis zur Entnahme neuer Bodenproben.

Table of content

Abbrevations

BZI   Soil Status Inventory (German: Bodenzustandsinventur)

C   Carbon

CEC   Cation exchange capacity

Cdef   Carbon deficit

CH_4 Methane

Csat   Carbon saturation potential

Cstock Carbon stock

COV   Covariate

DSM   Digital soil mapping

E   East

eBOD Austrian digital soil map

f<20 µm Mineral particle fraction <20 µm

GHG   Greenhouse Gas Emissions

fsilt   Fine silt fraction (2–20 µm)

N_2O Nitrous oxide

R   Programming language for statistical computing and graphics

RK   Regression Kriging

RF   Random Forest

SOC   Soil organic carbon

SOCS Soil organic carbon stocks Mg ha^{-1}

SOM   Soil organic matter

SSOC Stable soil organic carbon

TWI   Topographic Wetness Index

1 Introduction

Carbon (C) is one of the essential elements for life on earth (Hazen et al., 2012; Oró et al., 1990). Through the extraction and burning of fossil carbon, humanity released huge amounts of carbon dioxide (CO_2) into the atmosphere (Le Quéré et al., 2009; Zeebe et al., 2016). Besides the massive benefits for cheap energy and consequently increased wealth and development, it massively influenced the global carbon cycle and led to global warming, which is threatening our livelihood (Roston, 2010). The global C cycle divides into atmospheric, marine and terrestrial sub cycles, with different pools and different residence times of C, which are interlinked through C-dynamic processes (Battin et al., 2009; Dawson and Smith, 2007; Heimann, 1993; Post et al., 1982; Schimel, 1995). The largest terrestrial organic C pool is soil with ~1,500 Pg C to a depth of 1 m and ~ 2,400 Pg C to 2 m depth (Batjes, 1996), compared to that, the current amount of C stored in the atmosphere as CO_2 is currently one third (~830 Pg C) of it. The annual fossil fuel emissions with 10 Pg are less (Ciais et al., 2013). Therefore, it underlines the importance of preservation and if possible, an increase of the soils organic C pool. Through intensive agricultural use, land use change and degradation SOC pools losses are estimated at around 456 Gt until 2010 with no significant reduction until now (Houghton et al., 2012; Ruddiman, 2003). Additionally with intensive agriculture and change of land use higher methane (CH_4) and nitrous oxide (N_2O) as GHG emissions have to be taken into account (Paustian et al., 2016)

Part of the terrestrial pool is soil organic matter (SOM). It can have turnover times in soil ranging from weeks to several decades, depending on its chemical origin and environmental conditions (Schmidt et al., 2011). SOM can be subdivided in following measurable pools. biochemically-protected, silt- and clay-protected, microaggregate-protected and unprotected C pool (Six et al., 2002) and, in turn part of the different SOM pools are SOC. The internationally accepted operational definition of SOC is the organic carbon present in the fraction of the soil that passes through a 2 mm sieve (Whitehead and Tinsley, 1964), which is the fine earth fraction. The silt- and clay protected part of SOC is called stable soil organic carbon (SSOC) (Angers et al., 2011). Hassink et al. (1997) proposed the concept of soil carbon saturation potential (Csat) based on the particle fraction <20 µm (clay + fine silt), with a least-squares linear regression. This was confirmed by Angers (1998) and by Stewart et al. (2007). Feng et al. (2013) recalculated Csat for different soils with a upper boundary line. This showed that due to the variance in soil condition no universal formula for calculation should be used. Instead it should be calculated based on regional information. The Csat can then be used to calculate the soil carbon saturation deficit (Cdef) through subtraction with SSOC (Angers et al., 2011), in order to evaluate areas for long term carbon management and to use the calculated value of Cdef in policy decision making (O'Rourke et al., 2015). It was then further discussed how to use Cdef for the "4 per 1000 initiative" to increase global carbon stocks (SOCS). It was proposed to aim for an attainable model, based on Cstock values, because of the asymptoticly enrichment of SSOC (Barré et al., 2017; Lal, 2016a).

This study, however, will focus on the SSOC, Cdef and Csat parameters because of the lack of bulk density information in our data set and research area. In the following three different digital soil mapping (DSM) techniques were tested and the three parameters for agricultural and grassland soils of Lower Austria from $0-20$ cm were spatially predicted. The predictions with environmental covariates (COVs) are based on the soil forming factors principle (Jenny,

1

1941) which were recently adapted to the S.C.O.R.P.A.N. factors for DSM by (McBratney et al., 2003). This approach differentiates from the traditional approach of reference areas ("Landesmusterstücke und Referenzprofile") and is using spatial data. The final goal is to create maps that will help policymakers and advisors to define long-term and small-scale local measures for farmers to increase SSOC in the soil.

The following research questions are defined and tested within this study

- Do high resolution predictions of SSOC, Csat and Cdef with environmental covariates delivers usable maps of acceptable accuracy and resolution?
- Will covariates of the Austrian soil map significantly improve SSOC, Csat and Cdef predictions?
- Does land use have a major influence on Cdef?

2 Materials and Methods

2.1 Material

2.1.1 Study Area

The study area is confined by the province of Lower Austria. It has a total area of 19,186 km². In the year of 2018 675,295 ha were used as arable land and 167,626 ha as permanent grassland (Statistik Austria, 2018). In this study, only areas with this type of land use are utilized. The targeted areas and their spatial distribution can be seen in Fig.2-1. In the Weinviertel, north-east of Lower Austria, mainly arable land is dominant. South, at the foothills of the Alps, grassland is dominant. North-west in the Waldviertel arable and grassland is equally distributed.
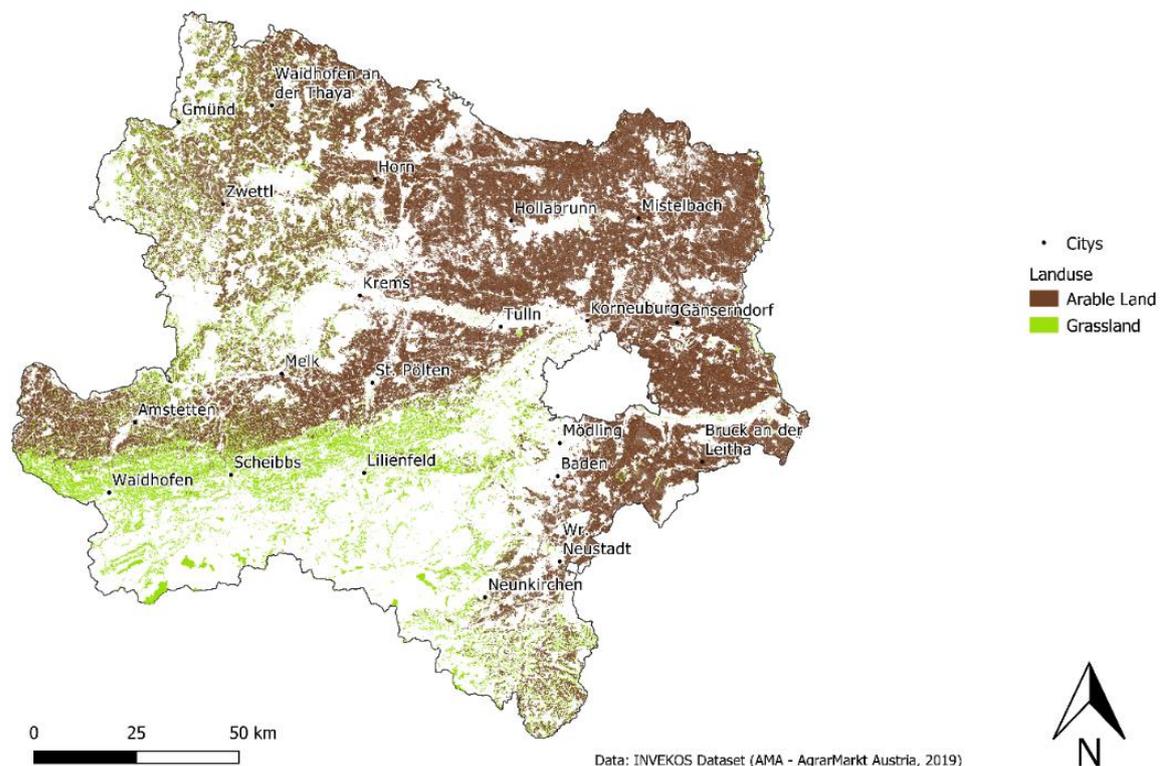

Figure 2-1 Land use in Lower Austria assigned to arable and grassland

Lower Austria's climate varies from a dry continental lowland climate in the northeast to a continental highland climate in the northwest. The Marchfeld and Viennese basin in the east have a Pannonian climate with low precipitation and hot summers, but only moderately cold winters. The remaining area can be defined as alpine transition climate and mountain humid climate in the southwest (Niederösterreich, 1994; Strauss et al., 2013). This can be seen in the following maps with average temperature of 8,5°C Fig.2 -2 and the average precipitation Fig.2-3 of the last Climate Normal from 1971 to 2000 (Hiebl et al., 2011).
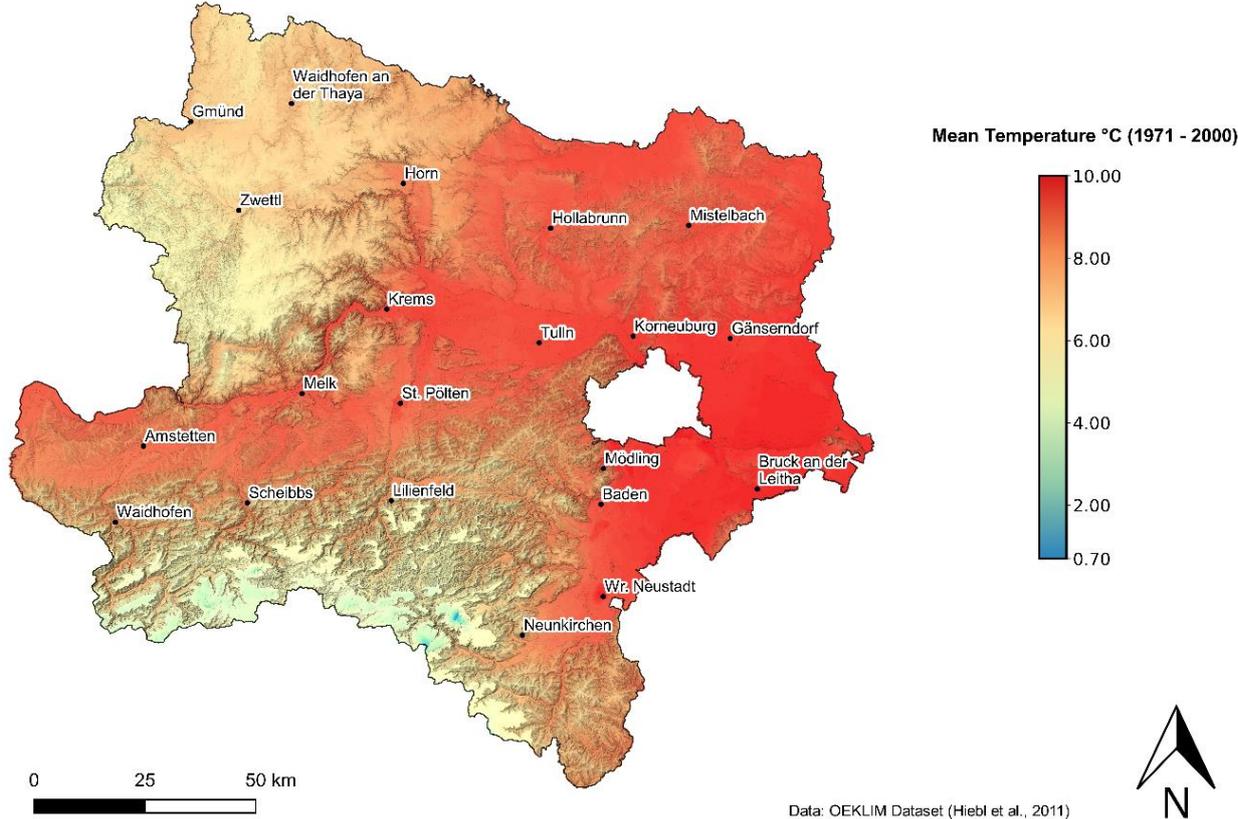

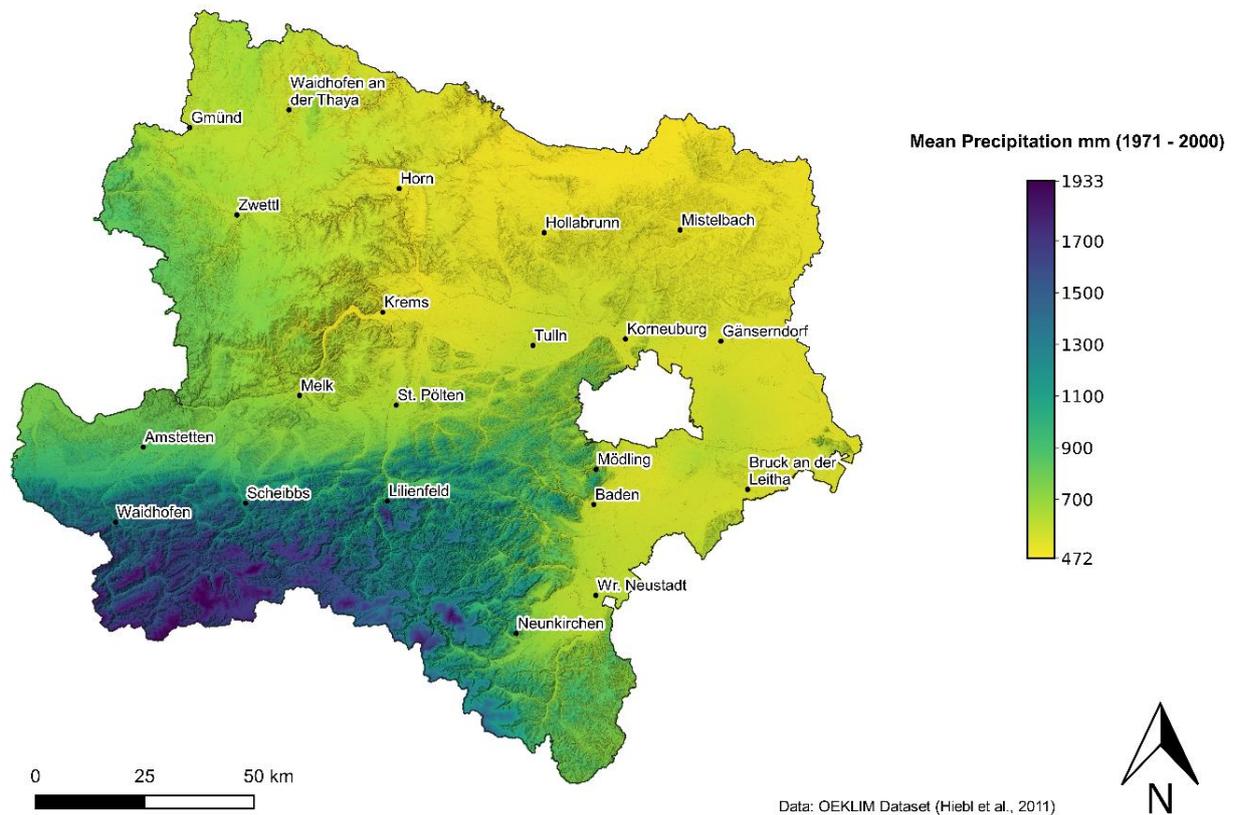**Figure 2-2 Mean Temperature map of Lower Austria**

Figure 2-3 Mean Precipitation map of Lower Austria

The geology of Lower Austria is very diverse. From the northwest it starts with the Bohemian Massif in the Waldviertel, followed by the molasse zone in the south and east and transitioning to the Alpine Massif in the south west. The Alpine Massif is divided into the Flysch zone, the Calcareous Alps and in the southern Greywacke zone. Located in the southwest is the central Alp Massif. The east of Lower Austria is characterized by the alluvium of the Danube with the Vienna Basin and Marchfeld nearby (Wessely and Draxler, 2006).

The described geological variation is reflected by the soil upon see Fig-2-4 through soil forming from parent rock. The Weinviertel and Marchfeld soil has mainly Tschernosem (Chernozems) with some parts of Feuchtschwarzerde (Gleyic Phaeozem) and Braunerde (Cambisols). Near rivers and the Danube there is alluvial soil (Fluvisols) present. The Waldviertel soils are mainly Podsols (Albic Podsols), Braunerden (Cambisols) and Parabraunerden (Luvisols) and some Pseudogleye in the northern part. The major soil types in the Flysh zone and the Mostviertel are Pseudogleye (Stagnosols) and Parabraunerden (Luvisols). In the mountainous parts int the south south-east of Lower Austria Rendzina are the common soils. In the Industrieviertel southeast of Vienna the soils are very divers, ranging from Feuchtschwarzerden (Gleyic Phaeozem) to Pararendzina, Tschernosems (Chernozems) and Podsols (Albic Podsols), amongst others. The translations in brackets are according to the international soil systematics, IUSS Working Group WRB, 2006.
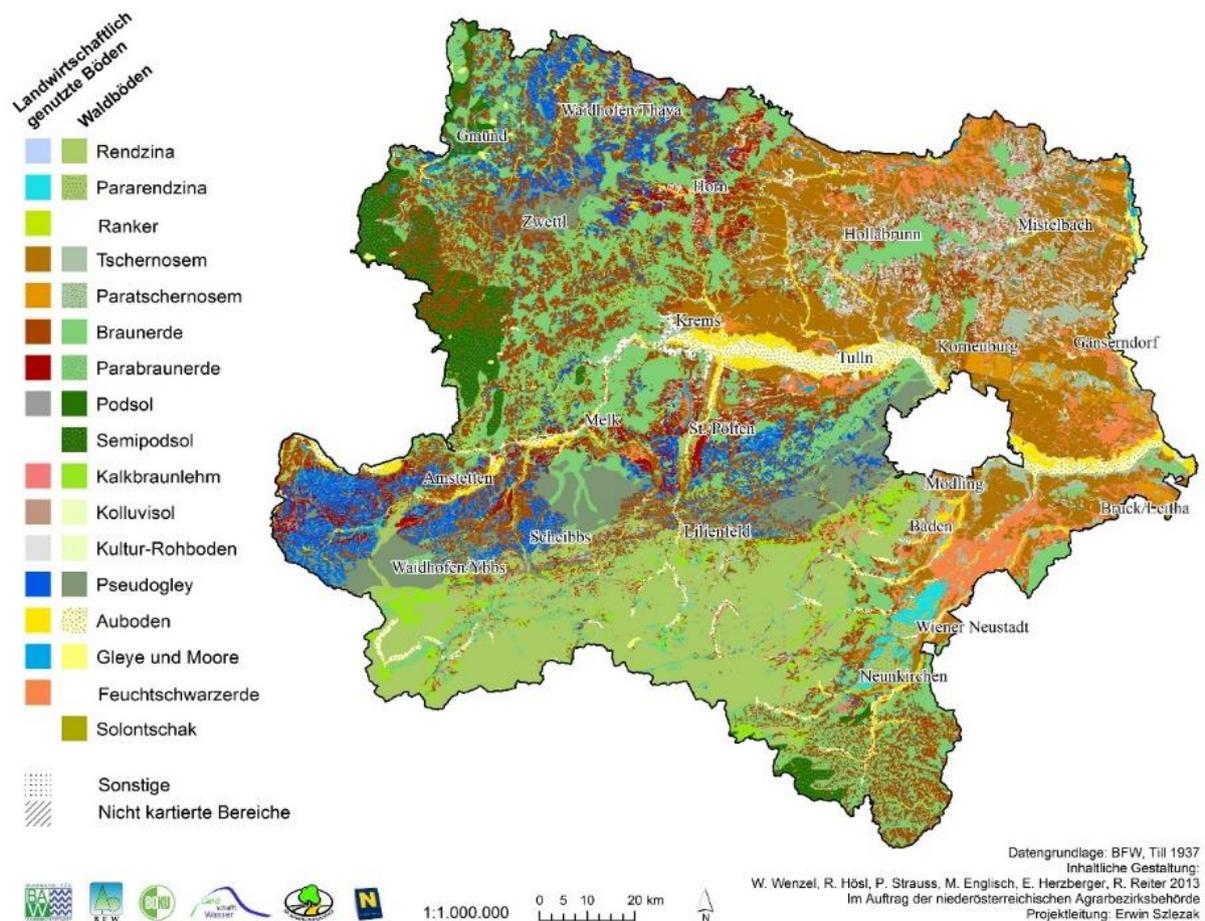
4

Figure 2-4 Soil map of Lower Austria

2.1.2 **Software**

Different software was used for data preparation and prediction. The data preparation of covariates was performed with QGIS 3.4, ArcGIS Map 10.6.1 and with R 3.6.1 in R-Studio. The prediction models were executed with R in R-Studio.

2.1.3 **Data**

To predict the different soil parameters SSOC, Cdef and Csat several data types are needed: first, point data with laboratory sample information and coordinates; second, different environmental spatial data sets covering the survey area; third, a boundary mask to enclose the survey area and last a mask defining the insides of the studies boundary, in this case defining arable land and grassland.

2.1.3.1 **Point Data**

The point data's source is the soil inventory of Lower Austria from 1991/92 (Niederösterreich, 1994). Sampling locations had been arranged in a regular, 3.9 x 3.9 km grid system. Out of 1730 sampling locations in agricultural land (i.e., arable and grassland), only 1449 had been realised, because parts of the grid knots were located in areas which were inaccessible or covered (sealed) by infrastructure. In this study, only samples of a depth of $0 - 20$ cm were considered, resulting in 1300 sample points in total.

In arable land (576 locations), soil samples had been collected from the depth increments $0-20$, $20-40$ and $40-50$ cm. In the centres of each grid cell (consisting of four regular nods) topsoils had been sampled from $0-20$ cm depth, resulting in 575 additional locations in arable land, and 149 samples in grassland areas. At the regular grid locations, samples had been pooled for each depth increment from four subsamples collected in open soil pits located within a circle area of 314 m^2 around the grid not. Sampling at additional locations had been pooled from 20 subsamples collected with a Pürckhauer corer. Soil sampling took place in the period from May 1990 till May 1992.

Samples had been air-dried and passed through a 2 mm screen. The fine earth fraction (< 2 mm) had been analysed for the following characteristics:

- textural classes sand ($2000-63$ µm), silt ($63-2$ µm) and clay (< 2 mm) using a combined sieving and sedimentation method after dispersion with sodium pyrophosphate, and – at humus contents > 50 g kg^{-1} – pre-treatment with H_2O_2 (ÖNORM L 1061);
- soil organic matter (SOM) using wet oxidation with $K_2Cr_2O_7$ solution and concentrated sulfuric acid and subsequent colorimetric measurement of the resulting Cr(III) (ÖNORM L 1081).

Soil sampling and analysis had been conducted by the former Bundesanstalt für Bodenwirtschaft, Vienna.

2.1.3.2 Environmental spatial data

For spatial prediction, different environmental variables are needed. These should be assigned to the S.C.O.R.P.A.N. factors for digital soil mapping (DSM) (McBratney et al., 2003) which are extended soil forming factors (Jenny, 1941). SOC relation to the environment is a broadly discussed field of interest (Lal, 2016b; Ramesh et al., 2019; Wiesmeier et al., 2019).

These factors are, for example, topography (Cardinael et al., 2017), climate (Muñoz- Rojas et al., 2017), soil type (Zhao et al., 2006), soil sampling depth (Li et al., 2017) mineralogical composition (Dwivedi et al., 2017), soil biota (Komarov et al., 2017), land use and management practices (Li et al., 2017;Wang et al., 2017a,b). Wiesmeier et al. (2019) summarized these drivers and indicators specially for SOC storage in context of prediction purposes.

Based on the scientific work described in the upper section and on the data availability the following covariates were selected and used (Table 1).

Table 1: Environmental variables used in the prediction of SSOC, Csat and Cdef

Variables	Definition and formula	Value	Resolution	Source
Topography				
Digital Elevation Model (DEM)	Elevation = Height above sea level	m meter	10x10m	Land Niederösterreich (https://www.data.gv.at/katalog/dataset/land-noe-digitales-hohenmodel-10-m)
Aspect	Inclination direction of the maximum rate of change of the value of one cell compared to its neighbouring cells	° degree to north	10x10m	Calculated from DEM
Slope	rate of change of elevation for each DEM cell. 1. Derivative of the DEM.	%	10x10m	Calculated from DEM
Curvature	2. Derivative of the DEM. 4.- Order polynomial of the 3 x 3 window around the cell. $$Z = Ax^2y^2 + Bx^2y + Cxy^2 + Dx^2 + Ey^2 + Fxy + Gx + Hy + I$$	dimensionless	10x10m	Calculated from DEM
Roughness	Rate of irregularity of the surface. The dispersion of vectors normal to surface areas (pixels). Normal vectors are defined by slope and aspect.	%	10x10m	Calculated from DEM
Topographic Wetness Index (TWI)	$$TWI = \ln\left(\frac{up-slope\ contributing\ area}{\tan(slope\ angle)}\right)$$ hydrological flow paths /potential runoff generation	dimensionless	10x10m	Calculated from DEM

Soil Variables (eBod mask only)				
Texture Classes	Soil texture classes	Category	Polygon	Austrian Digital Soil Map, Bfw (2019)
Sand	% of total mineral soil	%	Polygon	Austrian Digital Soil Map, Bfw (2019)
Silt	% of total mineral soil	%	Polygon	Austrian Digital Soil Map, Bfw (2019)
Clay	% of total mineral soil	%	Polygon	Austrian Digital Soil Map, Bfw (2019)
Soil Type	Austrian Soil Type System	Category	Polygon	Austrian Digital Soil Map, Bfw (2019)
pH	$-\log_{10} a\,(H^+)$	0 – 14	Polygon	Austrian Digital Soil Map, Bfw (2019)
Lime	% of soil	%	Polygon	Austrian Digital Soil Map, Bfw (2019)
SOM Balance	In relative categories from 1 to 6	Category	Polygon	Austrian Digital Soil Map, Bfw (2019)
SOM Value	In relative categories from 0 to 5	Category	Polygon	Austrian Digital Soil Map, Bfw (2019)
Permeability	In relative categories from 0 to 9	Category	Polygon	Austrian Digital Soil Map, Bfw (2019)
Bedrock	Named Bedrock Types	Category	Polygon	Austrian Digital Soil Map, Bfw (2019)
Field capacity	In relative categories from 1 to 4	Category	Polygon	Austrian Digital Soil Map, Bfw (2019)
Climate				
Precipitation	Mean yearly precipitation 1971 - 2000	mm	300x200 m	Austrian Climate Data, Hiebl et al. (2011)
Radiation	Mean yearly precipitation 1971 - 2000	kWh/m²	250x250 m	Austrian Climate Data, Hiebl et al. (2011)
Temperature	Mean yearly precipitation 1971 - 2000	°C	300x200 m	Austrian Climate Data, Hiebl et al. (2011)
Vegetation/anthropogenic factors				
Corine Land Cover 1990 (CLC)	Nomenclature with 27 classes	Category	Polygon	Umwelt Bundesamt (https://www.data.gv.at/katalog/dataset/9e60aeed-ddfa-4be8-b647-b67ab96880ff=)
Geology	Geological Map of Austria	Category	Polygon	Metallogenetischen Karte von Österreich 1:500.000" (GBA, L. Weber 1997).

2.1.3.3 Masks and Boundaries

The study area is confined by the Lower Austria's administrative boarders within a scale of 1:1000 (Land Niederösterreich, 2019). The area for prediction is determined through land use of grassland and agricultural land. Depending on the used covariate layers, two different masks were applied. One was the "INVEKOS Feldstücke" (AMA - AgrarMarkt Austria, 2019) data set. The other one was created from the land use layer of the digital soil map of Austria (Bfw, 2019). The second mask needed to use e-Bod data as covariates because of the different spatial expanses and missing areas compared to the INVEKOS boundaries. In contrast has the land use layer of the eBod data set has disadvantages, as not being up to date and not using single polygons to differentiate every field from each other. Both land use datasets had to be processed before use. Therefore, the different types of land use were generalised into grassland and agricultural land. Categories, not fitting to ones mentioned before were deleted. The R-Code for this can be found in the Appendix.

2.2 Data pre-processing

For the prediction, data pre-processing is needed. The pre-processing directly influences the quality and accuracy of the prediction.

2.2.1 Covariate pre-processing

Covariates are the main elements for successful DSM. To use different spatial layers from different sources, pre-processing of each COV-layer is needed in order to use them for prediction. This pre-processing is divided in several tasks. These are: selecting, acquiring and masking to the same extent, reprojection, rasterizing to one conform grid size and for categorical COVs, creation of Dummy Variables. This was done twice, creating a separate stack for each mask. Therefore R-Studio, Q-GIS and ArcGIS were used for data processing, depending on the program's performance for the task. The Dummy Variable creation is necessary for COVs with non-numerical categorical data, so they can be used with the different machine learning algorithms. A single raster layer for each class was created indicating the presence of the respective class type at each cell location.

2.2.2 Point Data spatial processing

The correct spatial positioning of the BZI datapoints was checked. Three points had wrong coordinates and were located outside of lower Austria. These points are erased from the dataset. After that the sample point location was compared with the two different masks. For each mask a set of points was created. Each set of points was then compared to the corresponding mask. More than 98% of the non-matching points, where less than 20 meters off. Therefore, points, which weren't inside the mask were relocated to the closest raster cell, but only if in a radius of 35 m. Considering the inaccuracies of GPS and the 10 m sampling radius, only points in a radius of 35 m were relocated. The R-Code can be found in the appendix at section "Point Calculation". Each set of points was then split into a training and a validation set. 75 % of the points were used for training, 25 % for testing/validation of the prediction. This operation was conducted with R-Studio via a self-written code.

In the following you can see the comparison of point SOC distribution for the training and validation sets. Figure 2-5 shows the distribution for the eBOD mask. Figure 2-6 shows the

9

distribution for the Invekos Mask. Each divided data set had nearly the same SOC distribution with only minor deviations. This approves both to be used use as train and test data.
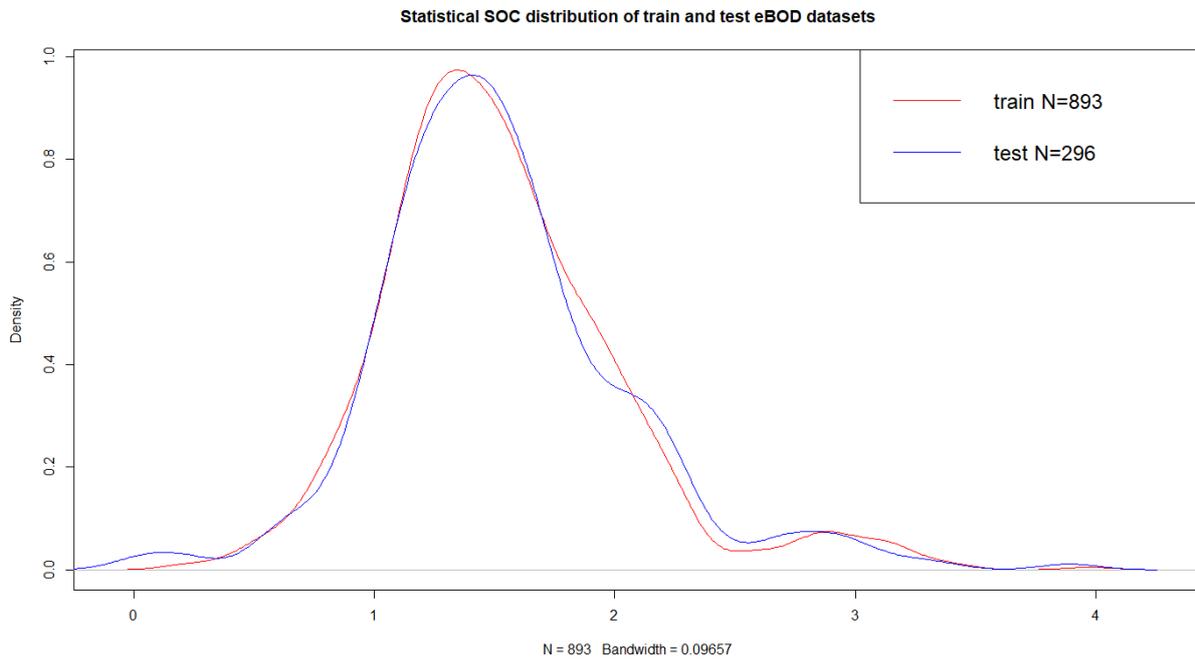

Figure 2-5 Statistical SOC Value distribution of the eBod datasets training and testing split
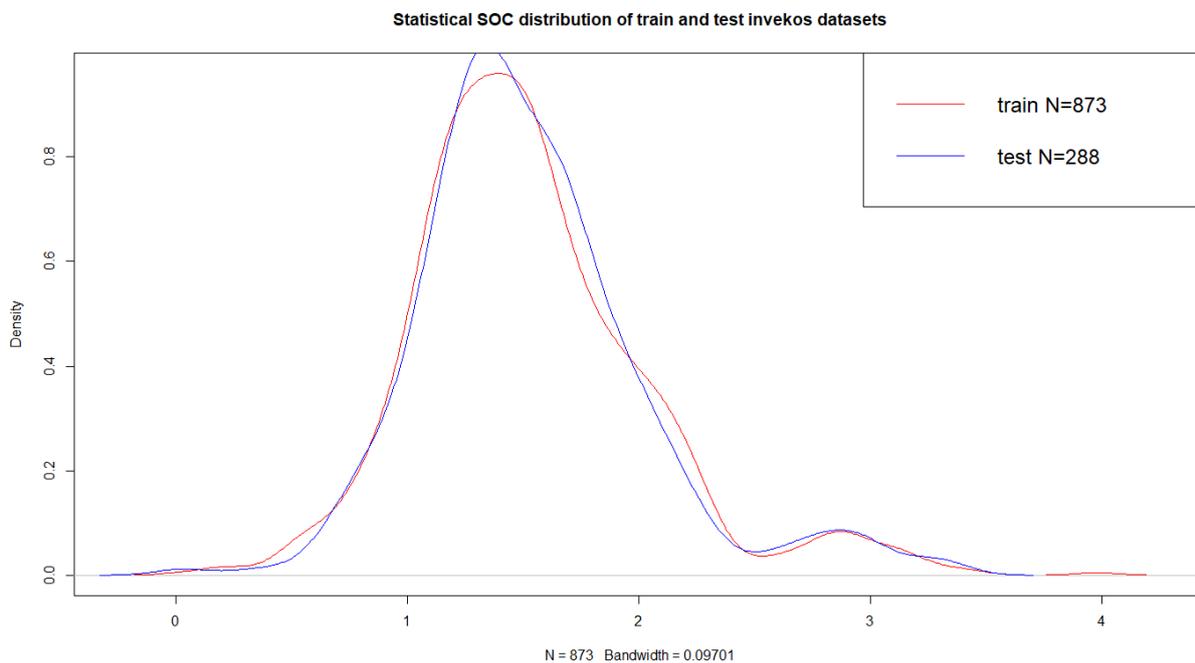

Figure 2-6 Statistical SOC Value distribution of the Invekos datasets training and testing split

2.2.3 **Calculation of Derived Soil Variables**

Before a prediction could be made, continuative soil variables of the original sample points had to be calculated. In the following the different calculations are explained. The calculation code can be found in the Appendix.

2.2.3.1 Calculation of SOC

The initial values of SOC for estimation are derived by dividing the measured soil organic matter (SOM) by the widely accepted Van Bemmelen factor of 1.724. To account for the differences between organic carbon measured by wet oxidation and combustion-based total elemental analysis, the data was corrected using a factor of 1.2 determined for Austrian soils (Gerzabek et al., 2005), yielding the SOC data as further used in this study.

2.2.3.2 Calculation of stable soil organic carbon (SSOC)

Stable soil organic carbon (SSOC) was calculated by multiplying SOC with the factor 0.85 (Angers et al., 2011) to correct for organic carbon not bound to the fine textural fraction $f < 20\,\mu m$ (i.e., particulate organic matter). This factor is based on the analysis of French topsoils and corresponds well with SSOC proportions in other European soils and data from literature reviews (Angers et al., 2011).

2.2.3.3 Calculation of the particle fraction < 20 μm (f<20μm)

Predicting Csat from soil textural data according to Hassink et al. (1997) or Feng et al. (2013) requires information on the fraction <20 μm ($f < 20\,\mu m$). As the Lower Austrian soil database only provides information on the clay ($f < 2\,\mu m$), silt ($f\,2-63\,\mu m$) and sand ($f63-2000\,\mu m$) fractions, Wenzel (unpublished) used published data from various other sources (Alge, 1993; Bröcker and Nestroy, 1995; Jandl, 1987; Katzensteiner et al., 2001; Mentler et al., 2001; Nelhiebel et al., 2001; Nestroy et al., 2001; Rampazzo et al., 2001; Schneider et al., 2001; Strauss et al., 2001) to establish a regression equation predicting $f < 2-20\,\mu m$ (g 100 g^{-1}) from measured $f < 63\,\mu m$ (g 100 g^{-1}):

$$f < 20\,\mu m \ = \ 0.3171\,(\pm 1.1547) * f < 63\,\mu m^{1.1647\,(\pm 0.0358)}$$

$$(n{=}258;\ R^2 = 0.8053;\ RMSE = 1.3491)$$

Equation 1 (Wenzel, unpublished)

The fraction f<63μm was calculated as the sum of the measured clay ($f < 2\,\mu m$) and silt ($f\,2-63\,\mu m$) fractions. The database for predicting $f < 20\,\mu m$ covers soils of different land use (arable, grassland, forest) and various major WRB soil groups (reference), including Leptosols, Fluvisols, Gleysols, Chernozems, Phaeozems, Podsols, Stagnosols, Umbrisols, Cambisols, Regosols. Most data originate from soils collected in Lower Austria or neighbouring provinces. It includes data from different soil horizons as preliminary calculations showed no relevant difference between data subsets of different soil depth.

2.2.3.4 Calculation of Csat

Wenzel (unpublished) employed three different regression models to calculate Csat from the particle size fraction $< 20\,\mu m$ ($f < 20\,\mu m$): Hassink´s least squares linear regression, the upper boundary line equation of Feng et al. (2013) using their parameters obtained for 2:1 mineral soils, and an upper boundary line equation calculated from his own grassland topsoil $(0-20\,cm)$ data. The regression line obtained with his data (Csat = 1.227(±0.0625) $f < 20\,\mu m$; $R^2{=}0.9872$; $R^2adj{=}0.7872$; RMSE=6.7003, p<0.005; n=6) indicates a considerably higher saturation potential than obtained with the parameters for 2:1 minerals reported by Feng et al.

(2013) , and those of Hassink et al. (1997). Figure 2-7 shows a plot of SSOC against f < 20 µm along with the carbon saturation potentials as calculated via the three models.
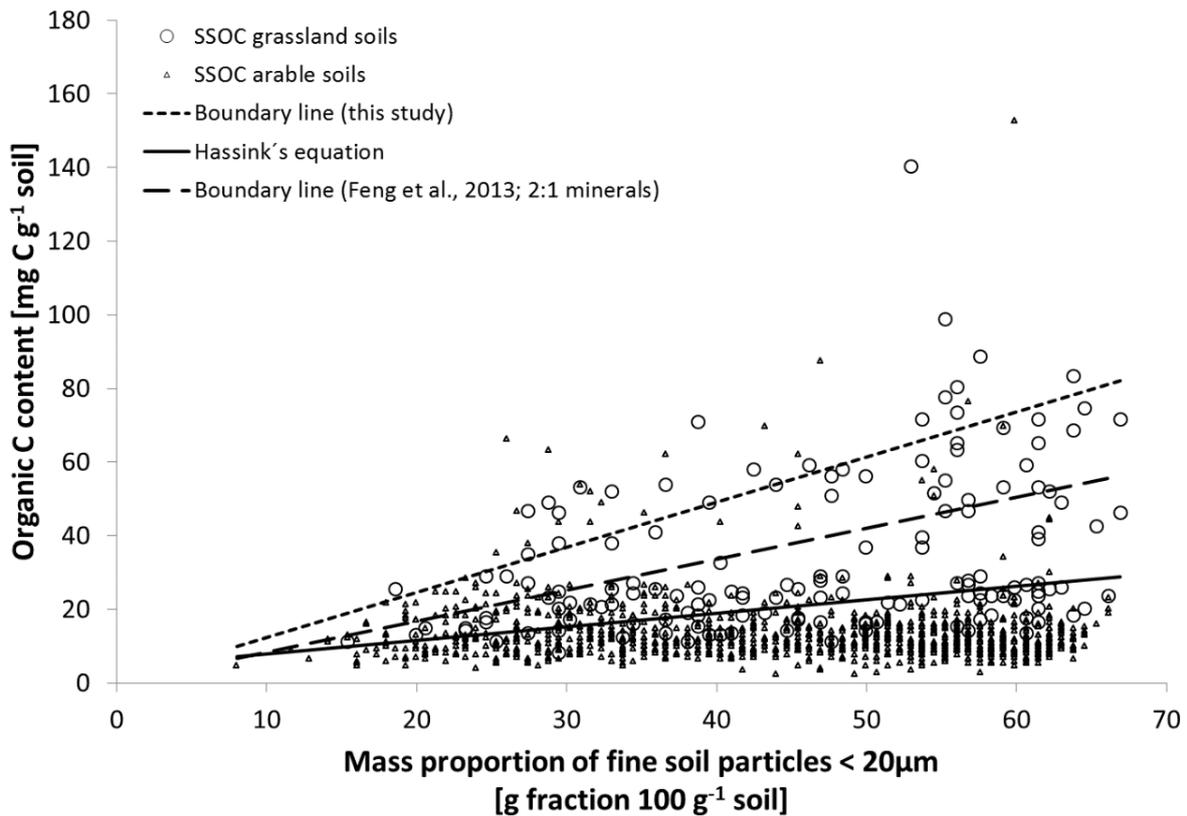

Figure 2-7 Csat regression line comparison (Wenzel, unpublished)

2.2.3.5 Calculation of Cdef

At last Cdef was calculated as the difference between Csat and SSOC values for each individual sampling point (Angers et al., 2011)

2.3 Mapping techniques

To spatially estimate SSOC, Csat and Cdef, several geostatistical methods are possible. In the scientific field of Pedometric mapping a huge variety of methods are used – like Support Vector Machines, Random Forest, Generalized Additive Models, Boosted Regression Trees, Artificial Neural Networks and more (Bhunia et al., 2018; Dai et al., 2014; Liu et al., 2015; Ma et al., 2019; Nayak et al., 2019; Padarian et al., 2019; Pouladi et al., 2019; Taghizadeh-Mehrjardi et al., 2020).

In this study Stepwise Regression Kriging, the Random Forest Regression Tree Method and Support Vector Machines were tested, which in general have good performances (Pouladi et al., 2019). They are also proposed in the Soil Organic Cookbook (Yigini et al., 2018) and two of them were recently used to make a high resolution Csat map of France (Chen et al., 2018). In addition, they are relatively easy to handle with R and they have the same requirements for environmental spatial layers. These are based on the McBratney et al. (2003) S.C.O.R.P.A.N. Principle for Digital soil Mapping, which is described by the soil environmental spatial relationship and its correlation. The original principle was developed by Jenny (1941) as the soil forming equation.

12

The main goal was to predict three different soil characteristics for Lower Austria. Therefore, the best set of covariates and the best prediction method should be used. To narrow down the amount of work, a pre-test was conducted to choose one prediction method. It was tested if there is a significant difference in the results by adding COVs from the eBOD, because of its own generalized soil description and consequently close relation to soil variables. Additionally, the impact on the results accuracy by factor COVs needing a conversion into dummy variables was determined.

In the following the three methods which were tested are described.

2.3.1 Stepwise Regression Kriging

Stepwise Regression Kriging is a hybrid method combining a multiple regression model with kriging of the prediction residuals. The relationship between the multiple COVs and the target variable to be predicted is modelled by the multiple regression equation. This equation is then used to predict the unknowns of the target variable via kriging the residuals (Minasny et al., 2017; Yigini et al., 2018). Kriging is a geostatistical interpolation method with the help of kriging weights. These are dependent on the applied kriging method (Wackernagel, 1995). Stepwise kriging means that the COVs regression model is refitted by removing or adding COVS and testing the model performance via F-test or T-test. Later collinearity among the models' COVs is tested using variance inflation factors. Failing COVs are excluded from the model. The fitted model, in our case, the regression equation, predicts the dependent variable using a linear function of the independent variables (Fig 2-8). The predictor consists of the values of the different COVs. The coefficient is calculated with multiple regression, reflecting the importance of each COV for the dependent variable.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

$Y = Dependent\ Variable;\ \beta_n = Coefficients;\ X_n = Predictors;\ \varepsilon = Residuals$

Equation 2 (Yigini et al., 2018)

Before using the model for kriging, the following assumptions had to be checked to improve the predictions accuracy: normality of the variable of interest; test COV for collinearity; delete one of the high correlating COVs.

2.3.2 Random Forest

Another tested method was Random Forest. It is a machine learning method based on many decision-tree models, which together assemble a "forest" (Breiman, 1998). DSM is an often used method to predict dependent variables (Chen et al., 2018; Pahlavan Rad et al., 2014; Poggio et al., 2013; Wiesmeier et al., 2011), like SOC through COVs. RF splits the main data set into uniform classes based on their distribution and variability. This is called Bagging, which is a bootstrapping technique. (Breiman, 2001). Each tree represents randomly selected sub-samples from the data. The number of COVs used for each tree (m_{try}) and the number of trees in the forest (n_{tree}) varies for each data set and is internally re-calculated for each prediction based on error stabilisation. Through out-of-bag cross validation the relative importance of each COVs can be determined internally. This gives the advantage of a relative high robustness against outliers and overfitting through averaging over large number of trees and yields estimates about the variable importance of the different predictors (Breiman, 2001). The

13

importance of the prediction factors is then used to give statistical weights to the different trees. Thereon the final prediction result is the average over all aggregated trees.

2.3.3 **Support Vector Machines**

SVM is a machine learning technique which maps input vectors, in the given case the COVs, in a non-linear relation through a high dimension feature space. Therefore it uses the help of kernel functions (Cortes and Vapnik, 1995). It is widely used to solve classification problems but also suitable to solve regression problems (Guevara et al., 2018). In the case at hand the linear model is set into a high dimensional feature space. It creates n-hyperplanes through n-dimensional spectral space, separating numerical data based on a kernel function. The planes shape is defined by the SVM parameters gamma, cost and epsilon, to divide the data and to set tolerances. Epsilon defines the insensitive-loss function. Higher epsilon values mean that larger errors while tuning the model, are not penalized. Cost regularizes the optimization model with the help of constrains of the Lagrange formulation. Gamma is the parameter of a Gaussian Kernel to handle non-linear classification. The best model is selected with the help of cross-validation of the mean squared error. Therefore, the best combination of cost and epsilon is selected. The support vector falls within each hyperplane, where the linear model is fitted to the vectors. Later this is used to predict the missing values (Yigini et al., 2018).

2.4 **Validation**

For validation the previously split test-dataset was used to calculate the error of the predictions at the test dataset location. Therefore, the predicted values of each map were extracted at the test-data point locations and then subtracted from the measured test-data values. Proceeding five map quality measures were calculated for each prediction: the mean error (ME), mean average error (MAE), mean squared error (MSA), root mean squared error (RMSE) and the amount of variance explained (R^2 or AVE). In addition, scatterplots were plotted for the three prediction methods. For the final prediction of all three soil parameters scatterplots were plotted as well as error bubble maps to visualize the spatial distribution of the prediction errors. Additionally, boxplots for comparison of the predictions values related to land use were created. This was accomplished in Q-GIS with the help of the zonal statistics function. For each land use polygon of the eBOD data the mean of enclosed raster cells of the prediction was calculated. For the final prediction the variable importance of used COVs is calculated. It indicates how much the removal of each unique COV reduces the model's accuracy and expresses the importance of each COV in the predictive model.
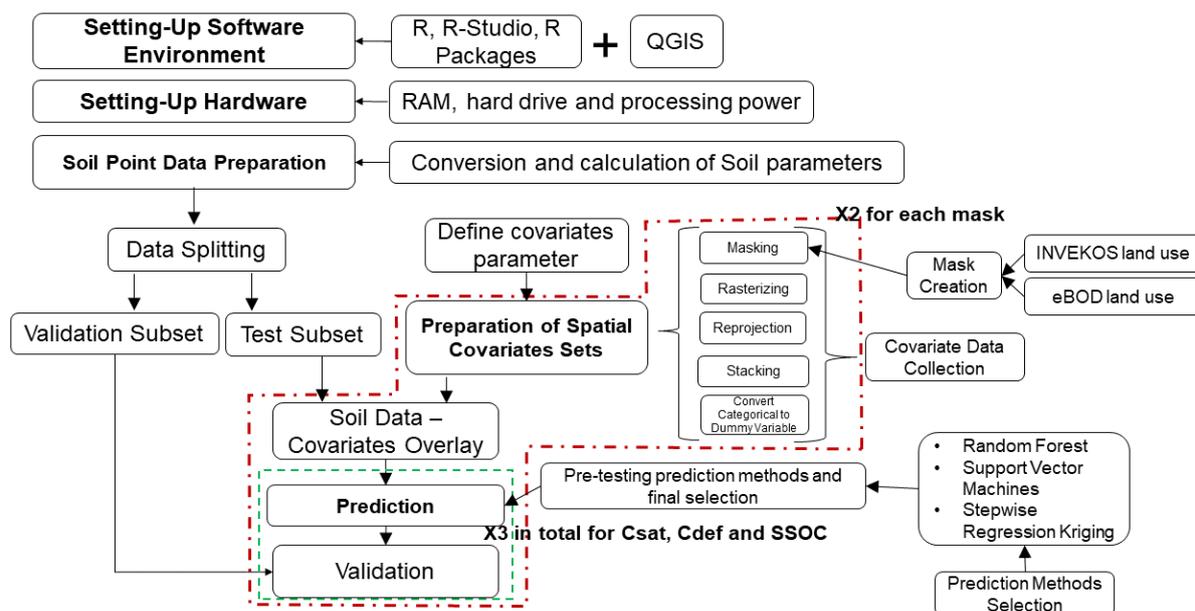
2.5 Workflow

Figure 2-8 Workflow Diagram of this studies prediction

In Figure 2-8 the order of the necessary tasks for this studies prediction can be seen.

3 Results

3.1 Prediction Method Selection

Digital Soil Mapping offers several different methods to predict soil characteristics. To narrow down available options pre-selected methods from the Soil Organic Cookbook were tested and compared (Yigini et al., 2018).

Used Covariates are: Digital Elevation Modell, Precipitation, Radiation, Temperature, Aspect, Slope, Curvature, Roughness and Topographic Wetness Index (TWI).

To validate the results, the prediction error (Table 2) was calculated with the help of previously separated test points.

Table 2 Validation of the prediction error (PE) for Regression Kriging (RK), Random Forest (RF) and Support Vector Machine (SVM)

	PE – RK	PE – RF	PE – SVM
Min.	-18.9435	-18.9608	-18.69129
1st Quantile	-0.7577	-0.8393	-0.93415
Median	0.1056	0.1249	0.05542
Mean	-0.3184	-0.4285	-0.43345
3rd Quantile.	0.6745	0.8309	0.76071
Max.	16.6593	3.7536	3.76603

Figure 3-1 shows the Regression line with the 1:1 line for each prediction, showing promising results for RF and SVM and more pronounced deviation for the RK prediction.
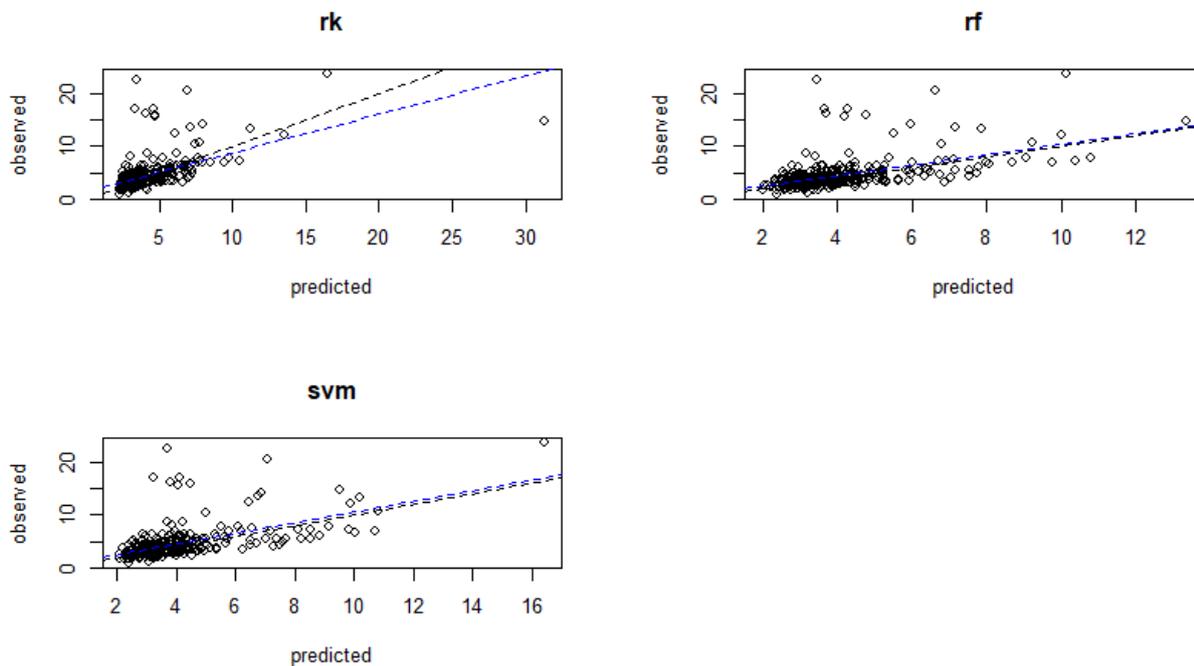

Figure 3-1 Scatterplot of predicted versus observed SSOC content for the three different predictions with Regression Kriging (rk), Random Forest (rf) and Support Vector Machines (svm), blackline represents the 1:1 line of prediction versus observed, the blue line represents the regression between observed and predicted values

Comparing the prediction error statistics of the three prediction methods in Table 2 shows that, regression kriging (RK) has the lowest mean prediction error of -0.3184, followed by random forest with -0.4285 and at last support vector machines with slightly worse results of -0.43345. It seems that RK overestimates in the prediction more than the other two methods, see Table 2 in the row Max prediction error. In Figure 3-1 the Scatterplot of the three predictions is displayed with the 1:1 line and the regression line. Random Forest and Support Vector Machine represents there better the observed values with their prediction than Regression Kriging. Random Forest was chosen over regression kriging due to superior processing speed and lesser pre-prediction preparation. It also performed slightly better than SVM.

3.2 Prediction with RF

Map quality measures are calculated from the prediction error to select the final set of COVs. Table 3 shows the statistics of the different RF predictions. The Soil map (eBod) (e) covariate layers significantly improve the R² of Csat and Cdef prediction compared to the predictions utilizing only the Invekos (i) layers. For example, Cdef_i_ofk_rf 0.475 compared to Cdef_e_ofk_rf 0.613. The R² shows the major improvement of the prediction by using the additional eBod covariates. R² continuously remains above 50 for Csat and Cdef predictions with the eBod layers, as seen in Fig. 3.2. and 3.3. in contrast to the SSOC results in Fig. 3.4. Comparing only the SSOC prediction the eBod covariates provided also a better result compared to the Invekos only predictions. But for the SSOC prediction, the eBod COVs didn't significantly improve the prediction compared to Invekos prediction. Also, the use of categorical COVs improved the prediction for all three variables. Potential problems with the converted Dummy Variables were ruled out via successful testing of predictions with and without categorical COVs. For example, to high cardinality of the variables can cause overfitting and

hence deterioration of the prediction results. These results deviate from the statistics results in 3.3, due to minor changes in the final predictions calculation and changes in the point calculation.

Table 3 Pre-Prediciton testing the different covariate combinations with Random Forest, colours indicating the performance of each error value within each prediction relatively, green = good, red = poor (e = with eBod COVs, i = Invekos COVs only, fk = with

	e	i	fk	ofk	ME	MAE	MSE	RMSE	R²
Cdef	x		x		-0.885	8.350	115.290	10.683	0.627
	x			x	-0.797	8.532	118.980	10.889	0.613
		x	x		-0.645	10.016	157.718	12.559	0.496
		x		x	-0.688	10.415	164.405	12.822	0.475
Csat	x		x		-1.054	8.388	117.979	10.807	0.599
	x			x	-0.925	8.492	121.165	10.989	0.586
		x	x		-0.874	10.433	169.146	13.006	0.453
		x		x	-0.858	10.820	179.065	13.382	0.421
SSOC	x		x		-0.459	1.321	9.605	3.084	0.304
	x			x	-0.452	1.394	10.455	3.228	0.237
		x	x		-0.435	1.392	7.126	2.670	0.302
		x		x	-0.464	1.454	7.791	2.791	0.237

Standard deviation of different mask for prediction

Variations in the different predictions can be found in the following density plots, visualizing the standard deviation. A difference was visible for the Cdef and Csat plots (Fig. 3-2; 3-3), which is based in the usage of eBod or Invekos set of COVs. In the SSOC prediction (Fig. 3-4) the differences are neglectable. The addition of soil map COV made a significant impact for Cdef and Csat.
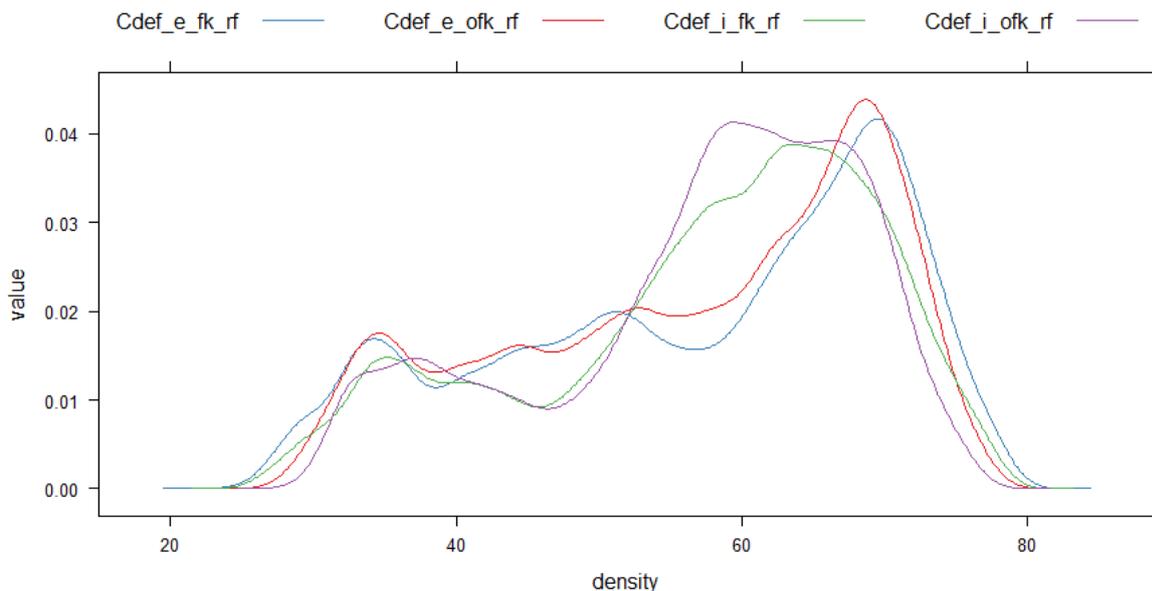

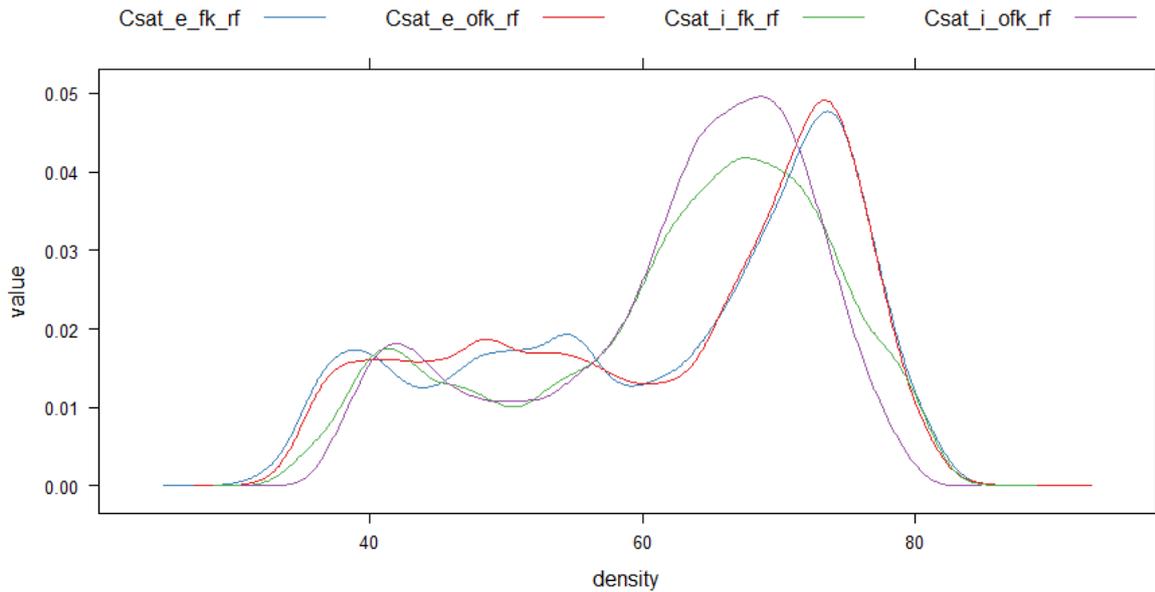**Figure 3-2 Density plot Cdef prediction with RF**

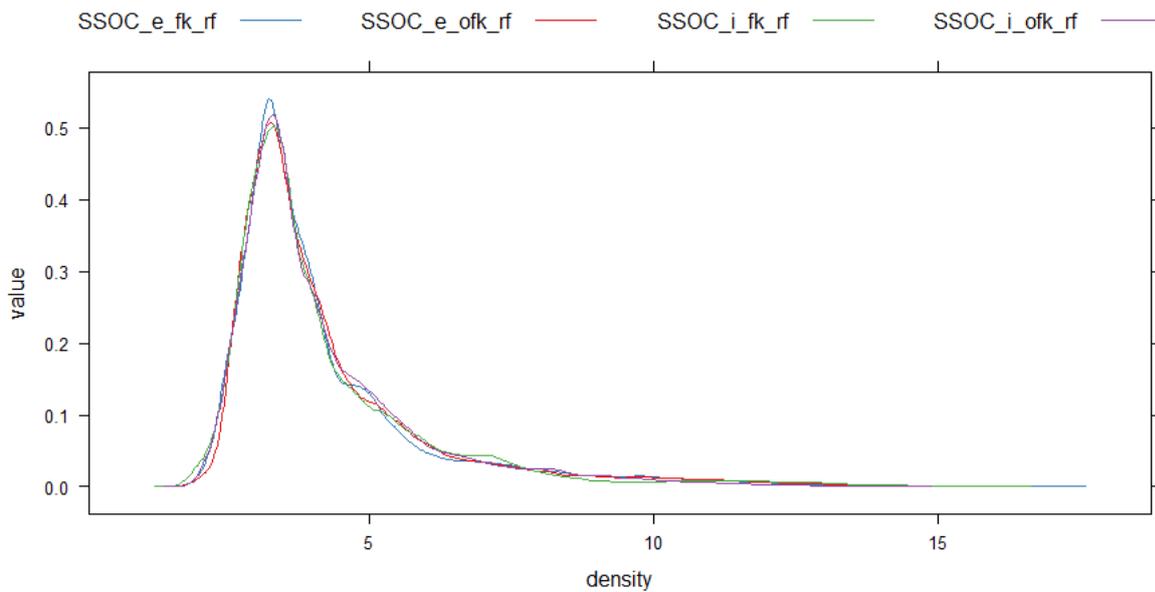**Figure 3-3 Density plot Csat prediction with RF**

Figure 3-4 Density plot SSOC prediction with RF

Out of these predictions the best COVs and mask combination was selected to predict the final SSOC, Csat and Cdef maps with random forest. Therefore, the eBOD mask with all available COVs was selected for the final prediction.

3.3   Final Predictions

Some changes in the point value calculation had been made to correct minor errors. Therefore, the test calculated error statistic deviates from the final results error statistic.

In Table 4 the error statistics of the best performing predictions can be seen. SSOC has a higher ME in comparison to Cdef and Cdef. Cdef has the highest MSE of 131.25, indicating a high scattering of the prediction values. The Csat prediction with an MSE of 48.81 shows the smallest scattering as wells as the best R² value compared to the other two predictions.

Table 4 Error Statistics of the final Cdef, Csat and SSOC prediction (ME = Mean Error, MAE = Mean Absolute Error, MSE = Mean Squared Error, RMSE = Root Means Squared Error, R² = Amount of Variance Explained)

Prediction	ME	MAE	MSE	RMSE	R²
Cdef	-0.115	6.772	131.247	11.398	0.517
Csat	-0.561	5.353	48.808	6.951	0.577
SSOC	-1.500	4.457	108.443	10.361	0.306

The scatterplots in Fig.3-5 show the 1:1 and the regression line for each prediction. Csat and Cdef predictions have only little variances from the 1:1 line. The SSOC prediction shows a higher slope compared to the 1:1 line, which is confirmed by the R² value in table 4. All three plots are showing a reasonable prediction through the models with variability and few outliers. The Csat prediction shows the widest scattering.
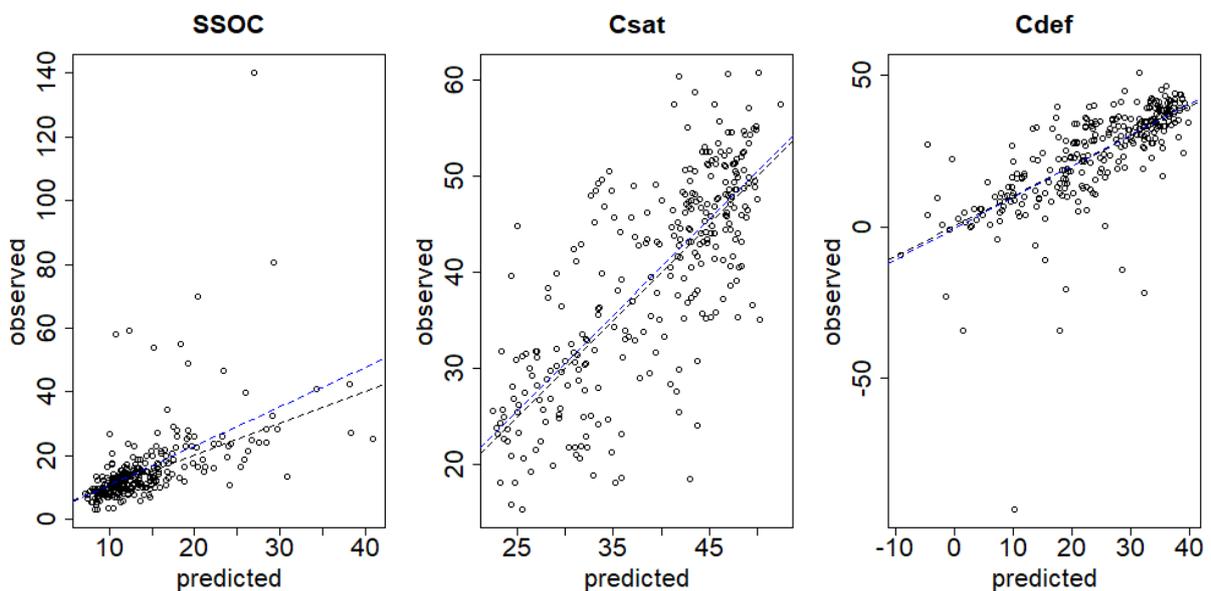

Figure 3-5 Scatterplot for SSOC, Csat and Cdef prediction, blackline represents the 1:1 line of prediction vs. observed, the blue line represents the regression of observed and predicted values

### 3.3.1 SSOC prediction with RF

The SSOC prediction results in Fig.6-4 show high values of $20 - 25$ g/kg and higher mainly in the southern parts of lower Austria and in the west near to the boarder of Upper Austria. Low SSOC values ranging from $< 10 - 12.5$ g/kg can be found in the Weinviertel. Slightly higher values of $15 - 20$ g/kg can be found southeast in the Industrieviertel and in the flood plains near the Danube as well in the Marchfeld east of Vienna and the plains next to Tulln. Also, in the central Waldviertel medium values can be found. In areas with grassland, there are mainly high SSOC values and in areas with arable land low SSOC values are present.

Stable soil organic carbon Map of Lower Austria for Agricultural and Grassland top soils 0-20cm (1990/91)

SSOC g/kg
- <= 10
- 10 - 12.5
- 12.5 - 15
- 15 - 17.5
- 17.5 - 20
- 20 - 22.5
- 22.5 - 25
- > 25

Modell: Random Forest
R² : 0,31
Mean Average Error: 4,46
Root Mean Squared Error: 10,36

Data Source:
Soil Inventory of Lower Austria
1990/91

Authors: F.J.K. Mayr & Walter W.W. Wenzel, 2020 ©

Universität für Bodenkultur Wien
University of Natural Resources
and Life Sciences, Vienna

0   50   100 km
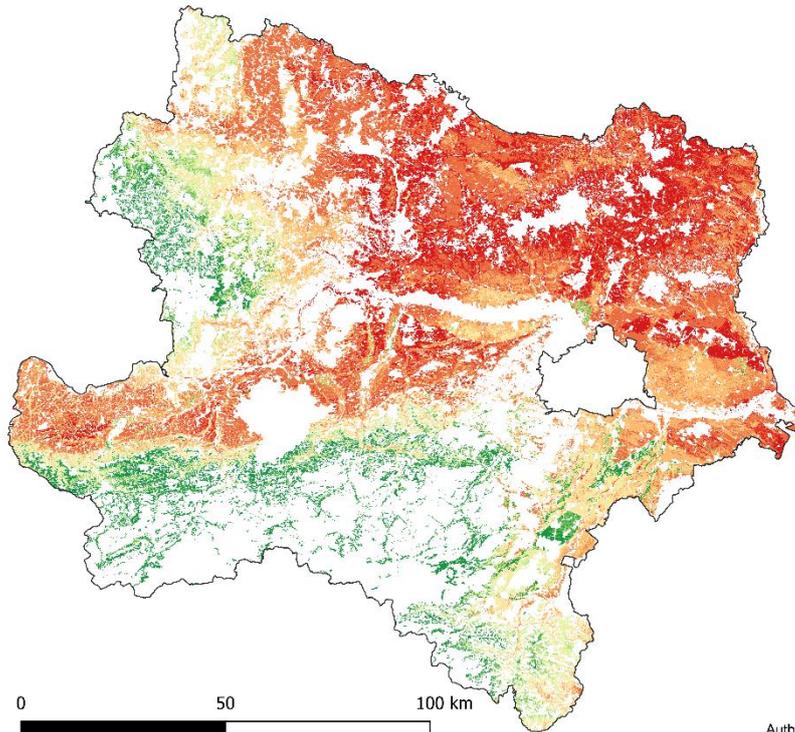
Figure 3-6 SSOC prediction for agricultural and grassland top soils 0-20cm (1990/92)

In Table 5 the 10 most important COVs for the prediction can be seen, with their percentage of importance within the RF prediction. The Digital Elevation Modell and Precipitation COVs have the highest influence on the SSOC prediction. On third place is the number five layer of the Soil Organic Value Layers of the Austrian Soil Map which directly shows an influence on SSOC prediction. The SOM Value Layers ranging from 0 to 5 indicating the relative amount of SOM in the study area. They are in separate layers because they are converted into Dummy Variables. The rest of the remaining COVS have similar influence ranging from $58 - 44\%$.

Table 5 SSOC - Variable Importance top 10 in % of the RF prediction

Covariate	%
Precipitation	100
Digital Elevation Model	76.55
SOM Value 5	75.00
Roughness	58.63
Clay	57.31
Temperature	55.66
SOM Value 4	55.18
Sand	47.86
Lime	44.18
Topographic Wetness Index	44.09

Arable land in Lower Austria has lower mean SSOC contents than grassland (Fig.3-7). The mean for arable land is 14.98 g/kg (SD = 5.1) and for grassland 20.29 g/kg (SD = 8.53). The Boxplot mean calculation description can be found at section 2.4.
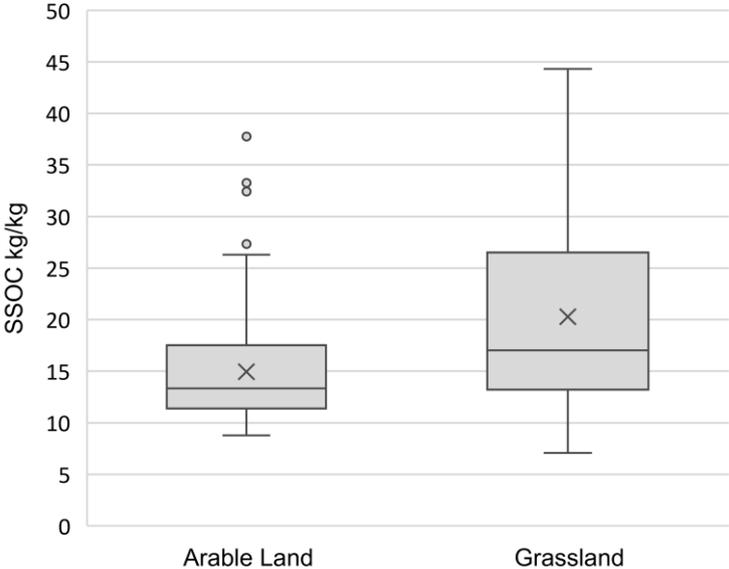

Figure 3-7 Boxplot of the mean predicted SSOC of arable and grassland soils

3.3.2 Csat prediction with RF

The carbon saturation potential map (Fig 3-8) shows high values from 40 g/kg up to over 50 g/kg of possible C saturation in nearly every part of the Weinviertel. In the north-western parts of the Waldviertel low Csat values are visible, ranging from 30 – 25 g/kg and lower. One outstanding area in the Marchfeld also displays low values from < 25 – 30 g/kg. In the southern part of the Industrieviertel low values ranging from 35 – 25 g/kg are observable. In the Alpine foothills of the Mostviertel higher values from 40 – 50 g/kg can be found.
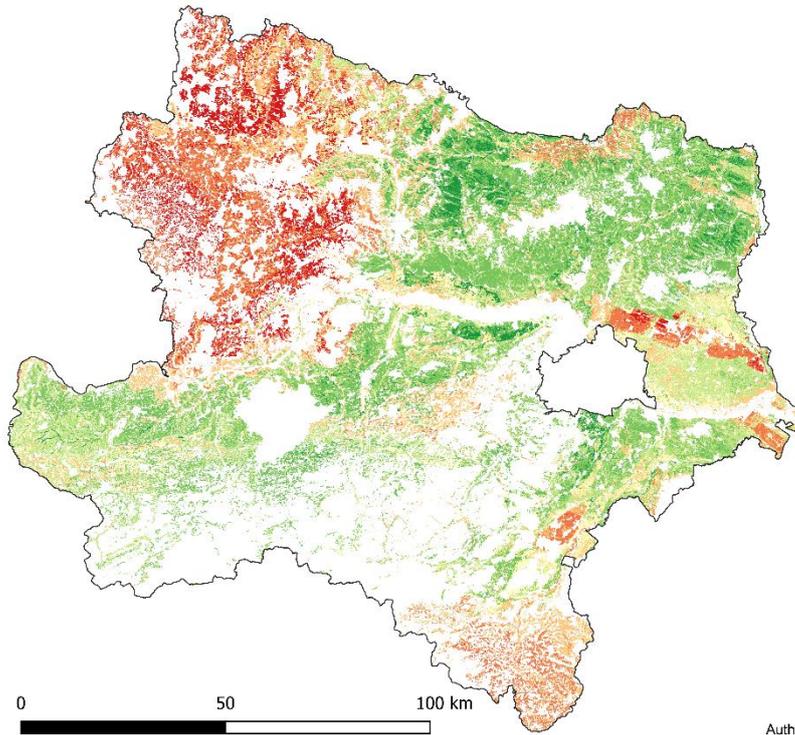
Figure 3-8 Csat prediction for agricultural and grassland top soils 0-20cm (1990/92)

In the following (Table 6) the 10 most important COVs for Csat prediction can be seen. Sand COV is of highest importance with 100%. Followed by the DEM and Clay with ~60%. Temperature and Precipitation also have a significant share in the prediction with 58% and ~56% in variable importance.

Table 6 Csat - Variable Importance top 10 in % of the RF prediction

Covariate	%
Sand	100.00
Digital Elevation Model	60.57
Clay	60.14
Temperature	58.05
pH	56.74
Precipitation	56.65
Silt	45.88
Roughness	42.95
Lime	34.35
Textur Classes 3	33.01

Csat mean values are comparable between arable and grassland, the latter expressing a broader range (Fig.3-9). The mean for arable land is 37.4 g/kg (SD = 6.76) and 38.5 g/kg (SD

= 6.63) for grassland. Soil texture is therefore relatively homogenously distributed over the two different types of land usage. The calculation method can be found in section 2.4
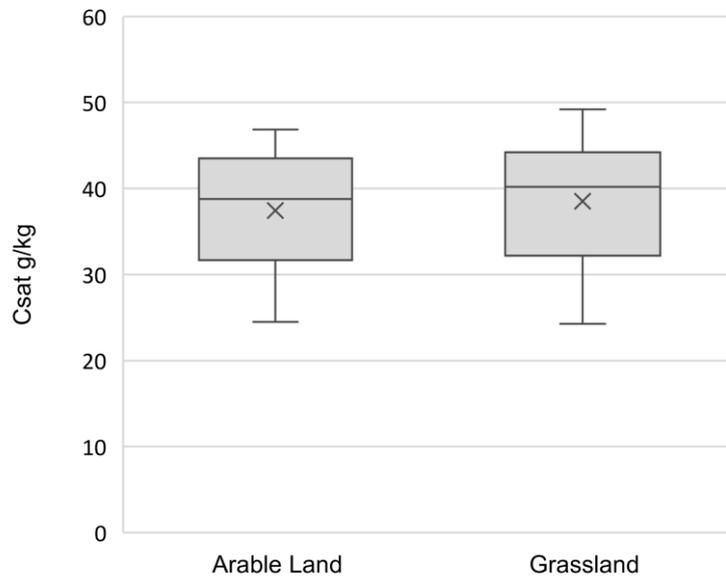
Figure 3-9 Boxplot of the mean predicted Csat of arable and grassland soils

3.3.3 Cdef prediction with RF

The Cdef predictions (Fig. 6-6) indicate the remaining storing potential of SSOC in Lower Austria. The highest deficit in the Weinviertel is located in the western parts with an additional C storage of 30 – 50 g/kg possible. In the northern parts of the Weinviertel lower levels can be found. They are ranging from 20 – 30 g/kg. In the Marchfeld Region in the east of Lower Austria, midrange deficits of 30 – 40 g/kg are visible. There is also a small area with around 20 g/kg in the soil present. In south-eastern parts the deficits are low, ranging from 30 – 20 g/kg. The area in the southwest of Lower Austria has higher values from 35 – 50 g/kg. The highest deficit can be found northwest of Vienna. Low levels of 25 g/kg and less are visible in the northwest area of Lower Austria.
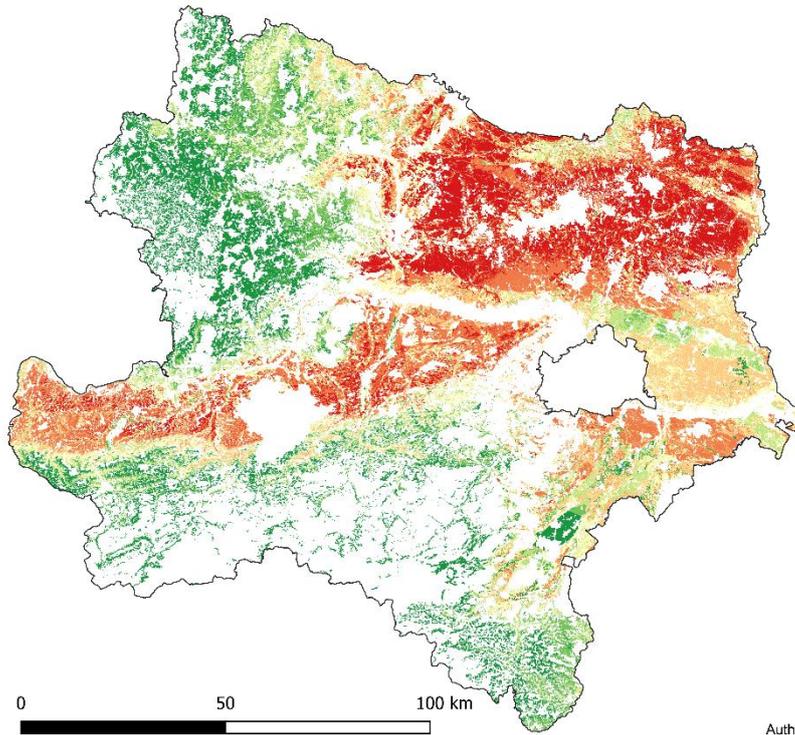
Figure 3-10 Cdef prediction for agricultural and grassland top soils 0-20cm (1990/92)

Table 7 represents the 10 most important COVs for the alongside their respective influence. The DEM has the biggest importance, followed by sand and temperature. Clay and silt are also in the top 10, as well as precipitation and radiation COVs. All three available climate COVs are in the top 10 underlining their importance. With the SOM-Value5, pH and Lime layer three other important COVs of the Austrian Soil map are included. This represents the combination of the SSOC and Csat COVS most important Variable Importance's.

Table 7 Cdef - Variable Importance top 10 in % of the RF prediction

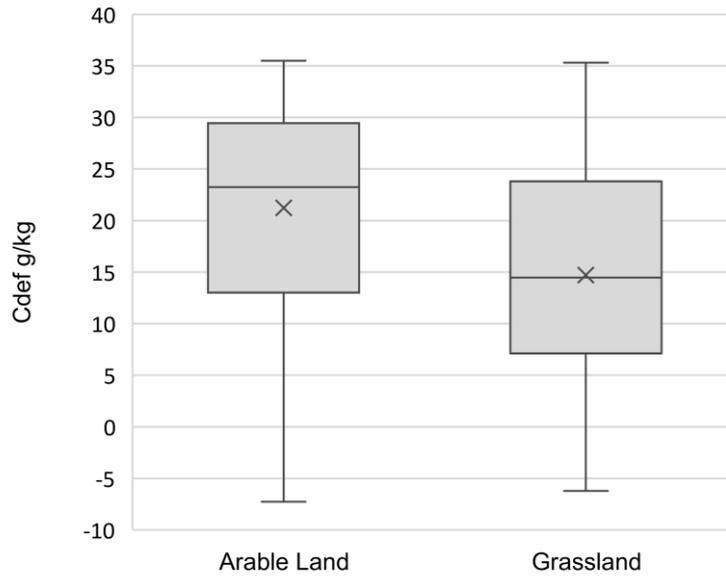| Covariate	%
Digital Elevation Model	100.00
Sand	95.33
Temperature	84.68
Precipitation	64.78
Clay	59.64
SOM Value 5	57.53
Silt	55.92
pH	52.42
Solar Radiation	52.10
Lime	50.16

Figure 3-11 Boxplot of the mean predicted Cdef of arable and grassland soils

Arable land in Lower Austria has higher mean Cdef values, indicating a strong influence of land use (Fig.3-11). The mean for arable land is 21.23 g/kg (SD = 9.80) and for grassland 14.8 g/kg (SD = 10.93). The Cdef means for the Boxplot in Figure 3-11 were calculated with the same method mentioned above.

3.3.4 Statistical Verification of land use influence on SSOC, Csat and Cdef

In the following Table 8 the results of the conducted unpaired Welch two sample T-tests for the soil values SSOC, Csat and Cdef in comparison with land use can be seen.

Table 8 Statistic results of an unpaired Welch two sample t-test for the land use Zonal Statistic mean values of SSOC, Csat and Cdef (Mean = mean of all polygon means, σ = standard deviation, H_0 = Null hypothesis ,H_1 = alternative hypothesis, α, t = Test variable calculated from the means, standard errors and sample size, df = degrees of freedom, p-value = largest probability under H_0, 95% con. I. = 95% confidence interval)

	SSOC		Csat		Cdef	
	arable land	grassland	arable land	grassland	arable land	grassland
Mean	14.98 g/kg	20.29 g/kg	37.4 g/kg	38.5 g/kg	21.23 g/kg	14.8 g/kg
σ	5.1	8.53	6.76	6.63	9.8	10.93
H_0	Mean Arable $_{SSOC}$ = Mean Grassland $_{SSOC}$		Mean Arable $_{Csat}$ = Mean Grassland $_{Csat}$		Mean Arable $_{Cdef}$ = Mean Grassland $_{Cdef}$	
H_1	Mean Arable $_{SSOC}$ ≠ Mean Grassland $_{SSOC}$		Mean Arable $_{Csat}$ ≠ Mean Grassland $_{Csat}$		Mean Arable $_{Cdef}$ ≠ Mean Grassland $_{Cdef}$	
α	0.0001		0.01		0.0001	
t	− 6.0361		−1.3051		4.8849	
df	211.74		243.49		245.81	
p-value	6.98E-06		1.93E-01		1.87E-06	
95% con. I.	−7.084697	−3.596506	−2.7878739	−2.7878739	3.839073	9.026812

The p-value of the SSOC T-test is smaller than α. Therefore, H0 can be rejected. The SSOC mean values for arable and grassland are not equal. H1 can be accepted, which concludes that there is a statistically significant difference between the means of the two land use types.

The p-value of the Csat T-test is bigger than α. Therefore, H0 is not rejected. The Csat mean values for arable and grassland showing no significantly difference between mean values regarding land use.

The p-value of the Cdef T-test is smaller than α. Therefore, H0 can be rejected. The Cdef means for arable and grassland are not equal. H1 can be accepted, which concludes that there is a statistically significant difference between the both means.

3.3.5 Prediction Error Bubble Map

The following figures 3-9, 3-10 and 3-11 show the prediction errors for the three different soil values based on the pre-selected test point. Red represents underestimation, green overestimation. The circle radius corresponds to the size of the Error. SSOC has mainly homogeneous under- and overestimation, spreading over the study area including some underestimating outliers, especially in areas of the Industrieviertel. Csat has homogeneous under- and overestimation. Cdef has some overestimating outliers with high values in the Industrieviertel.
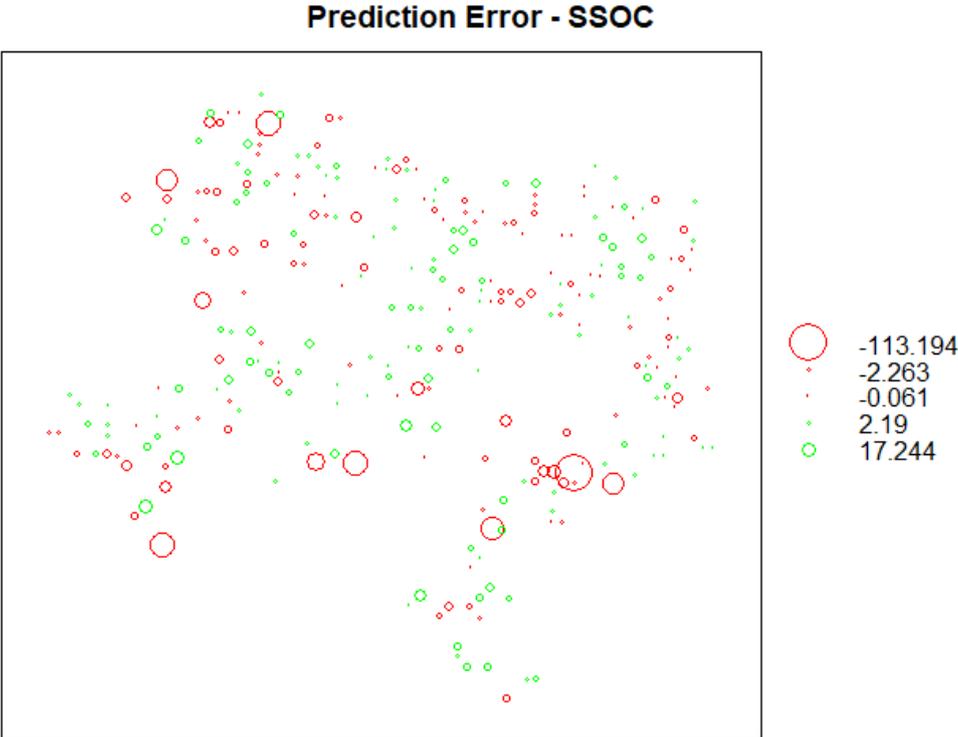
Figure 3-12 Prediction Error Bubble Map for SSOC

Prediction Error - Csat

○ -19.783
○ -4.856
· -0.881
○ 3.35
○ 24.454
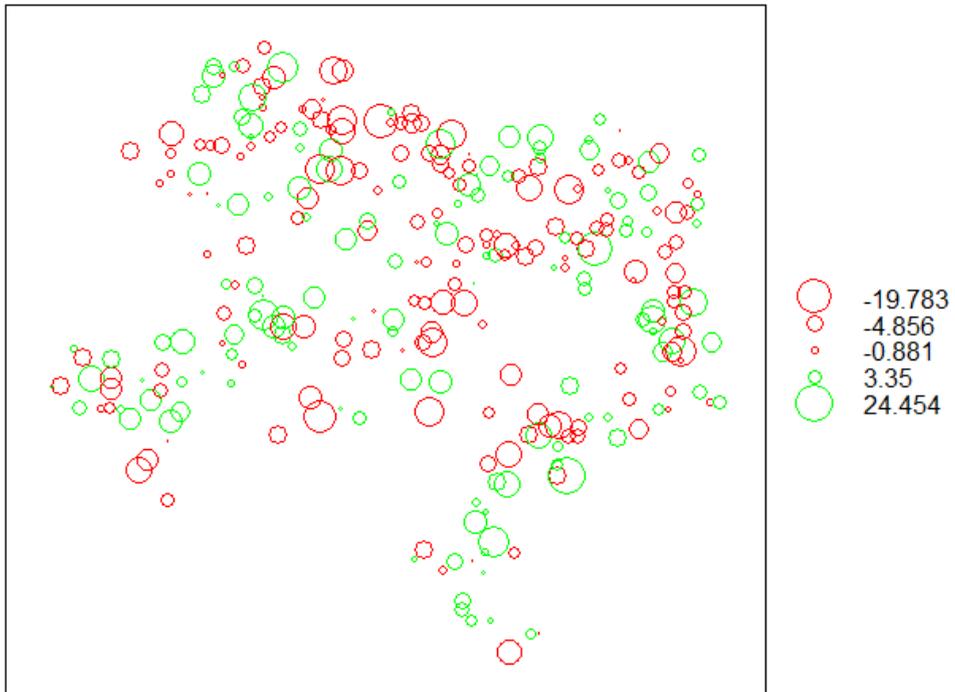
Figure 3-13 Prediction Error Bubble Map for Csat

Prediction Error - Cdef

○ -31.715
○ -5.174
· -1.473
○ 3.035
○ 103.97
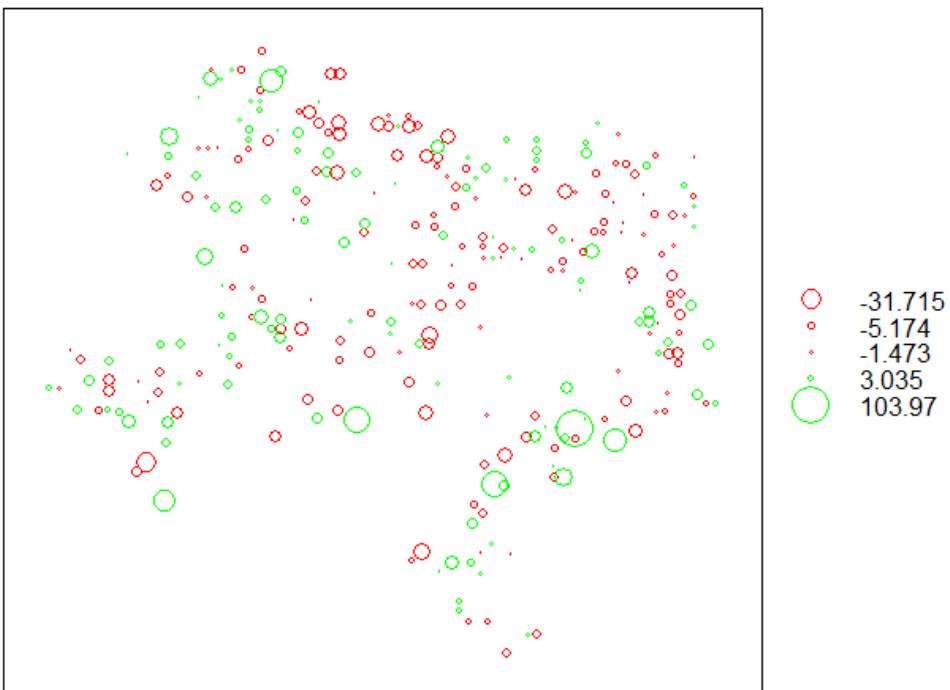
Figure 3-14 Prediction Error Bubble Map for Cdef

27

3.4 Discussion

3.4.1 Interpretation of the final calculated SSOC, Cdef and Csat results

The variable importance of all three predictions shows that the three texture classes and the climate COVs temperature, precipitation and radiation, have a major influence on the prediction. Additionally, the DEM and its derivates, like surface roughness and TWI strongly affect the prediction, validating the widely accepted soil forming factors. Other soil map variables like pH and lime should not be underestimated in their effects, too. A correlation between SSOC, Cdef and land use is demonstrated (Fig.3-7 and 3-11). This can be also seen visually by comparing the SSOC and Cdef prediction with Fig.2-1. It is also approved by the Welch t-test in 3.3.4, also resulting in high Cdef values for arable land and high SSOC values for grassland and verifying the assumption that arable land has in general lower SOC values in comparison to grassland (Martin et al., 2011; Meersmans et al., 2011; Wiesmeier et al., 2012). The boxplot statistic was carried out with a soil map land use layer which wasn't used in the prediction. For the prediction the Corine Land Cover data from 1990 (CLC90) was utilized due to its similar survey period compared to the soil sampling. The CLC90 COVs had an influence (Appendix 8.3), but only marginal. The broadly known and used "Universal Soil Loss Equation" (USLE), developed by the US department of agriculture estimates the average annual soil loss based on different environmental factors (Wischmeier and Smith, 1978). The equations factors are inter alia influenced by rainfall, topography and management of the soil. Therefore soil erosion has an implied influence towards the predicted SSOC and Cdef values (Li et al., 2017). This is reflected in the high Variable Importance of the DEM and its calculated COV-derivates like slope, roughness and TWI in combination with precipitation (see Table 5 and 7). The relatively high importance of the clay COV of SSOC can be related to its direct influence of its stabilisation and protection ability for organic matter (Hassink et al., 1997; Six et al., 2002). In the Csat prediction the COV top ten covariate importance (Table 6) reflects the strong correlation to texture with Sand Clay and Silt COVs, justified due to the Csat calculation from the particle size fraction $< 20\,\mu m$. These were calculated through a regression equation in 2.2.3.3 out of the sum of the silt and clay fraction ($f < 63\,\mu m$). Sand has the highest variable importance for the Csat prediction. It can be assumed that this is due to a negative correlation with the Csat prediction since it is the only texture fraction which is not included in the Csat calculation. In Table 6 the pH and lime layer show high importance for the prediction, these layers are part of the eBOD. This altogether underlines the importance of the Austrian soil map as COVs for prediction, especially regarding the texture layers.

A comparison of the prediction performance of this study with others shows similar results. The final prediction results of this study yields a R^2 of 0.517 for Cdef, 0.577 for Csat and 0.306 for SSOC. Chen et al. (2018) calculated a R^2 of 0.47 for the prediction of the C sequestration potential. Wiesmeier et al. (2011) attained better results with a R^2 of 0.74 for their SOCS prediction. Really good performances were achieved by Taghizadeh-Mehrjardi et al. (2020) with a SOC predictions with combined machine learning methods and advanced COV selection resulting in R^2's from 0.63 up to 0.9 for the best prediction within soil depths up to 30 cm. This shows that with the right data and methods results close to reality are achievable, improving accuracy of future soil maps. It must be considered that for each study different initial conditions and variations of SOC types were present.

3.4.2 SSOC Benefits through related management practices

Benefits of SOC sequestration

In order to understand the motivation behind the performed prediction, background knowledge about the benefits of carbon in soil is essential. Organic carbon (OC) is one of the most important indicators for healthy soil (Lefèvre et al., 2017). It plays a major role in the four soil services which are supporting, regulating, provisioning and cultural services. Lal (2011) summarizes the benefits of high soil quality if SOC-pools in agricultural lands are increased, resulting in co-benefits like improving resilience against climate change, increasing fertility and therefore productivity as well as increasing CO_2 storage capacities (Lal et al., 2015; Smith, 2008). This summary can be seen in Fig.3-15. It shows how physical, chemical and biological quality factors are positively influence SOC-pools in terms of quantity and quality.
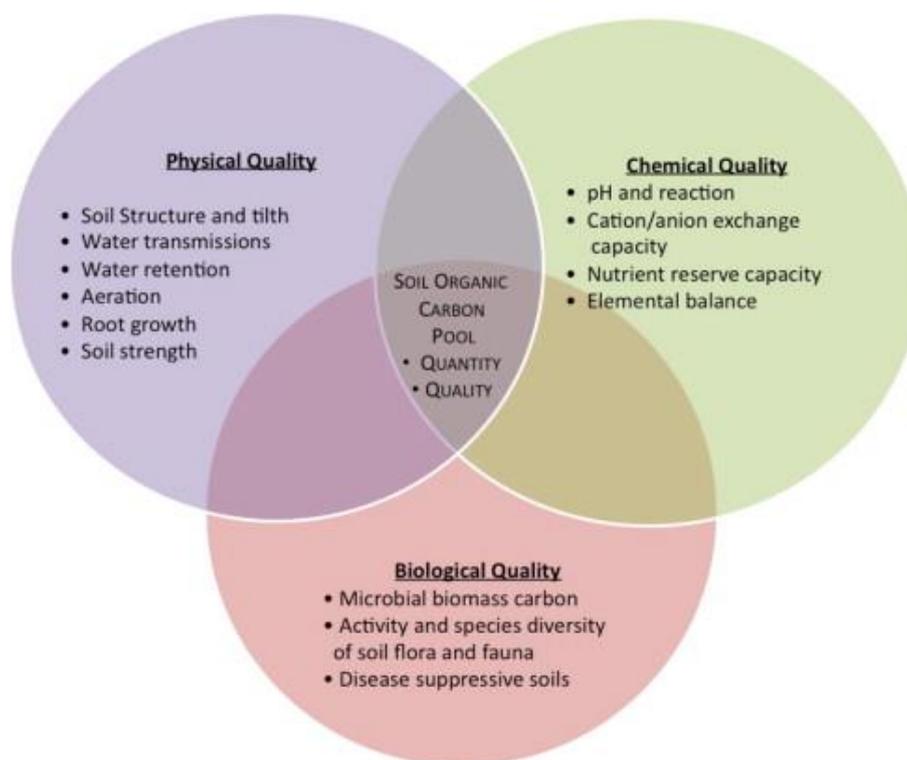
Figure 3-15 Soil quality improvements by increase in soil organic pool in agricultural soils (Lal, 2011)

The resistant or stable fraction of SOC mainly increases the soil's nutrient holding capacity (cation exchange capacity - CEC) and acts as a reservoir for plant nutrients (Banwart et al., 2015). Additionally, the SSOC's slow decomposing rates are especially interesting in terms of long-term SOC sequestration in soil (Lefèvre et al., 2017; Schmidt et al., 2011). Further positive features of enlarging the SOC pool are the increase of water holding capacity (Acín-Carrera et al., 2013; Karhu et al., 2011), greater soil permeability and therefore low runoff losses (Lal, 2004). Root growth is also benefiting (Kell, 2012; Mondal et al., 2020). With C input the availability of nutrients is increased through higher microbial activities (Fang et al., 2018). Other benefits are increasing the soils aggregate strength, leading to reduced erosion and acting as a buffer for sudden soil pH reactions (Lal, 2004). These factors emphasise the importance of increasing the quality and quantity of carbon in our managed soils (Lal, 2011). Additionally, there are also offsite functions with economic and environmental benefits, including reduced

sediment load in waterbodies, filtering pollutants and biodegradation of contaminants (Lal, 2004). The mentioned benefits are not a universal remedy for soils threats and climate change, but one important option to buffer their effects. For example, SSOC is prone to disturbances with multidimensional reasons, such as leaching, erosion and change in microbial activity caused by intensive land use, wrong management, climate change and extreme weather conditions. Exposure of C to O_2 results in microbial decomposition and therefore carbon loss happens. Accordingly, higher temperatures promote higher losses of SOC (Conant et al., 2011). This is one reason that the world's main soil SOC storage is located in the northern hemisphere in tundra, permafrost, wetlands and peat soils, which are most vulnerable to climate change (Scharlemann et al., 2014). One benefit named before relates to the role of SOC in CO_2 sequestration. Frank et al. (2015) calculated the mitigation potential of European cropland and concluded that we should put more emphasis on higher-effective mitigation measures like stopping deforestation, decreasing meat consumption to fight the increase of GHG. Prohibiting land use change is consequently more important for GHG mitigation than improvement of SSOC values regarding the small CO_2 storage capabilities for Lower Austria (Paustian et al., 2016; Smith, 2008). But Paustian et al. (2016) underlined that the focus with GHG-mitigation in context with agricultural soil should not be only on SOC storage but also on N_2O and CH_4 mitigation. He proposed a decision tree for cropland GHG mitigation practices, calling the management practices "climate smart agriculture". They estimated the total soil GHG mitigation potential worldwide, ranging from $5.3\,Pg\,CO_2(eq)\,yr^{-1}$ (without Economic constraints) to $1.5\,Pg\,CO_2(eq)\,yr^{-1}$ for agricultural land, being $50-10\%$ of the annual fossil fuel emissions (Ciais et al., 2013; Paustian et al., 2016), further accentuating the importance of correct soil management not only for SOC storage.

Management practices to increase SOC/SOCS

To achieve SOC accumulation in soil, different management practices are available. Lal (2016c) summarized the recent development and research in SOC sequestration and possible management techniques to accomplish a positive carbon budget, visible in Fig. 3-16.
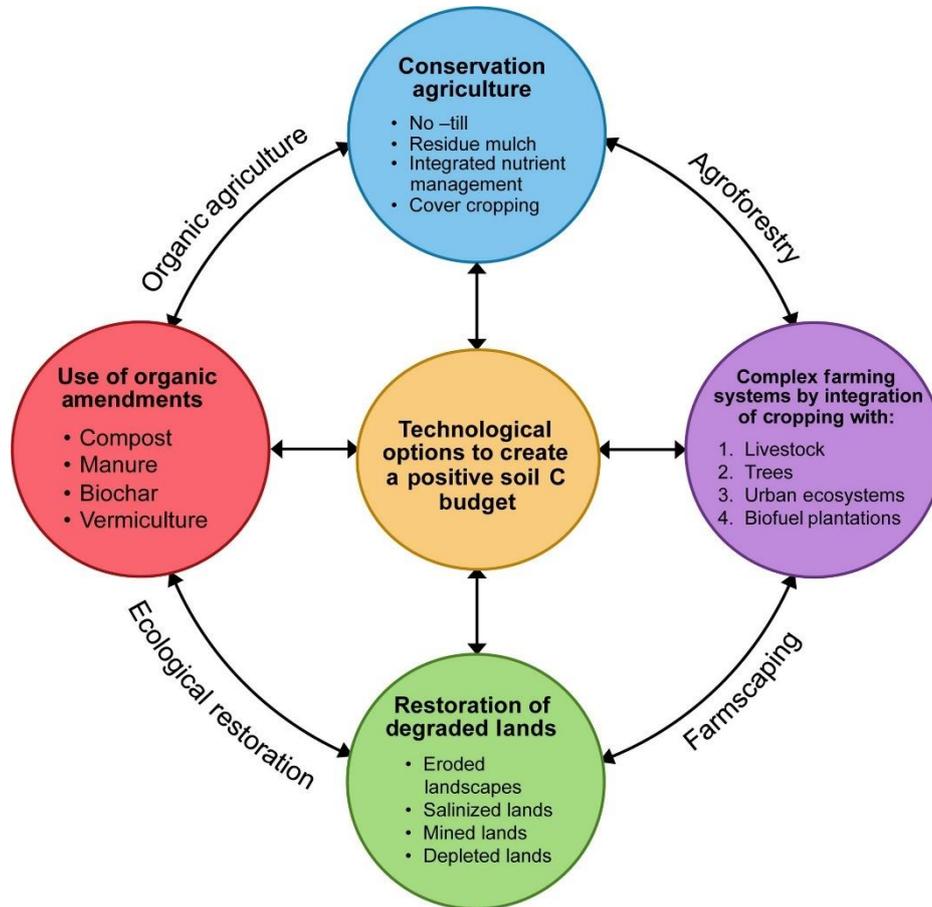
Figure 3-16 Technological options for soil carbon sequestration (Lal, 2016c)

These various options include applying organic agriculture with conservation agriculture, like cover cropping and no-till and the use of organic amendments like compost or biochar. Also, integrative farming with agroforestry and the restoration of degraded lands through farm scaping are options to alter the soil's content positively (Dawson and Smith, 2007; Lal, 2016c). With the use of cover crops and perennial cultures more root mass is introduced into soil, which again increases the carbon input through biomass (Rasse et al., 2005). All these options are creating synergies between each other. Only with a combined approach of used techniques and a broad application among farmers with the right policies, "the 4 per 1000" goal and a long-term improvement of soil quality can be achieved. The question remains, which the right incentives are to encourage farmers to apply the promoted techniques, even when resulting in a surplus of necessary labour. Carbon build-up in soil is a slow process, achievable only through long-term projects that needs time to reach lasting results, especially for SSOC (Olson et al., 2014), which together can make carbon build up to a tedious work and must be taken into account when changing the management practices of the land.

4 Conclusion and Outlook

The results show that usable prediction results can be achieved, even with soil samples lacking high-resolution satellite soil reflection images are available for the time of sampling.

The prediction result could significantly improve with the use of COVs from the Austrian soil map. Most of the available studies with similar spatial resolution are predicting smaller areas or having coarser resolutions. Therefore, it shows that predictions for larger areas with high resolution are possible with decent results.

The initially proposed research questions:

- Do high resolution predictions of SSOC, Csat and Cdef with environmental covariates delivers usable maps of acceptable accuracy and resolution?
- Will covariates of the Austrian soil map significantly improve SSOC, Csat and Cdef predictions?
- Does land use have a major influence on Cdef?

could all be evaluated and confirmed.

Recent soil sample data should ne used alongside new types of high-resolution covariates like spectral Sentinel satellite images, in different wavelengths for future predictions. Resulting in improved prediction accuracy. Combining different methods by stacking also yields promising results (Taghizadeh-Mehrjardi et al., 2020). Another possible improvement is the use of a feature selection technique called Boruta, which is based on shadow features and the iteration over them (Kursa and Rudnicki, 2010). Utilizing the latter would decrease the amount of COV and further reduce computation time and prediction noise. The creation of Dummy variables from categories to numeric can create the problem of underrepresentation of the converted COV in the prediction model. A slight generalisation of categorized COVs is a possible solution. Another option to improve computing time and storage problems is to decrease the final prediction resolution. For future studies, this should be focused more on improved concept planning, like selection of the needed types of statistics and pre-selection of the right parameters for comparison with the prediction results. Nevertheless, DSM methods should be used in future to improve several soil parameters of the Austrian Soil map and whenever there is a need for a high-resolution prediction of point data to spatial maps. This should have a major benefit for users of the Austrian soil map. In addition, it could be an improvement of the Austrian "Finanzbodenschätzung". In future analysis, not only SSOC, Cdef and Csat should be estimated but also soil carbon stocks (SOCS) SOCS sequestration potentials and SOCS deficits. This requires a precise information about the bulk density (BD), which should be included in future sampling campaigns. For BD, DSM methods can be used to create spatial maps which then can be utilized for SOCS, stock Csat and stock Cdef predictions. Without including BD another error biased has to be considered.

Therefore, the final prediction maps will give a very good baseline map for further research and comparison if new maps with recent soil data are created. It also proves that these novel methods can be applied successfully and should be used more frequently in future, to update existing and create new high-resolution prediction maps. It will be also a good starting point to make changes visible over time. At present, it will be used as a first reference map for advisory activities to improve agricultural management with practices stated above.

5 References

Acín-Carrera, M., José Marques, M., Carral, P., Álvarez, A.M., López, C., Martín-López, B., González, J.A., 2013. Impacts of land-use intensity on soil organic carbon content, soil structure and water-holding capacity. Soil Use Manage 29 (4), 547–556.

Alge, G., 1993. Bodenentwicklung auf quartären Sedimenten des westlichen Innviertels.

AMA - AgrarMarkt Austria, 2019. INVEKOS Feldstücke Österreich. https://www.data.gv.at/katalog/dataset/19b8d43f-4666-4a8c-98ef-c059fff9f192.

Angers, D., 1998. Water-stable aggregation of Québec silty clay soils: some factors controlling its dynamics. Soil and Tillage Research 47 (1-2), 91–96.

Angers, D.A., Arrouays, D., Saby, N.P.A., Walter, C., 2011. Estimating and mapping the carbon saturation deficit of French agricultural topsoils. Soil Use and Management 27 (4), 448–452.

Banwart, S.A., Noellemeyer, E., Milne, E. (Eds.), 2015. Soil carbon: Science, management and policy for multiple benefits. SCOPE series 71, Wallingford, Oxfordshire, 4 pp.

Barré, P., Angers, D.A., Basile-Doelsch, I., Bispo, A., Cécillon, L., Chenu, C., Chevallier, T., Derrien, D., Eglin, T.K., Pellerin, S., 2017. Ideas and perspectives: Can we use the soil carbon saturation deficit to quantitatively assess the soil carbon storage potential, or should we explore other strategies? Biogeosciences Discuss., 1–12.

Batjes, N.H., 1996. Total carbon and nitrogen in the soils of the world. Eur J Soil Sci 47 (2), 151–163.

Battin, T.J., Luyssaert, S., Kaplan, L.A., Aufdenkampe, A.K., Richter, A., Tranvik, L.J., 2009. The boundless carbon cycle. Nature Geosci 2 (9), 598–600.

Bfw, 2019. Digitale Bodenkarte von Österreich.

Bhunia, G.S., Shit, P.K., Maiti, R., 2018. Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon (SOC). Journal of the Saudi Society of Agricultural Sciences 17 (2), 114–126.

Breiman, L., 1998. Classification and regression trees, Repr ed. Chapman & Hall, Boca Raton, 358 pp.

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

Bröcker, F., Nestroy, O., 1995. Bodenkundliche Untersuchungen in der subalpinen und alpinen Stufe im Bereich der Kärntner Nockberge. Mitteilungen der Österreichischen Bodenkundlichen Gesellschaft.

Chen, S., Martin, M.P., Saby, N.P.A., Walter, C., Angers, D.A., Arrouays, D., 2018. Fine resolution map of top- and subsoil carbon sequestration potential in France. The Science of the total environment 630, 389–400.

Ciais, P., Chris, S., Govindasamy, B., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., Defries, R., Galloway, J., Heimann, M., 2013. Carbon and other biogeochemical cycles. Climate Change 2013: The Physical Science Basis, 465–570.

Conant, R.T., Ryan, M.G., Ågren, G.I., Birge, H.E., Davidson, E.A., Eliasson, P.E., Evans, S.E., Frey, S.D., Giardina, C.P., Hopkins, F.M., Hyvönen, R., Kirschbaum, M.U.F., Lavallee, J.M., Leifeld, J., Parton, W.J., Megan Steinweg, J., Wallenstein, M.D., Martin Wetterstedt,

J.Å., Bradford, M.A., 2011. Temperature and soil organic matter decomposition rates - synthesis of current knowledge and a way forward. Global Change Biology, 17(11), 3392-3404. Global Change Biol 17 (11), 3392–3404.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine Learning 20 (3), 273–297.

Dai, F., Zhou, Q., Lv, Z., Wang, X., Liu, G., 2014. Spatial prediction of soil organic matter content integrating artificial neural network and ordinary kriging in Tibetan Plateau. Ecological Indicators 45, 184–194.

Dawson, J.J., Smith, P., 2007. Carbon losses from soil and its consequences for land-use management. Science of The Total Environment 382 (2-3), 165–190.

Fang, Y., Nazaries, L., Singh, B.K., Singh, B.P., 2018. Microbial mechanisms of carbon priming effects revealed during the interaction of crop residue and nutrient inputs in contrasting soils. Global Change Biol 24 (7), 2775–2790.

Feng, W., Plante, A.F., Six, J., 2013. Improving estimates of maximal organic carbon stabilization by fine soil particles. Biogeochemistry 112 (1-3), 81–93.

Frank, S., Schmid, E., Havlík, P., Schneider, U.A., Böttcher, H., Balkovič, J., Obersteiner, M., 2015. The dynamic soil organic carbon mitigation potential of European cropland. Global Environmental Change 35, 269–278.

Gerzabek, M.H., Strebl, F., Tulipan, M., Schwarz, S., 2005. Quantification of organic carbon pools for Austria's agricultural soils using a soil information system. Can. J. Soil. Sci. 85 (Special Issue), 491–498.

Guevara, M., Olmedo, G.F., Stell, E., Yigini, Y., Aguilar Duarte, Y., Arellano Hernández, C., Arévalo, G.E., Arroyo-Cruz, C.E., Bolivar, A., Bunning, S., Bustamante Cañas, N., Cruz-Gaistardo, C.O., Davila, F., Dell Acqua, M., Encina, A., Figueredo Tacona, H., Fontes, F., Hernández Herrera, J.A., Ibelles Navarro, A.R., Loayza, V., Manueles, A.M., Mendoza Jara, F., Olivera, C., Osorio Hermosilla, R., Pereira, G., Prieto, P., Ramos, I.A., Rey Brina, J.C., Rivera, R., Rodríguez-Rodríguez, J., Roopnarine, R., Rosales Ibarra, A., Rosales Riveiro, K.A., Schulz, G.A., Spence, A., Vasques, G.M., Vargas, R.R., Vargas, R., 2018. No silver bullet for digital soil mapping: country-specific soil organic carbon estimates across Latin America. SOIL 4 (3), 173–193.

Hassink, J., Whitmore, A.P., Kubát, J., 1997. Size and density fractionation of soil organic matter and the physical capacity of soils to protect organic matter. European Journal of Agronomy 7 (1-3), 189–199.

Hazen, R.M., Hemley, R.J., Mangum, A.J., 2012. Carbon in Earth's interior: Storage, cycling, and life. Eos Trans. AGU 93 (2), 17–18.

Heimann, M., 1993. The Global Carbon Cycle. NATO ASI Series, Series I 15. Springer, Berlin, Heidelberg, 599 pp.

Hiebl, J., Reisenhofer, S., Auer, I., Böhm, R., Schöner, W., 2011. Multi-methodical realisation of Austrian climate maps for 1971–2000. Adv. Sci. Res. 6, 19–26.

Houghton, R.A., House, J.I., Pongratz, J., van der Werf, G.R., DeFries, R.S., Hansen, M.C., Le Quéré, C., Ramankutty, N., 2012. Carbon emissions from land use and land-cover change. Biogeosciences 9 (12), 5125–5142.

Jandl, R., 1987. Zur Genese ausgewählter Böden der österreichischen Nord- und Zentralalpen. Diplomarbeit. Institut für Bodenkunde der Universität für Bodenkultur, Vienna.

Jenny, H., 1941. Factors of soil formation: A system of quantitative pedology. McGraw-Hill publications in the agricultural sciences. McGraw-Hill, New York, 281 pp.

Karhu, K., Mattila, T., Bergström, I., Regina, K., 2011. Biochar addition to agricultural soil increased CH4 uptake and water holding capacity – Results from a short-term pilot field study. Agriculture, Ecosystems & Environment 140 (1-2), 309–313.

Katzensteiner, K., Ottner, F., Sieghart, M: Hager, H., Wresowar, M., Schume, H., 2001. Waldstandorte im mittleren Burgenland. Exkursionsführer zum Bodenkundlichen Kongress. Mitteilungen der Österreichischen Bodenkundlichen Gesellschaft (63).

Kell, D.B., 2012. Large-scale sequestration of atmospheric carbon via plant roots in natural and agricultural ecosystems: why and how. Phil. Trans. R. Soc. B 367 (1595), 1589–1597.

Kursa, M.B., Rudnicki, W.R., 2010. Feature Selection with the Boruta Package. J. Stat. Soft. 36 (11).

Lal, R., 2004. Soil carbon sequestration to mitigate climate change. Geoderma 123 (1-2), 1–22.

Lal, R., 2011. Sequestering carbon in soils of agro-ecosystems. Food Policy 36, S33-S39.

Lal, R., 2016a. Beyond COP 21: Potential and challenges of the "4 per Thousand" initiative. Journal of Soil and Water Conservation 71 (1), 20A-25A.

Lal, R. (Ed.), 2016b. Encyclopedia of soil science, Third edition ed. Culture and civilization in the Middle East 31. CRC Press Taylor & Francis Group, Boca Raton, London, New York, 50 pp.

Lal, R., 2016c. Soil health and carbon management. Food and Energy Security, 5(4), 212-222.

Lal, R., Negassa, W., Lorenz, K., 2015. Carbon sequestration in soil. Current Opinion in Environmental Sustainability 15, 79–86.

Land Niederösterreich, 2019. Verwaltungsgrenzen NÖ Linien 1:1000. https://www.data.gv.at/katalog/dataset/verwaltungsgrenzen-no-linien-1-1000.

Le Quéré, C., Raupach, M.R., Canadell, J.G., Marland, G., Bopp, L., Ciais, P., Conway, T.J., Doney, S.C., Feely, R.A., Foster, P., Friedlingstein, P., Gurney, K., Houghton, R.A., House, J.I., Huntingford, C., Levy, P.E., Lomas, M.R., Majkut, J., Metzl, N., Ometto, J.P., Peters, G.P., Prentice, I.C., Randerson, J.T., Running, S.W., Sarmiento, J.L., Schuster, U., Sitch, S., Takahashi, T., Viovy, N., van der Werf, G.R., Woodward, F.I., 2009. Trends in the sources and sinks of carbon dioxide. Nature Geoscience, 2(12), 831-836. Nature Geosci 2 (12), 831–836.

Lefèvre, C., Rekik, F., Alcantara Cervantes, V., Wiese, L., 2017. Soil organic carbon: The hidden potential. Food and Agriculture Organization of the United Nations, Rom, 77 pp.

Li, Z., Liu, C., Dong, Y., Chang, X., Nie, X., Liu, L., Xiao, H., Lu, Y., Zeng, G., 2017. Response of soil organic carbon and nitrogen stocks to soil erosion and land use types in the Loess hilly–gully region of China. Soil and Tillage Research 166, 1–9.

Liu, Y., Guo, L., Jiang, Q., Zhang, H., Chen, Y., 2015. Comparing geospatial techniques to predict SOC stocks. Soil and Tillage Research 148, 46–58.

Ma, Y., Minasny, B., Malone, B.P., McBratney, A.B., 2019. Pedology and digital soil mapping (DSM). Eur J Soil Sci 70 (2), 216–235.

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

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

Meersmans, J., van Wesemael, B., Goidts, E., van Molle, M., Baets, S. de, Ridder, F. de, 2011. Spatial analysis of soil organic carbon evolution in Belgian croplands and grasslands, 1960-2006. Global Change Biol 17 (1), 466–479.

Mentler, A., Rampazzo, N., Pamperl, S., Blum, W., 2001. Böden der Nördlichen Kalkalpen in mittleren Höhenlagen der Steiermark. Exkursionsführer zum Bodenkundlichen Kongress. Mitteilungen der Österreichischen Bodenkundlichen Gesellschaft (63).

Minasny, B., Malone, B.P., McBratney, A.B., Angers, D.A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z.-S., Cheng, K., Das, B.S., Field, D.J., Gimona, A., Hedley, C.B., Hong, S.Y., Mandal, B., Marchant, B.P., Martin, M., McConkey, B.G., Mulder, V.L., O'Rourke, S., Richer-de-Forges, A.C., Odeh, I., Padarian, J., Paustian, K., Pan, G., Poggio, L., Savin, I., Stolbovoy, V., Stockmann, U., Sulaeman, Y., Tsui, C.-C., Vågen, T.-G., van Wesemael, B., Winowiecki, L., 2017. Soil carbon 4 per mille. Geoderma 292, 59–86.

Mondal, S., Chakraborty, D., Bandyopadhyay, K., Aggarwal, P., Rana, D.S., 2020. A global analysis of the impact of zero-tillage on soil physical condition, organic carbon content, and plant root response. Land Degrad Dev 31 (5), 557–567.

Nayak, A.K., Rahman, M.M., Naidu, R., Dhal, B., Swain, C.K., Nayak, A.D., Tripathi, R., Shahid, M., Islam, M.R., Pathak, H., 2019. Current and emerging methodologies for estimating carbon sequestration in agricultural soils: A review. The Science of the total environment 665, 890–912.

Nelhiebel, P., Pecina, E., Baumgarten, A., Aust, G., 2001. Die Böden des Naturraumes Neusiedler See. Exkursionsführer zum Bodenkundlichen Kongress. Mitteilungen der Österreichischen Bodenkundlichen Gesellschaft (63).

Nestroy, O., Nelhiebel, P., Varallay, G., Bidlo, A., Surina, B: Bilek, P., 2001. Böden der grenznahen Gebiete Ungarns und der Slovakei. Exkursionsführer zum Bodenkundlichen Kongress. Mitteilungen der Österreichischen Bodenkundlichen Gesellschaft (63).

Niederösterreich, L., 1994. Niederösterreichische Bodenzustandsinventur. Wien : Amt d. NÖ Landesregierung, Abt. VI/4, 220 S., Ill., graph. Darst., Kt., 29 cm, Folie.

Olson, K.R., Al-Kaisi, M.M., Lal, R., Lowery, B., 2014. Experimental Consideration, Treatments, and Methods in Determining Soil Organic Carbon Sequestration Rates. Soil Science Society of America Journal, 78(2), 348-360. Soil Science Society of America Journal 78 (2), 348–360.

Oró, J., Miller, S.L., Lazcano, A., 1990. The origin and early evolution of life on Earth. Annual review of earth and planetary sciences 18, 317–356.

O'Rourke, S.M., Angers, D.A., Holden, N.M., McBratney, A.B., 2015. Soil organic carbon across scales. Global change biology 21 (10), 3561–3574.

Padarian, J., Minasny, B., McBratney, A.B., 2019. Using deep learning for digital soil mapping. SOIL 5 (1), 79–89.

Pahlavan Rad, M.R., Toomanian, N., Khormali, F., Brungard, C.W., Komaki, C.B., Bogaert, P., 2014. Updating soil survey maps using random forest and conditioned Latin hypercube sampling in the loess derived soils of northern Iran. Geoderma 232-234, 97–106.

Paustian, K., Lehmann, J., Ogle, S., Reay, D., Robertson, G.P., Smith, P., 2016. Climate-smart soils. Nature 532 (7597), 49–57.

Poggio, L., Gimona, A., Brewer, M.J., 2013. Regional scale mapping of soil properties and their uncertainty with a large number of satellite-derived covariates. Geoderma 209-210, 1–14.

Post, W.M., Emanuel, W.R., Zinke, P.J., Stangenberger, A.G., 1982. Soil carbon pools and world life zones. Nature 298 (5870), 156–159.

Pouladi, N., Møller, A.B., Tabatabai, S., Greve, M.H., 2019. Mapping soil organic matter contents at field level with Cubist, Random Forest and kriging. Geoderma 342, 85–92.

Ramesh, T., Bolan, N.S., Kirkham, M.B., Wijesekara, H., Kanchikerimath, M., Srinivasa Rao, C., Sandeep, S., Rinklebe, J., Ok, Y.S., Choudhury, B.U., Wang, H., Tang, C., Wang, X., Song, Z., Freeman II, O.W., 2019. Soil organic carbon dynamics: Impact of land use changes and management practices: A review, in: , vol. 156. Advances in Agronomy. Elsevier, pp. 1–107.

Rampazzo, N., Pamperl, S., Mentler, A., Blum, W., 2001. Sukzession von Böden an der Grenze Tertiär-Kristallin in Niederösterreich. Exkursionsführer zum Bodenkundlichen Kongress. Mitteilungen der Österreichischen Bodenkundlichen Gesellschaft (63).

Rasse, D.P., Rumpel, C., Dignac, M.-F., 2005. Is soil carbon mostly root carbon? Mechanisms for a specific stabilisation. Plant and Soil, 269(1-2), 341-356. Plant and Soil 269 (1-2), 341–356.

Roston, E., 2010. The carbon age: how life's core element has become civilization's greatest threat. Bloomsbury Publishing USA.

Ruddiman, W.F., 2003. The Anthropogenic Greenhouse Era Began Thousands of Years Ago. Climatic Change 61 (3), 261–293.

Scharlemann, J.P.W., Tanner, E.V.J., Hiederer, R., Kapos, V., 2014. Global soil carbon: understanding and managing the largest terrestrial carbon pool. Carbon Management 5 (1), 81–91.

Schimel, D.S., 1995. Terrestrial ecosystems and the carbon cycle. Global Change Biol 1 (1), 77–91.

Schmidt, M.W.I., Torn, M.S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I.A., Kleber, M., Kögel-Knabner, I., Lehmann, J., Manning, D.A.C., Nannipieri, P., Rasse, D.P., Weiner, S., Trumbore, S.E., 2011. Persistence of soil organic matter as an ecosystem property. Nature 478 (7367), 49–56.

Schneider, W., Danneberger, O.H., Baumgarten, A., Aust, G., 2001. Die Böden der Donauterassen östlich von Wien. Exkursionsführer zum Bodenkundlichen Kongress. Mitteilungen der Österreichischen Bodenkundlichen Gesellschaft (63).

Six, J., Conant, R.T., Paul, E.A., Paustian, K., 2002. Stabilization mechanisms of soil organic matter: Implications for C-saturation of soils. Plant and Soil 241 (2), 155–176.

Smith, P., 2008. Land use change and soil organic carbon dynamics. Nutr Cycl Agroecosyst 81 (2), 169–178.

Statistik Austria, 2018. Statistik der Landwirtschaft 2018.

Stewart, C.E., Paustian, K., Conant, R.T., Plante, A.F., Six, J., 2007. Soil carbon saturation: concept, evidence and evaluation. Biogeochemistry 86 (1), 19–31.

Strauss, F., Formayer, H., Schmid, E., 2013. High resolution climate data for Austria in the period 2008-2040 from a statistical climate change model. Int. J. Climatol. 33 (2), 430–443.

Strauss, P., Klaghoger, E., Schneider, W., 2001. Bodenerosion im niederösterreichischen Donauraum. Exkursionsführer zum Bodenkundlichen Kongress. Mitteilungen der Österreichischen Bodenkundlichen Gesellschaft (63).

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 (7), 1095.

Wackernagel, H. (Ed.), 1995. Multivariate Geostatistics: An Introduction with Applications. Springer, Berlin, Heidelberg, 257 pp.

Wenzel, W.W., unpublished. unpublished.

Wessely, G., Draxler, I., 2006. Niederösterreich: Geologie der österreichischen Bundesländer. Geol. Bundestanstalt, Wien, 416 pp.

Whitehead, D.C., Tinsley, J., 1964. EXTRACTION OF SOIL ORGANIC MATTER WITH DIMETHYLFORMAMIDE. Soil Science 97 (1), 34–42.

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

Wiesmeier, M., Spörlein, P., Geuß, U., Hangen, E., Haug, S., Reischl, A., Schilling, B., Lützow, M., Kögel-Knabner, I., 2012. Soil organic carbon stocks in southeast Germany (Bavaria) as affected by land use, soil type and sampling depth. Global Change Biol 18 (7), 2233–2245.

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

Wischmeier, W.H., Smith, D.D., 1978. Predicting rainfall erosion losses: A guide to conservation planning. United Sates Department of Agriculture. Agriculture handbook 537. U.S. Gov. Print. Off., Washington, D.C., 58 pp.

Yigini, Y., Olmedo, G.F., Reiter, S., Baritz, R., Viatkin, K. and Vargas, R. (Eds.), 2018. Soil Organic Carbon Mapping Cookbook 2nd Edition. FAO, Rome, 223 pp.

Zeebe, R.E., Ridgwell, A., Zachos, J.C., 2016. Anthropogenic carbon release rate unprecedented during the past 66 million years. Nature Geoscience, 9(4), 325-329. Nature Geosci 9 (4), 325–329.

6 Index of Tables

7 Index of Figures

8 Appendix

8.1 R-Code

pointdata-location-calculation.R

```r
#install.packages("remotes")
#remotes::install_github("SEEG-Oxford/seegSDM"
library(seegSDM)
library(sf)
library(raster)
library(rgdal)
library(tidyverse)
setwd('C:/Mapping/R')

#Load Data

point = st_read("C:/Mapping/Soil/BZI_fin.shp")
st_crs(point) = 31259

point = shapefile("C:/Mapping/Soil/BZI_fin.shp")
invekos = raster("F:/Masterarbeit Daten -
FIN/mask/maske_invekos/maske_invekos.tif")
eBOD = raster("F:/Masterarbeit Daten -
FIN/mask/maske_eBOD/maske_eBOD_fin.tif")
crs(point) <- CRS('+init=EPSG:31259')
crs(invekos) <- CRS('+init=EPSG:31259')
crs(eBOD) <- CRS('+init=EPSG:31259')

#calculation change eBOD to invekos for the other mask calculation

vals <- raster::extract(eBOD, point)
outside_mask <- is.na(vals)
outside_pts <- point[outside_mask, ]

c = 35 # distance to closest cell inside of the mask
#calculation
results = nearestLand(outside_pts, eBOD, c)
summary(results)
summary(is.na(results))
results = as.data.frame(results)
names(results)[names(results) == 'x'] <- 'X'
names(results)[names(results) == 'y'] <- 'Y'

#ad new coordinates to points and delete NA's not in in range
b_results = cbind.data.frame(outside_pts,results)
q = b_results %>% drop_na()
```

41

```r
q$geometry = NULL
q$x = NULL
q$y = NULL

#points to dataframe
w = cbind(point,st_coordinates(point))
w$geometry = NULL
w$x = NULL
w$y = NULL

#save only points that fit
outside_ptsdf = cbind(outside_pts,st_coordinates(outside_pts))
outside_ptsdf$geometry = NULL
outside_ptsdf$x = NULL
outside_ptsdf$y = NULL

d = anti_join(w,outside_ptsdf, by = 'NR_PROFL')

#hinzufügen von
t = full_join(d,q)

#create sf object
eBOD_point = st_as_sf(t,coords = c("X", "Y"), crs = 31259) # or
invekos_pointst = ...

#Safe as gpkg
st_write(eBOD_point, dsn= "pointsBZI.gpkg", layer = "eBOD_point", driver =
"GPKG")
st_write(invekos_pointst, dsn= "pointsBZI.gpkg", layer =
"invekos_pointst", driver = "GPKG")


#test point location
valst <- raster::extract(invekos, invekos_pointst)
outside_maskt <- is.na(valst)
outside_ptst <- invekos_pointst[outside_maskt, ]
summary(outside_maskt)

#display
plot(invekos)
plot(invekos_point, pch = ".", add = T)
```

soil-value-calculation.R

```r
#load data
setwd('C:/Users/IBF/Documents/Masterarbeit MAYR/Pointdata/fin')

#data
train.i = read.csv("dat_train_i.csv")
test.i = read.csv("dat_test_i.csv")
train.e = read.csv("dat_train_e.csv")
test.e = read.csv("dat_test_e.csv")

#delete old calc
dat = train.e # --> change for each data set
dat$SOC = NULL
dat$SSOC = NULL
dat$Csat = NULL
dat$Cdef = NULL
dat$part_fra = NULL
dat$'pred.p.f.<20u'= NULL
dat$'pred.p.f..20u'= NULL

colnames(dat)[106] <- "X"
colnames(dat)[107] <- "Y"

#calc new variables
dat$SOC = ((dat$HUMUS/1.724)*1.2*10
dat$SSOC = dat$SOC*0.85
dat$part_fra = 0.3171*(dat$KF_TON + dat$KF_SCHLU)^1.1647
dat$Csat = 1.227*dat$part_fra
dat$Cdef = dat$Csat - dat$SSOC
train.e = dat # --> change for each data set

#save point data
write.csv(train.i, file="dat_train_i_f.csv", row.names = FALSE,
fileEncoding = "UTF-8")
write.csv(train.e, file="dat_train_e_f.csv", row.names = FALSE,
fileEncoding = "UTF-8")
write.csv(test.i, file="dat_test_i_f.csv", row.names = FALSE, fileEncoding
= "UTF-8")
write.csv(test.e, file="dat_test_e_f.csv", row.names = FALSE, fileEncoding
= "UTF-8")
```

43

DataSplitting.R

```r
## DataSplitting

library(caret)
library(sf)

setwd('C:/Mapping/R')

dat_point = st_read("pointsBZI.gpkg",layer = "eBOD_point")
dat = cbind(dat_point,st_coordinates(dat_point))
dat = as.data.frame(dat)
dat$geom = NULL

train.ind <- createDataPartition(1:nrow(dat), p = .75, list = FALSE)
train <- dat[ train.ind,]
test  <- dat[-train.ind,]

plot(density (log(train$SOC)), col='red',
     main='Statistical distribution of train and test datasets')
lines(density(log(test$SOC)), col='blue')
legend('topright', legend=c("train", "test"),
       col=c("red", "blue"), lty=1, cex=1.5)

write.csv(train, file="dat_train_eBOD.csv", row.names = FALSE)
write.csv(test, file="dat_test_eBOD.csv", row.names = FALSE)
```

Dummy-Variables.R

```r
###Create DummyVariables###
###########################
#packages
library(sp)
library(raster)

#functions
dummyRaster <- function(rast){
  rast <- as.factor(rast)
  result <- list()
  for(i in 1:length(levels(rast)[[1]][[1]])){
    result[[i]] <- rast == levels(rast)[[1]][[1]][i]
    names(result[[i]]) <- paste0(names(rast),
                                 levels(rast)[[1]][[1]][i])
  }
  return(stack(result))
}


#load Data
cov = "C:/Users/IBF/Documents/Masterarbeit MAYR/layer_eBod/Faktoren"
cov = "C:/Users/IBF/Documents/Masterarbeit MAYR/layer_eBod/Faktoren main"
```

44

```r
files <- list.files(path = cov, pattern = "tif$",
                    full.names = TRUE)
covs <- stack(files)
names(covs)
crs(covs) <- CRS('+init=EPSG:31259')

# convert  from factor to dummy

#ebodfaktoren
COV2= stack()

for (name in names(covs)){
  print(Sys.time())
  tStart <- Sys.time()

  name <- dummyRaster(covs[[name]])
  COV2 = stack(COV2, name)

  tEND <- Sys.time()
  print(paste(tEND," Loop",name,"finished!"))
  print(Sys.time() - tStart)
}

#save as Raster!!
for (name in names(COV2)){
  print(Sys.time())
  tStart <- Sys.time()

  writeRaster(COV2[[name]],paste("C:/Users/IBF/Documents/Masterarbeit MAYR
/layer_eBod/faktoren_dummy/",names(COV2[[name]]),".tif", sep = ""))

  tEND <- Sys.time()
  print(paste(tEND," Loop",names(name),"finished!"))
  print(Sys.time() - tStart)
}

#load Data
covi = "C:/Users/IBF/Documents/Masterarbeit MAYR/layer_invekos/Faktoren"

files <- list.files(path = covi, pattern = "tif$",
                    full.names = TRUE)
covis <- stack(files)
names(covis)
crs(covis) <- CRS('+init=EPSG:31259')

# convert  from factor to dummy

#invekos faktoren
COVi= stack()

for (name in names(covis)){
  print(Sys.time())
  tStart <- Sys.time()

  name <- dummyRaster(covis[[name]])
```

```r
  COVi = stack(COVi, name)

  tEND <- Sys.time()
  print(paste(tEND," Loop",name,"finished!"))
  print(Sys.time() - tStart)
}

#save as Raster!!
for (name in names(COVi)){
  print(Sys.time())
  tStart <- Sys.time()

  writeRaster(COVi[[name]],paste("C:/Users/IBF/Documents/Masterarbeit MAYR
/layer_invekos/faktoren_dummy/",name,".tif", sep = ""))

  tEND <- Sys.time()
  print(paste(tEND," Loop",name,"finished!"))
  print(Sys.time() - tStart)
}

save(COVi, file = "covariates_invekos_dummy_clc_geo.RData")
save(COV, file = "covariates_ebod_dummy.RData")
save(COV2, file = "covariates_ebod_dummy_clc_geo.RData")
```

StepwiseRegressionKriging.R

```r
## Stepwise Regression Kriging
####################################
library(raster)
library(car)
library(rgdal)
library(sf)
library(dplyr)
library(ggplot2)
#############
library(xlsx)
#############
setwd('C:/Users/IBF/Documents/Masterarbeit MAYR/WD')

######################################################################################
##from here only one time. Load afterwards only the R files
#cov preperation only once!!! load later just the R file

#read data for calculation
cov = "C:/Users/IBF/Documents/Masterarbeit MAYR/layer_invekos"
dat = readOGR(dsn = "C:/Users/IBF/Seafile/Meine Bibliothek/Masterarbeit/Da
ta/Punktdaten/pointsBZI.gpkg", layer = "invekos_point")
crs(dat) <- CRS('+init=EPSG:31259')

#dat = shapefile("C:/Mapping/Soil/BZI.shp",use_iconv = TRUE, encoding="UTF
-8")

#plot sample locations in boundaries of lower autria
```

```r
test = raster("C:/Mapping/Mask/test.tif")
plot(test)
plot(datn, pch = ".", add = T)



files <- list.files(path = cov, pattern = "tif$",
                    full.names = TRUE)
covs <- stack(files)
crs(covs) <- CRS('+init=EPSG:31259')

save(covs, file = "covariates_invekos_RK.RData")



covs@layers[[5]] = as.factor(covs@layers[[5]])
covs@layers[[2]] = as.factor(covs@layers[[2]])


#saveRDS(covs, file = "covariates10m.rds")

#extract at sample points data from covs
datx <- extract(x = covs, y = datx, sp = TRUE)

summary(dat@data)
#delete old calc
dat$SSOC = NULL
dat$Csat = NULL
dat$Cdef = NULL
dat$part_fra = NULL
#calc new variables
dat$SOC = ((dat$HUMUS*1.724)*1.2)
dat$SSOC = dat$SOC*0.84
dat$'pred.p.f.<20u'= dat$KF_TON + (0.3171*dat$KF_SCHLU^1.1647)
dat$Csat.W = 1.227*dat$`pred.p.f.<20u`

#save as dataframe
dat <- as.data.frame(dat)
summary(dat)

# The points with NA values have to be removed
dat <- dat[complete.cases(dat),]

#reorder columns
dat <- dat[c(1:105,119,106,120,121,107:118)]
summary(dat)



coordinates(dat) <- ~ coords.x1 + coords.x2
crs(dat) <- CRS('+init=EPSG:31259')
#make categorial variables to factors
#dataframe
dat$geo = as.factor(dat$geo)
dat$clc = as.factor(dat$clc)

dat@data$geo = as.factor(dat@data$geo)
```

```r
dat@data$clc = as.factor(dat@data$clc)


#rename columns
names(dat)[116] <- "twi.1"

#Export as a *.csv table
write.csv(dat, "RegMatrix.csv", row.names = FALSE,fileEncoding = "UTF-8")
#write.xlsx(dat, file = "RegMatrix.xlsx", sheetName="RegMatrix",
#           col.names=TRUE, row.names=FALSE)
################################################################################
#covs laden
load("covariates_invekos.RData")
covs = cov.new
plot(raster::stack(covs))
dat = read.csv("RegMatrix.csv",fileEncoding = "UTF-8")
coordinates(dat) <- ~ x + y

datdf = as.data.frame(datx@data)
datdf = datdf[,c("SSOC",names(covs))]

###############
model.MLR.step.RData = load("model.MLR.step.RData")
model.MLR.step.RData.cd = load( "model.MLR.step.cd.RData")
datdf.cd = load("datdf.cd.RData")
################

#Kriging
#log transformation and model calc
model.MLR <- lm(log(SSOC) ~ ., data = datdf)

summary(model.MLR)

# Stepwise variable selection
model.MLR.step <- step(model.MLR, direction="both")
summary(model.MLR.step)
anova(model.MLR.step)
par(mfrow=c(2,2))
plot(model.MLR.step)
plot(model.MLR)

# Collinearity test using variance inflation factors
library(car)
vif(step(model.MLR, direction="both"))
vif(model.MLR.step)
# Problematic covariates should have sqrt (VIF) > 2
sqrt(vif(model.MLR.step))

# Removing "?" from the stepwise model
model.MLR.step <- update(model.MLR.step, . ~ . - temp)

# Test the vif again
sqrt(vif(model.MLR.step))

#check results
```

48

```r
summary(model.MLR.step)
anova(model.MLR.step)
par(mfrow=c(2,2))
plot(model.MLR.step)

# Outlier test using the Bonferroni test
outlierTest(model.MLR.step)

################################################################################
shapiro.test(log(datdf$SSOC))

names(log(datdf$SSOC))

datdf = data
cd = cooks.distance(model.MLR.step)
plot(cd)

# Defining outliers based on 4/n criteria;
cd.outlier <- ifelse(cd < 4/nrow(datdf), "keep","delete")
summarise(cd.outlier)
####

# Plot the Cook's Distance using the traditional 4/n criterion
sample_size <- nrow(datdf)
plot(cd, pch="*", cex=2, main="Influential Obs by Cooks distance")  # plot
cook's distance
abline(h = 4/sample_size, col="red")  # add cutoff line
text(x=1:length(cd)+1, y=cd, labels=ifelse(cd>4/sample_size, names(cd),"")
, col="red") # add labels

top_x_outlier <- 4
influential <- as.numeric(names(sort(cd, decreasing = TRUE)[1:top_x_outlie
r]))

datdf.cd <- datdf[-influential, ]
datdf.cd <- datdf[-69, ]

#log transformation and model calc
model.MLR.cd <- lm(log(SSOC) ~ ., data = datdf.cd)

summary(model.MLR.cd)

# Stepwise variable selection
model.MLR.step.cd <- step(model.MLR.cd, direction="both")
summary(model.MLR.step.cd)
anova(model.MLR.step.cd)
par(mfrow=c(2,2))
plot(model.MLR.step.cd)
plot(model.MLR.cd)

qqPlot(model.MLR.step.cd)
qqPlot(model.MLR.step)

# Outlier test using the Bonferroni test
outlierTest(model.MLR.step.cd)
```

```r
outlierTest(model.MLR.step)

# Collinearity test using variance inflation factors
library(car)
vif(step(model.MLR, direction="both"))
vif(model.MLR.step)

# Problematic covariates should have sqrt (VIF) > 2
sqrt(vif(model.MLR.step.cd))

# Removing "dem" from the stepwise model
model.MLR.step.cd <- update(model.MLR.step.cd, . ~ . - dem)

# Test the vif again
sqrt(vif(model.MLR.step.cd))

#check results
summary(model.MLR.step.cd)
anova(model.MLR.step.cd)
par(mfrow=c(2,2))
plot(model.MLR.step.cd)

cd = cooks.distance(model.MLR.step.cd)
plot(cd)

ggplot(data = datdf.cd, aes(x = row(datdf.cd), y = dist)) + geom_point() +
  geom_smooth(method = lm)

autoplot(datdf.cd)
plot(model.MLR.step.cd)
#### save models
save(model.MLR.step, file = "model.MLR.step.RData")
save(model.MLR.step.cd, file = "model.MLR.step.cd.RData")
save(datdf.cd, file = "datdf.cd.RData")
save(datdf, file = "datdf.RData")

################################################################################
# Promote covariates to spatial grid dataframe.
# Takes some time and a lot of memory!
# set memory limit
memory.limit(999999999)

#create sp object
covs.sp <- as(covs, "SpatialGridDataFrame") #not working connot allocate v
ector of size 11,7gb

#make categorial variables to factors incovs rasterstack if they are not a
lready

################################################################################
#removing not relevant layers (only if covs.sp is too big-- remove them be
fore creating sp data)

#check which layers are left
model.MLR.step$call$formula
```

50

```r
save(covs, file = "covariates_invekos.RData")

cov.new = dropLayer(covs, c(1,2,5,7))

covs.sp <- as(cov.new, "SpatialGridDataFrame")

covs.sp$geo = as.factor(covs.sp$geo)
covs.sp$clc = as.factor(covs.sp$clc)

#same levels for all data
levels(datdf$geo) = levels(covs.sp@data$geo)
levels(datdf$clc) = levels(covs.sp@data$clc)

levels(dat@data$geo) = levels(covs.sp@data$geo)
levels(dat@data$clc) = levels(covs.sp@data$clc)

##############################################################################
# Run regression-kriging prediction.
# This step can take hours!

#dat to sp object
coordinates(dat) <- ~ coords.x1 + coords.x2
crs(dat) <- CRS('+init=EPSG:31259')
crs(covs.sp) <- CRS('+init=EPSG:31259')

library(automap)
SSOC.krige <- autoKrige(formula =
                        as.formula(model.MLR.step$call$formula),
                    input_data = datx,
                    new_data = covs.sp,
                    verbose = TRUE,
                    block = c(1000, 1000))
SSOC.krige

# Convert prediction and standard deviation to rasters
# And back-tansform the vlaues
RKprediction <- exp(raster(SSOC.krige$krige_output[1]))
RKpredsd <- exp(raster(SSOC.krige$krige_output[3]))

plot(RKprediction)
plot(RKpredsd)

## Save results as tif files
writeRaster(RKprediction, filename = "results/SSOC_RK.tif",
            overwrite = TRUE)

writeRaster(RKpredsd, filename = "results/SSOC_RKpredsd.tif",
            overwrite = TRUE)

# save the model
saveRDS(model.MLR.step, file="results/RKmodel.Rds")

#Cross-validation

# autoKrige.cv() does not removes the duplicated points
```

51

```r
# we have to do it manually before running the cross-validation
datx2 = datx[which(!duplicated(datx@coords)), ]

SSOC.krige.cv <- autoKrige.cv(formula =
                                as.formula(model.MLR.step$call$formula),
                            input_data = datx2, nfold = 10)
summary(SSOC.krige.cv)
```

Random Forest.R

```r
## randomForest
library(raster)
library(rgdal)
library(randomForest)
library(caret)
library(snow)

setwd('C:/Users/IBF/Documents/Masterarbeit MAYR/WD')

load("covariates_invekos_RF.RData")
crs(covs) <- CRS('+init=EPSG:31259')

#data invekos
train = read.csv("C:/Users/IBF/Documents/Masterarbeit MAYR/Pointdata/fin/f
inal/dat_train_i_f1.csv")

#data eBod
train = read.csv("C:/Users/IBF/Documents/Masterarbeit MAYR/Pointdata/fin/f
inal/dat_train_e_f2.csv")

#e-BOD Covs Laden
cov = "C:/Users/IBF/Documents/Masterarbeit MAYR/layer_eBod/main_new"
files <- list.files(path = cov, pattern = "tif$",
                  full.names = TRUE)
covs <- stack(files)
crs(covs) <- CRS('+init=EPSG:31259')

#dummy eBod covs laden
cov = "C:/Users/IBF/Documents/Masterarbeit MAYR/layer_eBod/faktoren_dummy"
files <- list.files(path = cov, pattern = "tif$",
                  full.names = TRUE)
covd <- stack(files)
crs(covs) <- CRS('+init=EPSG:31259')
#combine stacks
covs = stack(covs,covd)

#covs laden  invekos
cov = "C:/Users/IBF/Documents/Masterarbeit MAYR/layer_invekos"
files <- list.files(path = cov, pattern = "tif$",
                  full.names = TRUE)
covs <- stack(files)
crs(covs) <- CRS('+init=EPSG:31259')
covs = dropLayer(covs,2)
```

52

```r
covs = dropLayer(covs,4)

#dummy covs laden
cov = "C:/Users/IBF/Documents/Masterarbeit MAYR/layer_invekos/faktoren_dummy"
files <- list.files(path = cov, pattern = "tif$",
                    full.names = TRUE)
covd <- stack(files)
crs(covs) <- CRS('+init=EPSG:31259')

#combine stacks
covs = stack(covs,covd)

#extract covs data to point

datx <- train[c(104:105,110)]

coordinates(datx) <- ~ X + Y
crs(datx) <- CRS('+init=EPSG:31259')

datn <- extract(x = covs, y = datx, sp = TRUE)
crs(datn) <- CRS('+init=EPSG:31259')
options(max.print=1000000)

summary(datn@data)

datntest <- as.data.frame(datn)

datntest <- datntest[complete.cases(datntest),]
datn = datntest
coordinates(datn) <- ~ X + Y
crs(datn) <- CRS('+init=EPSG:31259')

#We have to convert the columns with categorical variables in the soil samples
#data.frame to dummies as well. For doing this we can use function
#model.matrix(). After this, we use cbind() to merge the resulting data.frame.
# Convert soilmap column to dummy, the result is a matrix
# To have one column per category we have to add -1 to the formula
dat_clcmap_dummy <- model.matrix(~clc -1, data = datn@data)
dat_geomap_dummy <- model.matrix(~geo -1, data = datn@data)

# Convert the matrix to a data.frame
dat_clcmap_dummy <- as.data.frame(dat_clcmap_dummy)
dat_geomap_dummy <- as.data.frame(dat_geomap_dummy)

datn@data <- cbind(datn@data, dat_clcmap_dummy, dat_geomap_dummy)
names(datn@data)

################################################################

####make categorial variables to factors
#same levels for all data
datx@data$geo = as.factor(datx@data$geo)
```

53

```r
datx@data$clc = as.factor(datx@data$clc)

geo.levels = levels(as.factor(geo.levels[[1]]$ID))
clc.levels = levels(as.factor(clc.levels[[1]]$ID))

levels(datx@data$geo) = geo.levels
levels(datx@data$clc) = clc.levels

summary(datn)

datf<- datf[complete.cases(datf),]

######################################################
writeOGR(datx, dsn= "datx.gpkg", layer = "datx_v3", driver = "GPKG", overw
rite_layer = TRUE)

dat <- readOGR("dat.gpkg", "dat")

coordinates(datx) <- ~ coords.x1 + coords.x2
crs(datx) <- CRS('+init=EPSG:31259')


################### RF PREDICTION #############################
# For its use on R we need to define a model formula
# For not normal distributed data transformation is needed, like log.
fm = as.formula(paste("log(SSOC) ~", paste0(names(covs[[-167]]),collapse =
"+")))

# Default 10-fold cross-validation
ctrl <- trainControl(method = "cv", savePred=T)

# Search for the best mtry parameter
rfmodel <- caret::train(fm, data=datn@data, method = "rf", trControl = ctr
l,importance=TRUE)

# This is a very useful function to compare and test different
# prediction algorithms type names(getModelInfo()) to see all the
# possibilitites implemented on this function

# Variable importance plot, compare with the correlation matrix
# Select the best prediction factors and repeat
varImpPlot(rfmodel[11][[1]])

varImp(rfmodel)

##############################
# Check if the error stabilizes
plot(rfmodel[11][[1]])

#Make a prediction across all Lower Austria
#Note that the units are still in log
pred <- predict(covs, rfmodel)

# Back transform predictions log transformed data if it was transformed
pred <- exp(pred)
```

54

```r
# Save the result as a tiff file
writeRaster(pred, filename = "results/FIN2/SSOC_e_fk_rf_fin2.tif",
            overwrite=FALSE)
#back transform again for rf-qr
pred <- log(pred)
crs(pred) <- CRS('+init=epsg:31259')

#Plot prediction
pred = raster("results/FIN2/SSOC_i_fk_rf.tif")
```

SVM.R

```r
## SupportVectorMachines

# plot the names of the covariates
train = read.csv("dat_train.csv")

names(covs)
daty <- train[c(109:111)]
daty = cbind(daty,datdf[,2:9])
coordinates(daty) <- ~ coords.x1 + coords.x2
crs(daty) <- CRS('+init=EPSG:31259')

# variable selection using correlation analysis
selectedCovs <- cor(x = as.matrix(daty@data[,1]),
                    y = as.matrix(datx@data[,2:9]))

# print correlation results
selectedCovs

library(reshape)
x <- subset(melt(selectedCovs), value != 1 | value != NA)
x <- x[with(x, order(-abs(x$value))),]

idx <- as.character(x$Var2[1:5])

dat2 <- daty[c('Csat.W', idx)]
names(dat2)

COV <- covs[[idx]]

# Selected covariates
names(COV)

# Categorical variables in svm models
dummyRaster <- function(rast){
  rast <- as.factor(rast)
  result <- list()
  for(i in 1:length(levels(rast)[[1]][[1]])){
    result[[i]] <- rast == levels(rast)[[1]][[1]][i]
    names(result[[i]]) <- paste0(names(rast),
                                 levels(rast)[[1]][[1]][i])
```

55

```r
  }
  return(stack(result))
}

# convert soilmap from factor to dummy
# soilmap_dummy <- dummyRaster(covs$soilmap)

# convert LCEE10 from factor to dummy
LCEE10_dummy <- dummyRaster(covs$LCEE10)

# Stack the 5 COV layers with the 2 dummies
COV <- stack(COV, LCEE10_dummy)
# COV <- stack(COV, soilmap_dummy, LCEE10_dummy)

# print the final layer names
names(COV)

# convert soilmap column to dummy, the result is a matrix
# to have one column per category we had to add -1 to the formula
# dat_soilmap_dummy <- model.matrix(~soilmap -1, data = dat@data)
# convert the matrix to a data.frame
# dat_soilmap_dummy <- as.data.frame(dat_soilmap_dummy)


# convert LCEE10 column to dummy, the result is a matrix
# to have one column per category we had to add -1 to the formula
dat_LCEE10_dummy <- model.matrix(~LCEE10 -1, data = dat@data)
# convert the matrix to a data.frame
dat_LCEE10_dummy <- as.data.frame(dat_LCEE10_dummy)

dat@data <- cbind(dat@data, dat_LCEE10_dummy)
# dat@data <- cbind(dat@data, dat_LCEE10_dummy, dat_soilmap_dummy)

names(dat@data)

# Fitting a svm model and parameter tuning
library(e1071)
library(caret)

#  Test different values of epsilon and cost
tuneResult <- tune(svm, Csat.W ~.,  data = daty@data[,c("Csat.W",names(COV
))],
                   ranges = list(epsilon = seq(0.05,0.5,0.02), cost = c(2,
5,7,15,20)))

plot(tuneResult)

# Choose the model with the best combination of epsilon and cost
tunedModel <- tuneResult$best.model

print(tunedModel)


# Use the model to predict the SOC in the covariates space
CsatSVM <- predict(COV, tunedModel)
```

56

```r
# Save the result
writeRaster(CsatSVM, filename = "results/CsatSVM_svm.tif",
            overwrite=TRUE)

plot(CsatSVM)

# Variable importance in svm. Code by:
# stackoverflow.com/questions/34781495

w <- t(tunedModel$coefs) %*% tunedModel$SV     # weight vectors
w <- apply(w, 2, function(v){sqrt(sum(v^2))})  # weight

w <- sort(w, decreasing = T)
print(w)
```

validation_prediction.R

```r
#Validation
library(sp)
library(raster)
library(sjmisc)
library(xlsx)

#Load test data points
#invekos
test_i <- read.csv("F:/Masterarbeit Daten -
FIN/points/final_point/dat_test_i_F.csv")
coordinates(test_i) <- ~ X + Y
crs(test_i) <- CRS('+init=EPSG:31259')

#eBod
test_e <- read.csv("F:/Masterarbeit Daten -
FIN/points/final_point/dat_test_e_f.csv")
coordinates(test_e) <- ~ X + Y
crs(test_e) <- CRS('+init=EPSG:31259')

#read predictions invkos
i_pred = "C:/Mapping/FIN2/i"
files <- list.files(path = i_pred, pattern = "tif$",
                    full.names = TRUE)
ipred <- stack(files)
crs(ipred) <- CRS('+init=EPSG:31259')

#read predicitons eBod
e_pred = "C:/Mapping/FIN2/e"
e_pred = "C:/Mapping/Results_final"
files2 <- list.files(path = e_pred, pattern = "tif$",
                     full.names = TRUE)
epred <- stack(files2)
crs(epred) <- CRS('+init=EPSG:31259')
crs(ipred) <- CRS('+init=EPSG:31259')

#Statistic Dataframe
stat = data.frame(row.names = c("ME","MAE","MSE","RMSE","AVE"))

#calc validation
#invekos
for (name in names(ipred)){
  # extract points for SSOC,Csat,Cdef

  test_i <- extract(x = ipred[[name]], y = test_i, sp = TRUE)
```

```r
    #alte variante:    stat[[paste("PE",name ,sep = ".")]] <- test_i[[name]]
- test_i$Cdef

    # prediction error
    if (str_contains(name, "SSOC", ignore.case = FALSE, logic = NULL, switch
= FALSE)){

      #PE
      test_i[[paste("PE",name ,sep = ".")]] <- test_i[[name]] - test_i$SSOC
      #ME
      stat[1,name] = mean(test_i[[paste("PE",name ,sep = ".")]], na.rm=TRUE)
      #MAE
      stat[2,name] = mean(abs(test_i[[paste("PE",name ,sep = ".")]]),
na.rm=TRUE)
      #MSE
      stat[3,name] = mean(test_i[[paste("PE",name ,sep = ".")]]^2,
na.rm=TRUE)
      #RMSE
      stat[4,name] = sqrt(sum(test_i[[paste("PE",name ,sep = ".")]]^2,
na.rm=TRUE) /
                              length(test_i[[paste("PE",name ,sep = ".")]]))
      #AVE/ r2
      stat[5,name] = 1 - sum(test_i[[paste("PE",name ,sep = ".")]]^2,
na.rm=TRUE) /
        sum( (test_i$SSOC - mean(test_i$SSOC, na.rm = TRUE))^2,na.rm = TRUE)

      #adj-r2
      #stat[6,name] = 1 - (n - 1) /(n - (k +1)) * (1- stat[5,name])




    }
    else if (str_contains(name, "Cdef", ignore.case = FALSE, logic = NULL,
switch = FALSE)){

      #PE
      test_i[[paste("PE",name ,sep = ".")]] <- test_i[[name]] - test_i$Cdef
      #ME
      stat[1,name] = mean(test_i[[paste("PE",name ,sep = ".")]], na.rm=TRUE)
      #MAE
      stat[2,name] = mean(abs(test_i[[paste("PE",name ,sep = ".")]]),
na.rm=TRUE)
      #MSE
      stat[3,name] = mean(test_i[[paste("PE",name ,sep = ".")]]^2,
na.rm=TRUE)
```

```r
    #RMSE
    stat[4,name] = sqrt(sum(test_i[[paste("PE",name ,sep = ".")]]^2,
na.rm=TRUE) /
                          length(test_i[[paste("PE",name ,sep = ".")]]))
    #AVE
    stat[5,name] = 1 - sum(test_i[[paste("PE",name ,sep = ".")]]^2,
na.rm=TRUE) /
      sum( (test_i$Cdef - mean(test_i$Cdef, na.rm = TRUE))^2,na.rm = TRUE)

  }
  else if (str_contains(name, "Csat", ignore.case = FALSE, logic = NULL,
switch = FALSE)){

    #PE
    test_i[[paste("PE",name ,sep = ".")]] <- test_i[[name]] - test_i$Csat
    #ME
    stat[1,name] = mean(test_i[[paste("PE",name ,sep = ".")]], na.rm=TRUE)
    #MAE
    stat[2,name] = mean(abs(test_i[[paste("PE",name ,sep = ".")]]),
na.rm=TRUE)
    #MSE
    stat[3,name] = mean(test_i[[paste("PE",name ,sep = ".")]]^2,
na.rm=TRUE)
    #RMSE
    stat[4,name] = sqrt(sum(test_i[[paste("PE",name ,sep = ".")]]^2,
na.rm=TRUE) /
                          length(test_i[[paste("PE",name ,sep = ".")]]))
    #AVE
    stat[5,name] = 1 - sum(test_i[[paste("PE",name ,sep = ".")]]^2,
na.rm=TRUE) /
      sum( (test_i$Csat - mean(test_i$Csat, na.rm = TRUE))^2,na.rm = TRUE)
  }
}

#eBod
for (name in names(epred)){
  # extract points for SSOC,Csat,Cdef

  test_e <- extract(x = epred[[name]], y = test_e, sp = TRUE)

  #alte variante:   stat[[paste("PE",name ,sep = ".")]] <- test_i[[name]]
- test_i$Cdef

  # prediction error
  if (str_contains(name, "SSOC", ignore.case = FALSE, logic = NULL, switch
= FALSE)){
```

60

```r
    #PE
    test_e[[paste("PE",name ,sep = ".")]] <- test_e[[name]] - test_e$SSOC
    #ME
    stat[1,name] = mean(test_e[[paste("PE",name ,sep = ".")]], na.rm=TRUE)
    #MAE
    stat[2,name] = mean(abs(test_e[[paste("PE",name ,sep = ".")]]),
na.rm=TRUE)
    #MSE
    stat[3,name] = mean(test_e[[paste("PE",name ,sep = ".")]]^2,
na.rm=TRUE)
    #RMSE
    stat[4,name] = sqrt(sum(test_e[[paste("PE",name ,sep = ".")]]^2,
na.rm=TRUE) /
                          length(test_e[[paste("PE",name ,sep = ".")]]))
    #AVE
    stat[5,name] = 1 - sum(test_e[[paste("PE",name ,sep = ".")]]^2,
na.rm=TRUE) /
      sum( (test_e$SSOC - mean(test_e$SSOC, na.rm = TRUE))^2,na.rm = TRUE)

  }
  else if (str_contains(name, "Cdef", ignore.case = FALSE, logic = NULL,
switch = FALSE)){

    #PE
    test_e[[paste("PE",name ,sep = ".")]] <- test_e[[name]] - test_e$Cdef
    #ME
    stat[1,name] = mean(test_e[[paste("PE",name ,sep = ".")]], na.rm=TRUE)
    #MAE
    stat[2,name] = mean(abs(test_e[[paste("PE",name ,sep = ".")]]),
na.rm=TRUE)
    #MSE
    stat[3,name] = mean(test_e[[paste("PE",name ,sep = ".")]]^2,
na.rm=TRUE)
    #RMSE
    stat[4,name] = sqrt(sum(test_e[[paste("PE",name ,sep = ".")]]^2,
na.rm=TRUE) /
                          length(test_e[[paste("PE",name ,sep = ".")]]))
    #AVE
    stat[5,name] = 1 - sum(test_e[[paste("PE",name ,sep = ".")]]^2,
na.rm=TRUE) /
      sum( (test_e$Cdef - mean(test_e$Cdef, na.rm = TRUE))^2,na.rm = TRUE)

  }
  else if (str_contains(name, "Csat", ignore.case = FALSE, logic = NULL,
switch = FALSE)){
```

```r
    #PE
    test_e[[paste("PE",name ,sep = ".")]] <- test_e[[name]] - test_e$Csat
    #ME
    stat[1,name] = mean(test_e[[paste("PE",name ,sep = ".")]], na.rm=TRUE)
    #MAE
    stat[2,name] = mean(abs(test_e[[paste("PE",name ,sep = ".")]]),
na.rm=TRUE)
    #MSE
    stat[3,name] = mean(test_e[[paste("PE",name ,sep = ".")]]^2,
na.rm=TRUE)
    #RMSE
    stat[4,name] = sqrt(sum(test_e[[paste("PE",name ,sep = ".")]]^2,
na.rm=TRUE) /
                            length(test_e[[paste("PE",name ,sep = ".")]]))
    #AVE
    stat[5,name] = 1 - sum(test_e[[paste("PE",name ,sep = ".")]]^2,
na.rm=TRUE) /
       sum( (test_e$Csat - mean(test_e$Csat, na.rm = TRUE))^2,na.rm = TRUE)
  }
}

#Show result
stat

#Save Results
write.csv(stat, "F:/BokuDrive/Seafile/Meine
Bibliothek/Masterarbeit/Results/Statistik/rf_base.csv", row.names = T)
write.xlsx(stat, file = "F:/BokuDrive/Seafile/Meine
Bibliothek/Masterarbeit/Results/Statistik/rf_test.xlsx", sheetName=name,
           col.names=TRUE, row.names=TRUE, append=TRUE, showNA=TRUE,
password=NULL)
```

8.2 Maps

Stable soil organic carbon Map of Lower Austria for Agricultural and Grassland top soils 0-20cm (1990/91)

Universität für Bodenkultur Wien
University of Natural Resources
and Life Sciences, Vienna

SSOC g/kg

- <= 10
- 10 - 12.5
- 12.5 - 15
- 15 - 17.5
- 17.5 - 20
- 20 - 22.5
- 22.5 - 25
- > 25

Modell: Random Forest
R² : 0,31
Mean Average Error: 4,46
Root Mean Squared Error: 10,36

Data Source:
Soil Inventory of Lower Austria
1990/91

Authors: F.J.K. Mayr & Walter W.W. Wenzel, 2020 ©

N

0   50   100 km
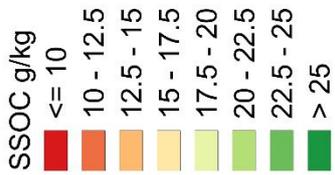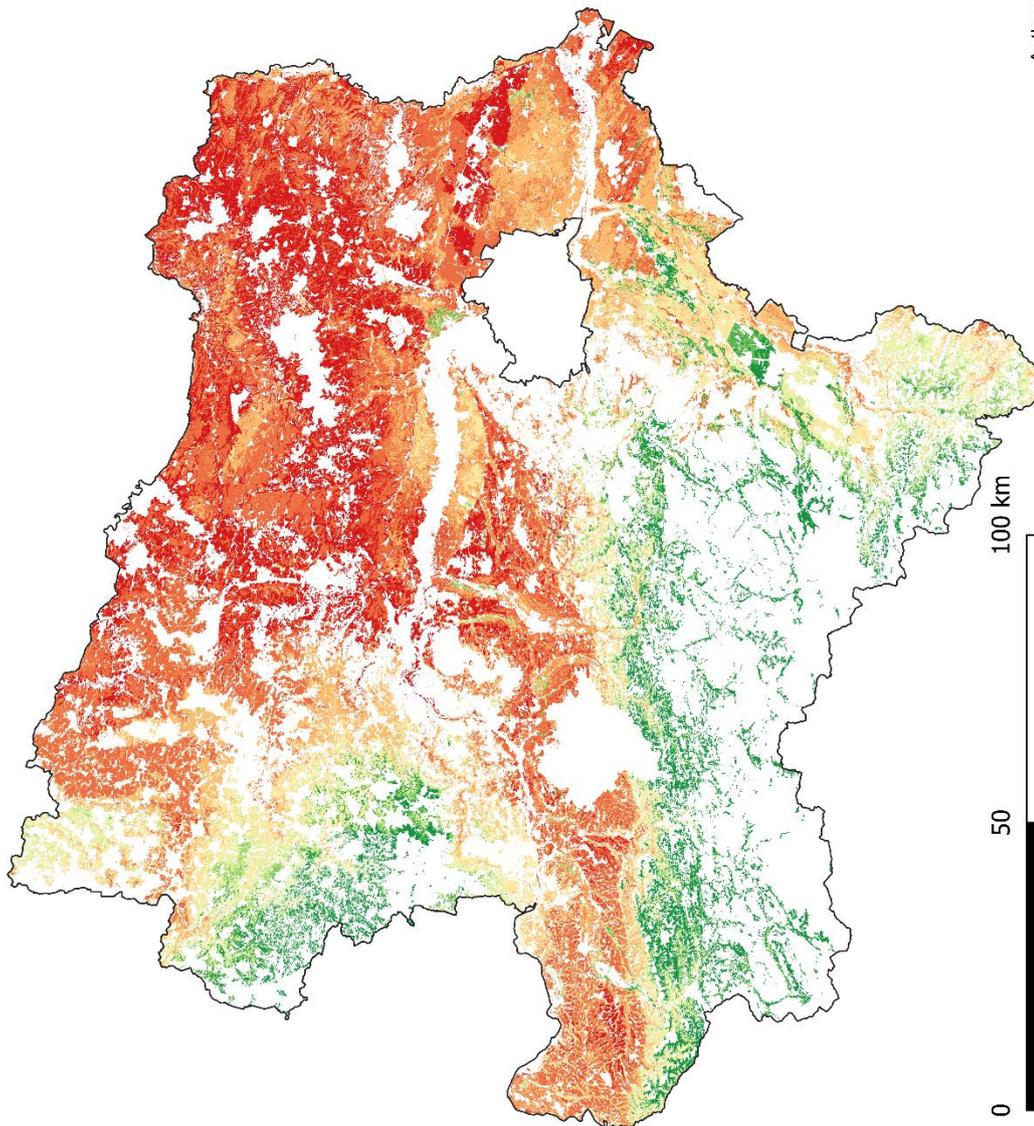
**Carbon saturation potential Map of Lower Austria for
Agricultural and Grassland top soils 0-20cm (1990/91)**

N

Csat g/kg

<= 25
25 - 30
30 - 35
35 - 40
40 - 45
45 - 50
> 50

Universität für Bodenkultur Wien
University of Natural Resources
and Life Sciences, Vienna

Modell: Random Forest
R²: 0.58
Mean Average Error: 5,35
Root Mean Squared Error: 6,95

Data Source:
Soil Inventory of Lower Austria
1990/91

Authors: F.J.K. Mayr & Walter W.W. Wenzel, 2020 ©

0   50   100 km

Carbon deficit Map of Lower Austria for Agricultural and Grassland top soils 0-20cm (1990/91)

Universität für Bodenkultur Wien
University of Natural Resources
and Life Sciences, Vienna

Cdef g/kg

<= 10
10 - 15
15 - 20
20 - 25
25 - 30
30 - 35
35 - 45

Modell: Random Forest
R²: 0.52
Mean Average Error: 6,77
Root Mean Squared Error: 11.4

Data Source:
Soil Inventory of Lower Austria
1990/91

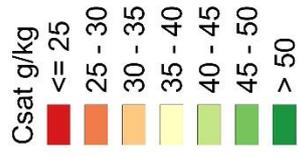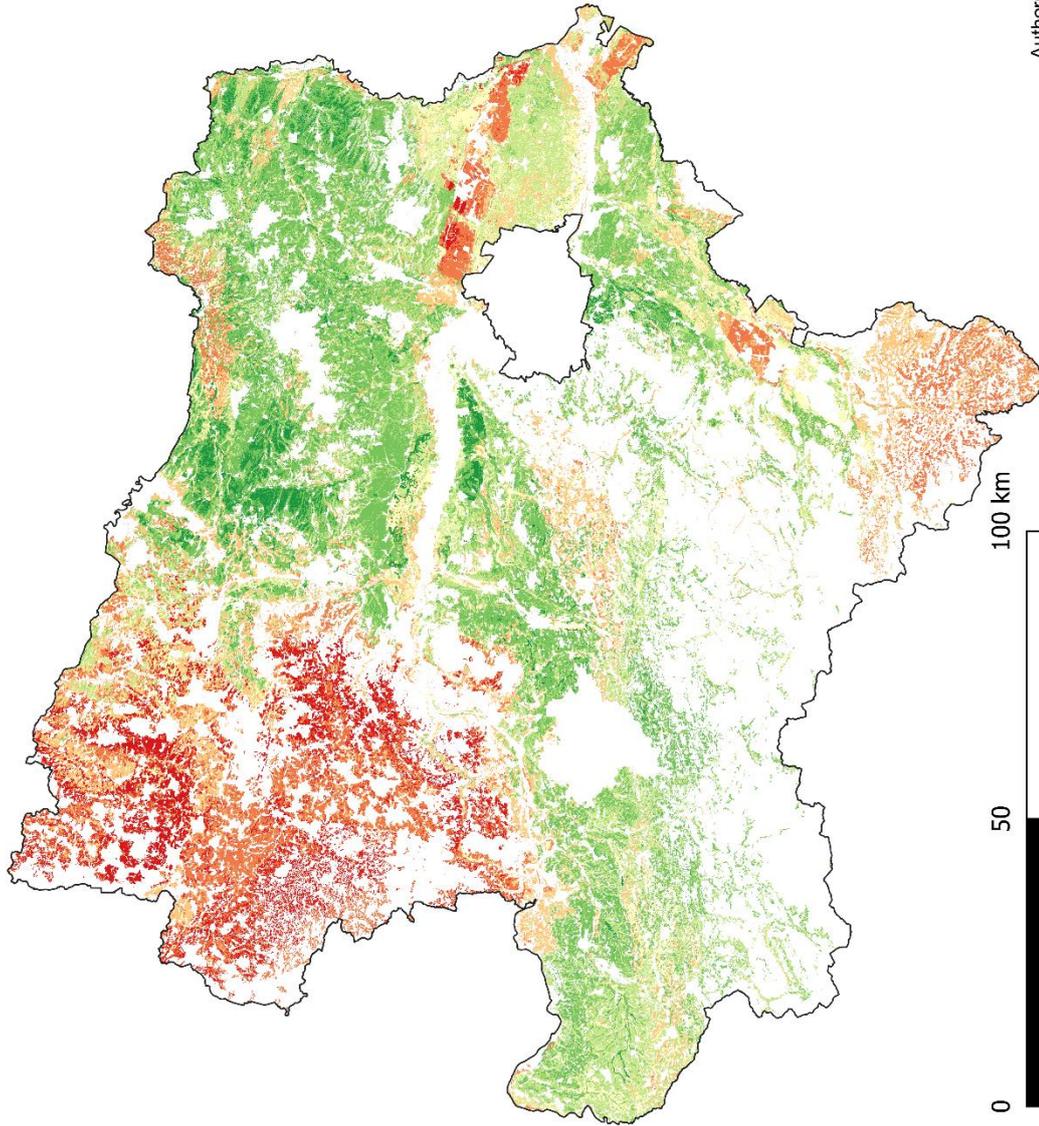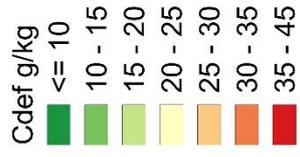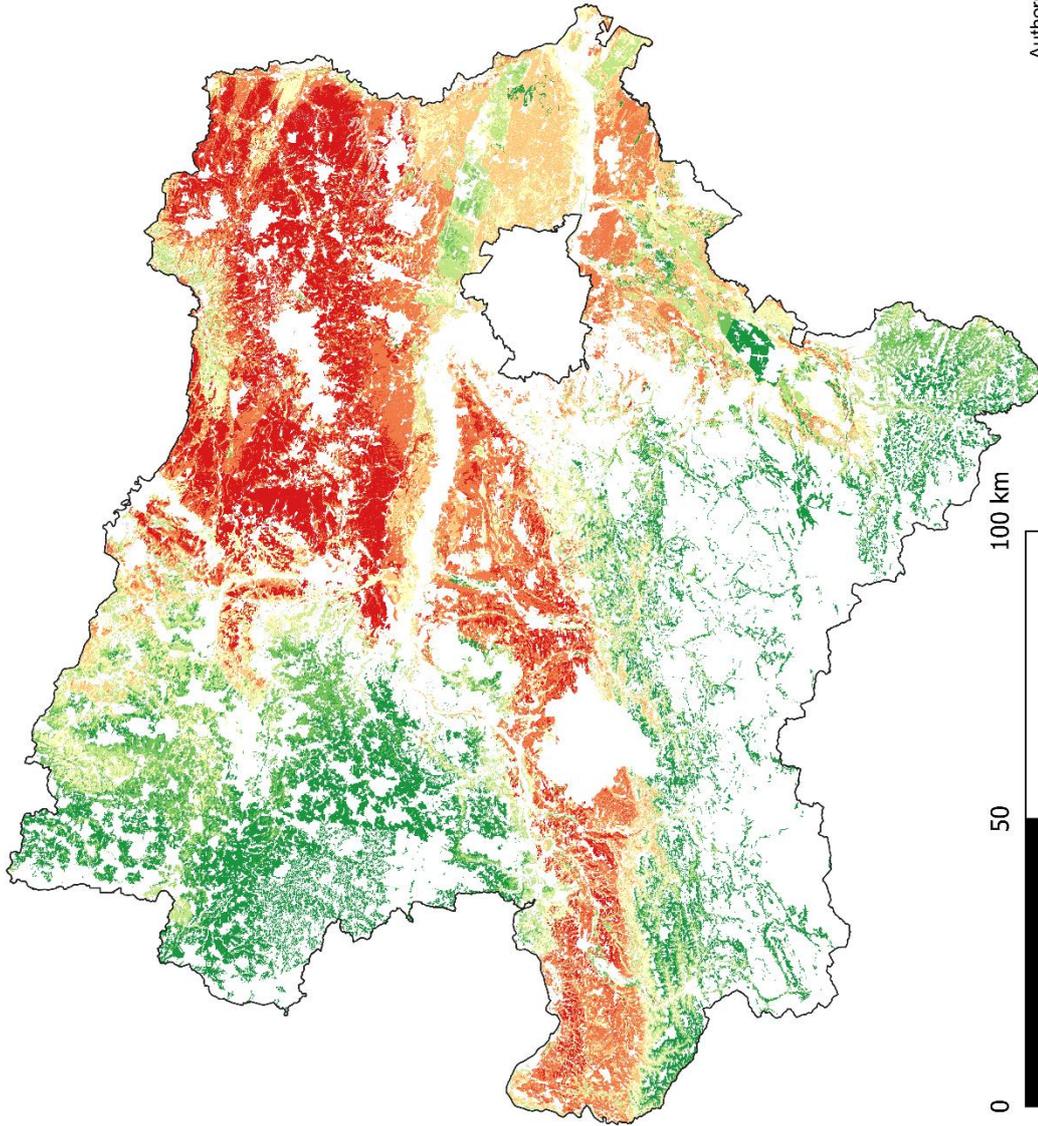Authors: F.J.K. Mayr & Walter W.W. Wenzel, 2020 ©

N

0   50   100 km

8.3 Variable Importance

SSOC Variable Importance		Csat Variable Importance		Cdef Variable Importance	
Covariate	*%*	*Covariate*	*%*	*Covariate*	*%*
precip	100	sand	100	dem	100
dem	76.5527268	dem	60.574723	sand	95.325731
somvalue5	75.0023404	clay	60.136600	temp	84.676963
rough	58.6292003	temp	58.048935	precip	64.780144
clay	57.3136506	pH	56.740718	clay	59.642449
temp	55.657249	precip	56.645336	somvalue5	57.530740
somvalue4	55.1831137	silt	45.881634	silt	55.918255
sand	47.8589583	rough	42.945729	pH	52.422782
lime	44.1774668	lime	34.350070	radi	52.098749
twi	44.0923748	texturclasses3	33.013253	lime	50.155368
somvalue3	42.4906863	twi	31.456931	bedrock7100	50.106341
pH	39.9232541	permeability7	28.928916	rough	46.959214
radi	36.3426845	sombalance4	28.0252	permeability7	44.057502
somvalue2	35.829687	texturclasses1	27.97583	permeability3	42.24065
soiltype4	35.1515577	soiltype12	26.076351	clc211	40.667341
soiltype2	34.2336632	bedrock7100	25.981095	somvalue4	39.469574
silt	32.4829002	bedrock5400	25.668579	soiltype14	38.372046
clc231	31.7315555	clc242	24.971506	clc231	38.352772
clc211	31.3523299	sombalance5	24.589005	somvalue3	36.907542
soiltype10	30.3478505	clc211	24.101214	texturclasses3	36.09729
aspect	28.8803555	permeability3	23.825894	aspect	36.047535
bedrock4240	26.334034	radi	23.523219	bedrock4240	35.723104
sombalance6	25.3846964	permeability9	22.052794	sombalance4	35.578854
texturclasses1	24.2553683	fieldcap2	21.553936	twi	34.18112
bedrock7100	23.6263905	texturclasses10	20.923392	soiltype4	34.068285
permeability3	23.3970099	permeability5	20.688719	bedrock5610	32.863101
permeability5	23.0643749	soiltype4	20.64484	soiltype30	32.48738
sombalance5	22.8100819	soiltype14	20.577035	texturclasses1	32.377113
soiltype15	22.7633746	soiltype28	20.535354	bedrock5000	31.535925
bedrock1410	22.2790604	aspect	19.296865	somvalue2	31.304935
sombalance2	21.8196703	bedrock5210	19.186024	permeability6	30.661482
soiltype12	21.6918904	somvalue1	19.153972	bedrock5210	30.61894
soiltype35	21.6479116	soiltype2	19.06918	permeability9	30.146665
fieldcap3	21.4737695	bedrock5820	18.914592	bedrock3100	28.826493
soiltype17	21.2049308	bedrock5610	18.688512	clc243	28.423352
sombalance4	21.203624	soiltype7	18.545285	fieldcap2	27.446608
permeability7	21.1435029	somvalue2	18.486003	sombalance6	27.380504
soiltype7	20.9327907	soiltype35	18.288604	bedrock5310	27.29065
bedrock5610	20.9117419	soiltype26	18.084433	sombalance2	27.035561
texturclasses4	20.8745154	sombalance2	17.853002	soiltype7	26.998145
bedrock5400	20.6377945	soiltype15	17.622328	clc112	26.783182
texturclasses3	20.4620854	bedrock3	17.614137	sombalance5	25.454971
clc242	20.0886464	bedrock3100	17.36668	clc313	25.170467
curvature	20.0286849	bedrock6	17.362765	bedrock2140	24.825679

clc313	19.7681373	fieldcap3	17.299363	fieldcap1	24.605976
fieldcap4	19.3354911	bedrock15	17.269519	soiltype6	24.388815
somvalue1	18.2677439	texturclasses2	17.145539	bedrock1210	24.212217
soiltype5	17.6507635	bedrock2113	16.446683	bedrock1410	23.849084
texturclasses7	17.5222678	sombalance3	16.358505	soiltype2	23.599985
fieldcap2	16.599699	texturclasses5	16.196514	soiltype12	23.469525
soiltype11	16.5230704	clc231	16.177492	curvature	23.164825
bedrock1000	15.4518973	bedrock9	16.097436	bedrock5820	22.479188
sombalance1	15.3679716	fieldcap1	16.070997	soiltype35	22.440647
texturclasses5	15.2334883	texturclasses7	15.64931	sombalance3	22.351431
fieldcap1	15.2220272	curvature	15.523229	soiltype13	22.262779
soiltype14	14.52349	bedrock2140	15.479385	soiltype26	21.893032
clc112	14.5200418	bedrock1000	15.44781	clc242	21.849166
bedrock3	14.3985351	bedrock1210	15.107576	clc312	21.796239
bedrock5820	14.3659871	somvalue4	14.995406	bedrock9	21.693688
bedrock1211	14.0806248	bedrock2111	14.829097	bedrock2113	21.405996
bedrock6100	14.072025	soiltype9	14.78957	fieldcap3	21.343849
permeability1	14.041231	clc313	14.741663	texturclasses11	21.125361
bedrock2100	13.9942559	fieldcap4	14.357204	bedrock1211	21.051244
bedrock5	13.8907964	clc221	14.051359	somvalue15	20.856966
soiltype19	13.456461	bedrock2100	13.998573	texturclasses5	20.826818
bedrock2111	12.9754548	bedrock5000	13.877387	bedrock5612	20.765417
soiltype27	12.9411838	somvalue5	13.845203	soiltype19	19.446793
bedrock3110	12.5944349	sombalance6	13.617069	bedrock15	19.438728
soiltype25	12.4497353	bedrock5310	13.465891	soiltype25	19.183113
soiltype28	12.3633841	soiltype17	13.444028	soiltype28	19.172648
soiltype3	12.2194426	bedrock5	13.383443	texturclasses10	18.989427
texturclasses6	12.1806122	bedrock1211	13.376805	texturclasses2	18.949904
sombalance3	11.8403703	soiltype1	13.333504	soiltype10	18.937924
bedrock3100	11.8218281	bedrock5612	13.141395	permeability5	18.767024
soiltype9	11.7810801	clc312	13.00443	texturclasses7	18.678153
permeability4	11.4513707	bedrock5300	12.943579	bedrock3	18.510044
bedrock2140	11.4026516	soiltype3	12.807795	soiltype8	18.330142
bedrock2113	11.1246568	bedrock0	12.730327	bedrock7	17.673571
bedrock5000	11.0191549	bedrock1	12.730327	texturclasses4	17.496248
bedrock15	10.9813802	bedrock10	12.730327	bedrock0	17.333647
soiltype1	10.5960304	bedrock11	12.730327	bedrock1	17.333647
bedrock0	10.0710451	bedrock1120	12.730327	bedrock10	17.333647
bedrock1	10.0710451	bedrock12	12.730327	bedrock11	17.333647
bedrock10	10.0710451	bedrock1310	12.730327	bedrock1120	17.333647
bedrock11	10.0710451	bedrock14	12.730327	bedrock12	17.333647
bedrock1120	10.0710451	bedrock2120	12.730327	bedrock1310	17.333647
bedrock12	10.0710451	bedrock4	12.730327	bedrock14	17.333647
bedrock1310	10.0710451	bedrock4230	12.730327	bedrock2120	17.333647
bedrock14	10.0710451	bedrock5500	12.730327	bedrock4	17.333647
bedrock2120	10.0710451	bedrock7110	12.730327	bedrock4230	17.333647
bedrock4	10.0710451	bedrock8	12.730327	bedrock5500	17.333647

67

bedrock4230	10.0710451	bedrock8100	12.730327	bedrock7110	17.333647
bedrock5500	10.0710451	clc111	12.730327	bedrock8	17.333647
bedrock7110	10.0710451	clc121	12.730327	bedrock8100	17.333647
bedrock8	10.0710451	clc122	12.730327	clc111	17.333647
bedrock8100	10.0710451	clc124	12.730327	clc121	17.333647
clc111	10.0710451	clc131	12.730327	clc122	17.333647
clc121	10.0710451	clc132	12.730327	clc124	17.333647
clc122	10.0710451	clc141	12.730327	clc131	17.333647
clc124	10.0710451	clc142	12.730327	clc132	17.333647
clc131	10.0710451	clc321	12.730327	clc141	17.333647
clc132	10.0710451	clc324	12.730327	clc142	17.333647
clc141	10.0710451	clc333	12.730327	clc321	17.333647
clc142	10.0710451	clc411	12.730327	clc324	17.333647
clc321	10.0710451	clc511	12.730327	clc333	17.333647
clc324	10.0710451	clc512	12.730327	clc411	17.333647
clc333	10.0710451	permeability0	12.730327	clc511	17.333647
clc411	10.0710451	permeability2	12.730327	clc512	17.333647
clc511	10.0710451	permeability8	12.730327	permeability0	17.333647
clc512	10.0710451	soiltype16	12.730327	permeability2	17.333647
permeability0	10.0710451	soiltype18	12.730327	permeability8	17.333647
permeability2	10.0710451	soiltype20	12.730327	soiltype16	17.333647
permeability8	10.0710451	soiltype21	12.730327	soiltype18	17.333647
soiltype16	10.0710451	soiltype22	12.730327	soiltype20	17.333647
soiltype18	10.0710451	soiltype24	12.730327	soiltype21	17.333647
soiltype20	10.0710451	soiltype25	12.730327	soiltype22	17.333647
soiltype21	10.0710451	soiltype29	12.730327	soiltype24	17.333647
soiltype22	10.0710451	soiltype31	12.730327	soiltype29	17.333647
soiltype23	10.0710451	soiltype32	12.730327	soiltype31	17.333647
soiltype24	10.0710451	soiltype33	12.730327	soiltype32	17.333647
soiltype29	10.0710451	soiltype34	12.730327	soiltype33	17.333647
soiltype31	10.0710451	soiltype36	12.730327	soiltype34	17.333647
soiltype32	10.0710451	soiltype37	12.730327	soiltype36	17.333647
soiltype33	10.0710451	soiltype38	12.730327	soiltype37	17.333647
soiltype34	10.0710451	soiltype39	12.730327	soiltype38	17.333647
soiltype36	10.0710451	soiltype40	12.730327	soiltype39	17.333647
soiltype37	10.0710451	soiltype41	12.730327	soiltype40	17.333647
soiltype38	10.0710451	soiltype42	12.730327	soiltype41	17.333647
soiltype39	10.0710451	soiltype43	12.730327	soiltype42	17.333647
soiltype40	10.0710451	soiltype44	12.730327	soiltype43	17.333647
soiltype41	10.0710451	soiltype45	12.730327	soiltype44	17.333647
soiltype42	10.0710451	soiltype46	12.730327	soiltype45	17.333647
soiltype43	10.0710451	soiltype47	12.730327	soiltype46	17.333647
soiltype44	10.0710451	soiltype48	12.730327	soiltype47	17.333647
soiltype45	10.0710451	somvalue0	12.730327	soiltype48	17.333647
soiltype46	10.0710451	somvalue15	12.730327	somvalue0	17.333647
soiltype47	10.0710451	soiltype5	12.249146	soiltype5	16.933658
soiltype48	10.0710451	bedrock1410	12.19096	bedrock5400	16.80058

68

somvalue0	10.0710451	texturclasses11	12.165626	soiltype3	15.48858
somvalue15	10.0710451	clc112	11.630711	bedrock3110	15.280317
clc311	10.0643601	soiltype19	11.415358	soiltype17	15.240806
texturclasses8	9.6921596	clc311	11.303383	bedrock6	14.994646
soiltype13	9.3938525	soiltype8	11.303021	clc311	14.670504
texturclasses10	9.3582713	bedrock3110	11.099981	bedrock2111	13.946986
permeability6	9.1878554	bedrock7	10.793483	bedrock1000	13.810327
bedrock5210	8.8831298	somvalue3	10.78154	soiltype23	13.810327
texturclasses11	8.8154082	soiltype23	10.460309	soiltype15	13.502736
permeability9	8.7680095	soiltype6	10.301763	fieldcap4	12.775681
bedrock5300	8.1178966	bedrock6100	9.972388	soiltype1	12.465177
soiltype30	8.0715253	permeability1	9.605501	soiltype11	12.228605
texturclasses2	6.6774434	soiltype27	9.517494	soiltype27	11.843831
bedrock1210	6.5322299	sombalance1	9.278922	bedrock4120	11.706976
clc221	6.475714	soiltype13	9.242461	permeability1	11.611292
bedrock5310	6.4094464	soiltype30	8.968657	clc221	11.332763
bedrock7	6.3395369	soiltype10	8.882178	bedrock5	11.091602
bedrock5612	6.0745995	soiltype11	8.403527	bedrock5300	10.885785
bedrock9	5.8658104	texturclasses4	8.048875	somvalue1	10.507573
soiltype26	4.8873496	bedrock4120	8.01893	permeability4	10.481742
clc243	4.7527929	bedrock4240	7.997258	bedrock6100	10.136586
bedrock4120	4.4687831	texturclasses6	7.523689	sombalance1	10.026569
bedrock13	3.9144303	clc243	7.358701	bedrock2100	9.876023
bedrock6	3.4770136	bedrock13	7.125124	bedrock13	9.483656
soiltype8	2.6941659	permeability4	5.019059	texturclasses6	8.708259
bedrock4700	1.4254446	texturclasses8	4.319355	bedrock4700	6.974117
soiltype6	0.5561257	bedrock4700	1.7028	soiltype9	6.648818
clc312	0	permeability6	0	texturclasses8	0