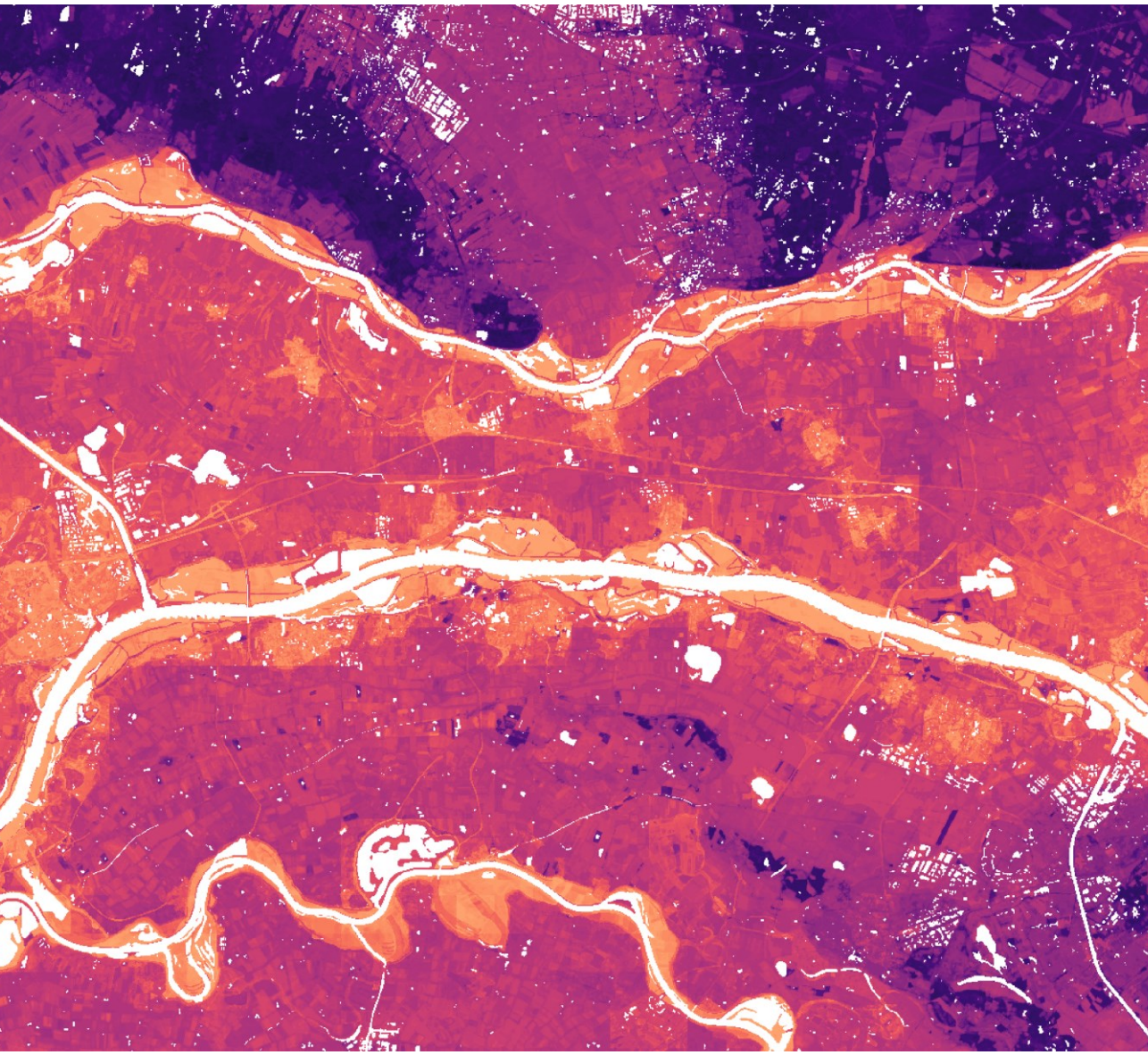
BIS-4D

A high-resolution soil modelling and mapping platform for the Netherlands in 3D space and time



Anatol Helfenstein

Propositions

1. The "ten challenges for the future of pedometrics" are as much political as they are scientific.
   (this thesis)

2. Soil maps wield power despite being erroneous representations of reality.
   (this thesis)

3. Soil carbon certificates delay climate mitigation.

4. Use precision agriculture to upscale agroecology, not conventional farming.

5. Scientific publishing companies' unprecedented profit margins result from three-fold robbing of the public.

6. Only assessing the projects, and not the partners, of university collaborations jeopardizes the university's integrity and values.

7. Privileged men should educate themselves about intersectional feminism.

8. Decolonization is crucial for a just and global energy transition.

Propositions belonging to the thesis, entitled

BIS-4D: A high-resolution soil modelling and mapping platform for the Netherlands in 3D space and time

Anatol Helfenstein
Wageningen, 11 June, 2024

BIS-4D:
A high-resolution soil modelling and mapping platform for the Netherlands in 3D space and time

Anatol Helfenstein

Thesis committee

Promotor:
Prof. Dr Gerard B.M. Heuvelink
Special Professor Pedometrics & Digital Soil Mapping
Wageningen University & Research

Co-promotors:
Dr Vera L. Mulder
Associate Professor, Soil Geography & Landscape Group
Wageningen University & Research

Dr Mirjam J.D. Hack-ten Broeke
Researcher & team leader, Soil, Water & Land Use Team
Wageningen University & Research

Other members:
Prof. Dr Ioannis Athanasiadis, Wageningen University & Research
Prof. Dr Mogens H. Greve, Aarhus University, Denmark
Dr Madlene Nussbaum, Utrecht University
Dr Gera van Os, Aeres University of Applied Sciences, Dronten

This research was conducted under the auspices of the C.T. de Wit Graduate School of Production Ecology & Resource Conservation (PE&RC)

BIS-4D:
A high-resolution soil modelling and mapping platform for the Netherlands in 3D space and time

Anatol Helfenstein

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr C. Kroeze,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 11 June 2024
at 4 p.m. in the Omnia Auditorium.

Anatol Helfenstein
BIS-4D: A high-resolution soil modelling and mapping platform for the Netherlands in 3D space and time
196 pages

PhD thesis, Wageningen University, Wageningen, the Netherlands (2024)
With references, with summary in English

DOI https://doi.org/10.18174/658444

Contents

Page

Contents v

Chapter 1 General introduction 1
- 1.1 Background . . . 2
  - 1.1.1 A brief history of soil science: a spatial perspective . . . 2
  - 1.1.2 Soil mapping . . . 3
- 1.2 GlobalSoilMap . . . 7
- 1.3 Soils and soil maps in the Netherlands . . . 8
  - 1.3.1 Soil geography in the Netherlands . . . 8
  - 1.3.2 Soil surveys and maps in the Netherlands . . . 10
- 1.4 Knowledge gaps and research challenges . . . 10
  - 1.4.1 Mapping soil properties at high-resolution using machine learning . . . 10
  - 1.4.2 Assessing map accuracy . . . 12
  - 1.4.3 Mapping in 3D . . . 13
  - 1.4.4 Space-time mapping in 3D+T . . . 14
  - 1.4.5 Operational mapping platforms for delivering soil information . . . 17
- 1.5 Problem definition . . . 18
- 1.6 Aim, objectives and research questions . . . 19
- 1.7 Thesis structure . . . 20

Chapter 2 Tier 4 maps of soil pH at 25 m resolution for the Netherlands 23
- 2.1 Introduction . . . 25
- 2.2 Materials and methods . . . 28
  - 2.2.1 Soil point datasets . . . 28
  - 2.2.2 Covariate selection . . . 30
  - 2.2.3 Covariate preprocessing . . . 32
  - 2.2.4 Model tuning and calibration . . . 33
  - 2.2.5 Maps of predicted soil pH, uncertainty and accuracy thresholds . . . 34
  - 2.2.6 Evaluation of map accuracy using statistical validation . . . 35
  - 2.2.7 Software and computational framework . . . 39
- 2.3 Results . . . 41
  - 2.3.1 Model tuning, calibration and variable importance . . . 41
  - 2.3.2 Soil pH maps: mean predictions . . . 41
  - 2.3.3 Soil pH maps: quantiles, PI90 and accuracy thresholds . . . 42
  - 2.3.4 Evaluation of map accuracy using statistical validation . . . 44
- 2.4 Discussion . . . 49

2.4.1 Map accuracy using statistical validation strategies 49
2.4.2 Comparison to other soil pH maps 51
2.4.3 Uncertainty using QRF 52
2.4.4 Accuracy thresholds for Tier 4 GSM maps and user applications . . 53
2.5 Conclusion 54
2.6 Appendix A. GSM accuracy thresholds 55

Chapter 3 Three-dimensional space and time mapping reveals soil organic matter decreases across anthropogenic landscapes in the Netherlands 57
3.1 Introduction 59
3.2 Results and Discussion 61
3.2.1 SOM decrease in peatlands 61
3.2.2 SOM changes in reclaimed lands due to land subsidence 63
3.2.3 Little SOM change in mineral soils 64
3.2.4 Model accuracy assessment 65
3.2.5 3D+T mapping: a new paradigm for SOM monitoring 67
3.3 Materials and Methods 72
3.3.1 Soil point data 72
3.3.2 Covariates 74
3.3.3 Overlay and regression matrix 79
3.3.4 Model selection, tuning and calibration 79
3.3.5 Variable importance 80
3.3.6 Prediction maps 80
3.3.7 SOM changes based on soil type and land use 81
3.3.8 Model accuracy assessment 81
3.4 Code and data availability 83

Chapter 4 A nature-inclusive future with healthy soils? Mapping soil organic matter in 2050 in the Netherlands 85
4.1 Introduction 87
4.2 Methods 88
4.2.1 Nature-inclusive scenario for 2050 88
4.2.2 3D+T SOM model 89
4.2.3 2050 scenario modelling 91
4.3 Results 92
4.3.1 SOM trends at the national scale 92
4.3.2 SOM trends in mineral soils 93
4.3.3 SOM trends in peatlands 95
4.3.4 Model uncertainty 95
4.4 Discussion 95
4.5 Conclusion 99

Chapter 5 BIS-4D: Mapping soil properties and their uncertainties at 25 m resolution in the Netherlands **101**
5.1 Introduction . . . 103
5.2 Materials and Methods . . . 106
5.2.1 Soil point data . . . 106
5.2.2 Covariates . . . 112
5.2.3 Model selection, tuning and calibration . . . 113
5.2.4 Variable importance . . . 115
5.2.5 Prediction maps . . . 116
5.2.6 Accuracy assessment . . . 116
5.2.7 BIS-4D updates: pH and SOM . . . 117
5.2.8 Software and computational framework . . . 117
5.3 Results and Discussion . . . 118
5.3.1 Accuracy assessment . . . 118
5.3.2 Strengths . . . 124
5.3.3 Limitations and improvements . . . 126
5.3.4 Assessment scale . . . 129
5.3.5 BIS-4D user manual . . . 130
5.4 Data and code availability . . . 131

Chapter 6 Synthesis **133**
6.1 Introduction . . . 134
6.2 Overview and implications of findings . . . 135
6.2.1 Assessing map accuracy . . . 135
6.2.2 Mapping in 3D . . . 138
6.2.3 Mapping in 3D+T . . . 139
6.2.4 3D+T mapping for scenario modelling . . . 141
6.2.5 An operational platform for high-resolution soil property mapping . 142
6.3 Future research and practical implementation . . . 144
6.3.1 Improving predictions in 3D . . . 144
6.3.2 A new paradigm: scalable soil estimation services . . . 145
6.3.3 Modern soil surveys to meet modern demands . . . 147
6.4 Conclusion . . . 149

References **151**

Summary **181**

Acknowledgements **185**

About the author **189**

PE&RC Training and Education Statement **193**

Chapter 1

General introduction

1.1 Background

Soil functions as the Earth's living skin, serving as the substrate for numerous atmospheric gases to undergo biogeochemical cycling and for water to be filtered and retained within the global hydrological cycle. It represents a vast and constantly changing storehouse of carbon and serves as the fundamental medium for much of our agricultural output. Profound shifts, driven primarily by alterations in the climate, land use and agricultural practices, will inevitably impact these interconnected roles, potentially yielding far-reaching consequences for society now and in the future. Ultimately, the manner in which we oversee the management of Earth's soil, both directly and indirectly, will influence the future viability of humankind and other species (Amundson et al., 2015).

1.1.1 A brief history of soil science: a spatial perspective

For millennia, human knowledge of soils was linked to the development of agriculture and entire civilizations, prompting efforts to understand and map soil spatial variation. Not all areas have the same degree of soil spatial variation, but the soils of an area are rarely completely homogeneous. The first known soil maps were created in China about 4000 years ago (Gong et al., 2003). The Aztecs developed a soil classification system with up to 45 classes for purposes of taxation, soil management, medicinal usage and construction (Williams, 1976; Williams & Jorge, 2008; Hartemink et al., 2013). The first scientific soil map was created by Stanisław Staszic for Eastern Europe in 1806 based on ideas and classification approaches from agrogeology (Grigelis et al., 2011). This was followed by soil maps in Germany, France, Austria, the Netherlands, Belgium, Russia and the United States in the second half of the 19$^{\text{th}}$ century (Hartemink et al., 2013). Towards the end of the 19$^{\text{th}}$ century, Vasily V. Dokuchaev and other Russian researchers began examining soil spatial variation on a continental scale, correlating differences in soil properties with climatic variation by conducting soil surveys (Dokuchaev, 1899). A soil survey is the systematic examination, description, classification and mapping of the soils in a given area (Soil Science Society of America, 2008). Dokuchaev came to the conclusion that soil was the result of five soil-forming factors: climate, organisms (including plants and humans), relief (topography), parent material (geology), and time. The ideas of Dokuchaev were gradually adopted and expanded by soil scientists elsewhere, including Hans Jenny, who described the relationship between soil properties and the five soil-forming factors using a generic mathematical relationship (Jenny, 1941):

$$s = f(cl,\ o,\ r,\ p,\ t) \tag{1.1}$$

where the magnitude of a soil property s is determined by, or is a function of (f), the climate cl, organisms o, topography r, parent material p and time t. Between the World Wars, numerous governments launched regional soil survey and mapping initiatives, par-

ticularly spurred by events like the 'dust bowl' in the 1930s. As Franklin D. Roosevelt, president of the United States, wrote in 1937: "The nation that destroys its soil destroys itself" (Roosevelt, 1937).

After World War II, governments increasingly wanted to create soil maps and soil inventories on a national scale. Soil surveys and maps were primarily aimed at enhancing agricultural output and management in support of the growing post-war population (Arrouays et al., 2021). Largely as a result of this, the Common Agricultural Policy was launched in Europe in 1962 and large-scale land use planning and soil mapping activities emerged (Arrouays et al., 2021).

Today, we acknowledge a broader array of needs concerning soil resources. While ensuring food security for an expanding global population remains paramount, we also prioritize soil health and a wide spectrum of soil-related ecosystem services, such as biodiversity preservation, water retention and purification, and the soil's potential to act as a carbon sink to mitigate climate change (Keesstra et al., 2016; Lehmann et al., 2020). With these additional ecosystem services on the agenda, the need for spatial soil information has become even more important (Arrouays et al., 2021).

The broad range of needs concerning soils are in part due to a broader range of potential uses of spatial soil information at different scales. For example, on a local scale, a farmer may want to decide which crops to grow on which fields based on the water storage and drainage capacity of the soil, among other factors. On a national scale, information on soil texture and soil organic matter (SOM) are necessary for greenhouse gas reporting of the Land Use, Land Use Change and Forestry sector for the United Nations Framework Convention on Climate Change (Arets et al., 2020). In Europe, soil information is beneficial for the successful implementation of policies and initiatives such as the Common Agricultural Policy, Zero Pollution and the Green Deal (Panagos et al., 2022b). On the global scale, soils play a crucial role for several of the United Nations Sustainable Development Goals, such as "Zero hunger" and "Life on land" (United Nations, 2015). Soil maps also serve as inputs for a variety of Earth system science models across all these scales.

1.1.2 Soil mapping

The need for spatial soil information is apparent, but it is impossible to know what the soil is like everywhere. Even if assuming that a measurement could fully capture a given soil property, only a finite number of locations can be visited to conduct such measurements. Therefore, the best option for predicting a soil property at all unvisited locations, or mapping it, is to spatially interpolate by using a model informed by data from measurement locations. However, it remains challenging to accurately predict a soil property and quantify the associated uncertainty, or error, of predictions in space. This challenge further increases when considering that soil is not two-dimensional (2D), but

also varies across depth and over time. Soil properties are variable in three-dimensional (3D) space and time (3D+T). In fact, soil often varies vertically more so than horizontally, and certainly over much shorter distances. In addition, the degree to which soil properties are dynamic mostly depends on the time-scale considered. While all soil properties are dynamic over geological time-scales, soil properties related to soil texture, for example clay content, can be considered static within the lifespan of a human. However, some chemical soil properties like SOM and pH are dynamic within human lifespans, while some hydrological soil properties such as soil moisture are dynamic even within one day.

With the establishment of national soil survey institutes and systematic soil mapping throughout the 20th century, regional and national maps of soil classes were created using soil classification systems with nomenclature designed to capture the soil spatial variation in that country. Conventional soil mapping builds on the assumption that an area can be divided into discrete units or subareas that have similar soil-forming factors and hence have similar soil properties. Thus, the result of conventional soil mapping is a polygon map that delineates these discrete units by drawing crisp boundaries between them and assigning a soil type, or soil class, to each polygon (Fig. 1.1). Conventional soil mapping makes use of external information such as aerial photographs and geological maps, but it relies most on the soil surveyor's expertise. Soil surveyors assembled existing information, for example aerial photos and geological maps, conducted a preliminary field visit, designed the field sampling campaign and collected field data via soil augering (boreholes) or soil pits, whereby pedogenetic soil horizons and soil classes were designated. Thereafter, they created a soil map by delineating borders between soil classes on a paper map (Fig. 1.1), described and sampled representative soil profiles and produced the final map and report. Field estimations were sometimes coupled with soil samples that were brought to the laboratory for more detailed analysis, which mostly involved measuring soil properties using wet chemical analysis.

Beginning in the 1990s and over the past few decades, soil mapping has revolutionized and is today mainly achieved through a process known as digital soil mapping (DSM), rather than via conventional soil mapping. DSM is the computer-assisted mapping of soil classes or soil properties, using mathematical and statistical models to infer the relationship between a response, the soil class or soil property, and the predictors, the spatially exhaustive environmental explanatory variables, also termed covariates (McBratney et al., 2003; Scull et al., 2003). The covariates are typically related to the main soil forming factors (Dokuchaev, 1899; Jenny, 1941), although other soil information besides the response, and the spatial position, or geographical space, can themselves also be used as covariates (McBratney et al., 2003). The main steps of DSM can be summarized as follows: collect georeferenced soil point data, e.g. from a database, collect and prepare covariates, create a regression matrix by extracting the covariate information at each soil point location, calibrate a regression model, predict at all locations, thereby creating a soil map and comparing observations with predictions using statistical validation (Fig. 1.2).

Figure 1.1: A conventional soil map in the making, where lines mark the boundaries between different soil classes and abbreviations indicate soil class codes. The map shows a region near Cuijk, the Netherlands along the Meuse River (Wageningen Environmental Research archives).

DSM was sparked by technological advancements, digitization of paper maps, and new, readily available geospatial data of the Earth's surface, for example using remote sensing. Thus, DSM relies on spatial databases, geographic information systems and availability of covariates. Compared to DSM, conventional soil mapping has the advantage that pedological knowledge of soil surveyors is easily incorporated. However, the advantages of DSM over conventional soil mapping are numerous: mapping is more automated, less laborious, more reproducible and easier to update and quantify map accuracy (Heuvelink et al., 2010; Kempen et al., 2009, 2012b, 2015).

DSM relies strongly on statistical and mathematical models, or the function f' in Eq. 1.1,

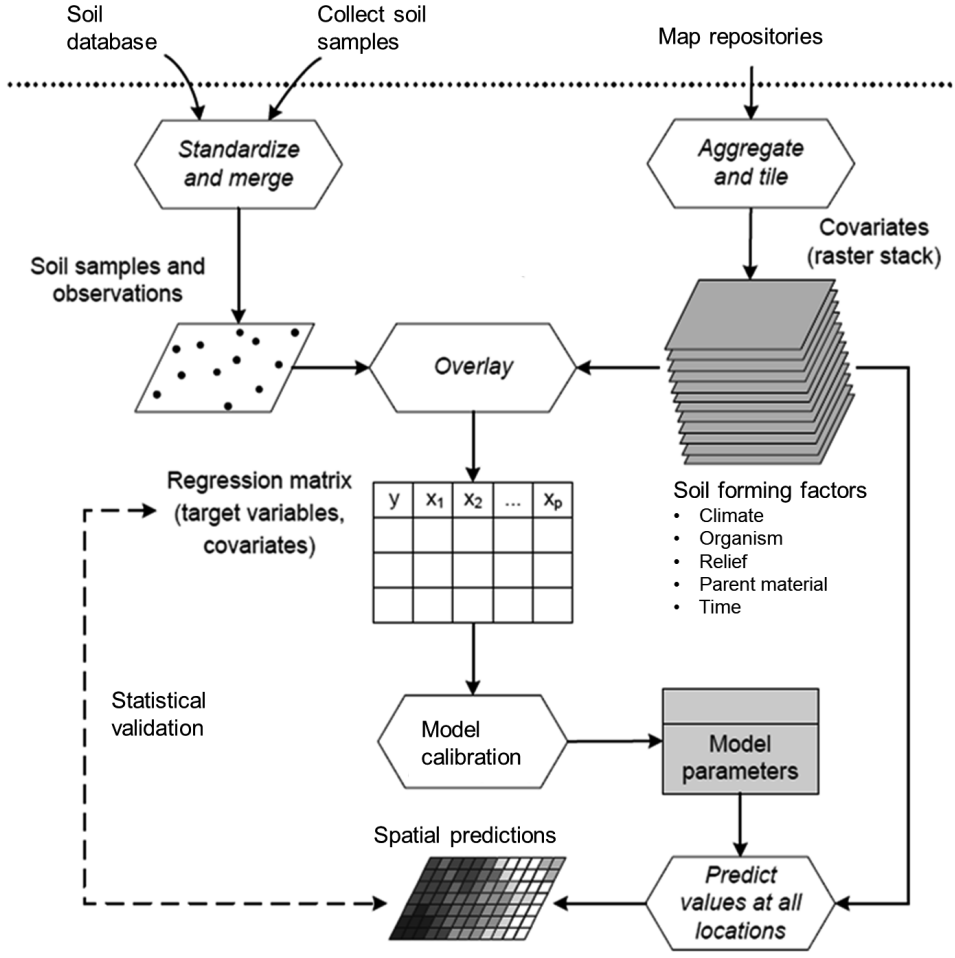


Figure 1.2: The general workflow in DSM consists of collecting georeferenced soil point data, collecting and preparing covariates, creating a regression matrix by extracting the covariate information at each soil point location, model calibration, predicting at all locations and comparing observations with predictions using statistical validation (adjusted from Hengl et al., 2017b).

wherein results are typically expressed as numerical values. This stands in contrast to conventional soil mapping, which primarily relies on qualitative descriptions mainly for mapping soil classes. While Eq. 1.1 serves as a conceptual model in conventional soil mapping, concrete mathematical functions are defined in DSM. Most DSM models use an empirical, data-driven model for f' because the goal is usually prediction accuracy, rather than model inference, or interpretation (James et al., 2021, Sect. 2.1.1). This

means that predictive modelling DSM approaches can only detect correlations, not causal relationships. To model causal relationships, a mechanistic model is needed for function f'. The most commonly used predictive models in DSM can be divided into geostatistical models, such as ordinary kriging (e.g. Burgess & Webster, 1980a) or block kriging (e.g. Burgess & Webster, 1980b), linear models (e.g. Walker et al., 1968; Jones, 1973), regression kriging (e.g. Knotters et al., 1995; Odeh et al., 1994, 1995; Goovaerts, 1999; Bishop & McBratney, 2001; Hengl et al., 2004), nonlinear models and nonlinear models plus kriging of residuals. Nonlinear models commonly use machine learning methods such as neural networks (e.g. Behrens et al., 2005), random forest (e.g. Grimm et al., 2008), support vector machine (e.g. Kovačević et al., 2010) and cubist (e.g. Viscarra Rossel et al., 2015; Miller et al., 2015). However, nonlinear models also includes Bayesian models (e.g. Steinbuch et al., 2018), Gaussian process regression (e.g. Ballabio et al., 2019) and integrated nested Laplace approximation with stochastic partial differential equation (Poggio et al., 2016; Huang et al., 2017). DSM has become a relatively common practice thanks to the increasing availability of spatial data, soil databases, computer software and algorithms. Today, DSM is a core discipline within pedometrics, the application of mathematical and statistical methods to study the distribution and genesis of soils (de Gruijter et al., 1994; McBratney et al., 2018). DSM has been widely adopted to meet the demands for accurate and high-resolution soil information for a wide range of purposes from the field to the global scale.

1.2 GlobalSoilMap

Besides technical advances, DSM has also been widely adopted thanks to the GlobalSoilMap (GSM) initiative launched in 2006 (Arrouays et al., 2014a). Due to the initiative's rapid success and arising opportunity to connect soil scientists internationally, GSM soon also formed its own working group of the International Union of Soil Sciences and was later adopted by the Global Soil Partnership of the Food and Agriculture Organization of the United Nations. GSM was formed across eight geographic nodes around the world to strive for a common goal: a high-resolution spatial soil information system of selected soil properties and their uncertainties at six standard depths for the entire world (Arrouays et al., 2014b).

Perhaps the most prominent GSM product on a global scale is SoilGrids, which provides global maps of physical, chemical and derived soil properties and soil classes developed by ISRIC – World Soil Information. SoilGrids has since its initial release (Hengl et al., 2014) been improved and updated twice (Hengl et al., 2017b; Poggio et al., 2021). Other examples of global soil maps include Global Gridded Surfaces of Selected Soil Characteristics (IGBP-DIS; Global Soil Data Task Group, 2000), WISE30sec (Batjes, 2016), S-World (Stoorvogel et al., 2017) and SoilKsatDB for soil saturated hydraulic conductivity (Gupta et al., 2021). DSM approaches have also been used to map soils at the continental scale,

for example in Africa (Hengl et al., 2015, 2017c, 2021) and Europe (Ballabio et al., 2016, 2019; Panagos et al., 2022c,a).

Much progress has also been made in mapping soils on a national scale. Examples of countries with soil maps using DSM methods include Russia (Mukhortova et al., 2021; Chinilin & Savin, 2023), China (Liang et al., 2019; Liu et al., 2020), the United States (Hempel et al., 2014b; Ramcharan et al., 2018; Chaney et al., 2019), Brazil (Gomes et al., 2019), Australia (Grundy et al., 2015; Viscarra Rossel et al., 2015), India (Dharumarajan et al., 2019, 2020), Iran (Taghizadeh-Mehrjardi et al., 2020; Zeraatpisheh et al., 2020), Nigeria (Akpa et al., 2014), Chile (Padarian et al., 2017), France (Mulder et al., 2016a,b), Scotland (Poggio & Gimona, 2014, 2017b,a) and Denmark (Adhikari et al., 2013, 2014a,b). A global soil organic carbon map (GSOC) was also created based on national soil organic carbon maps from 110 countries (Brus et al., 2017; FAO, 2018), which are maintained by the Food and Agriculture Organization of the United Nations (FAO, 2017). National DSM products also contribute to GSM because if all countries develop national soil maps then global coverage of spatial soil information can also be achieved.

1.3 Soils and soil maps in the Netherlands

1.3.1 Soil geography in the Netherlands

As elsewhere in the world, spatial soil information is also important in the Netherlands (land area = 33 481 km^2), which exhibits a number of distinct soil geographic regions (Fig. 1.3). The soils in the Netherlands are comprised of sandy soils (42%), marine clays (24%), fluvial clays and loams (8%), organic soils (15%) and soils that developed on loess deposits (1.4%) (Edelmann, 1950; Jongmans et al., 2013; Hartemink & Sonneveld, 2013). None of the soils in the Netherlands are derived from consolidated rock and more than 90% have groundwater within 150 cm of the surface during the winter (Knotters et al., 2018; Hartemink & Sonneveld, 2013). Generally, the soils in the Netherlands can be grouped by their substrate material into clay, loam and peat in the low-lying areas stemming from the Holocene epoch and sand and loess in the slightly elevated areas from the Pleistocene epoch, although peat also formed in raised bogs and brook valleys during the Holocene in the elevated areas.

Located in the midst of Europe's largest delta, the Rhine-Meuse-Scheldt delta, most soils in the low-lying areas of the Netherlands are naturally very fertile (Edelmann, 1950; Römkens & Oenema, 2004). They feature river and marine clay zones, as well as thick layers of peat that can measure several meters deep, or a mix of both clay and peat that is a result of river flooding and marine transgressions (Brouwer et al., 2018, 2023). Over centuries, drainage, excavation, and oxidation have diminished the widespread presence of peat, often leaving marine clay at the surface which was originally at the base of the peat (Erkens et al., 2016). These low-lying areas are separated from the North Sea by

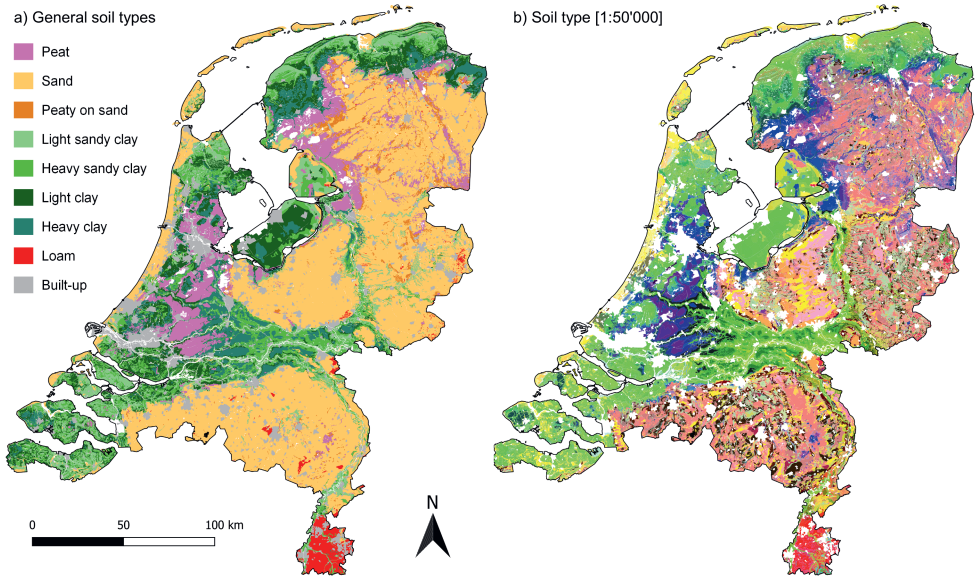


Figure 1.3: a) General soil types (Wageningen UR- Alterra, 2006) and b) the 2023 version of the national soil map of the Netherlands at 1:50 000 scale (de Vries et al., 2003). The soil map and detailed legend with all soil (sub-)classes is available at bodemdata.nl.

a relatively narrow strip of coastal sand dunes. The dunes not only protect the low country from flooding, but also provide a barrier between the salt water of the North Sea and the fresh groundwater inland. About half of the country is below sea level and would be submerged in the absence of dunes, dikes and pumping stations (AHN, 2023). Beginning with the "Beemster polder" in 1612, roughly 17% of the Dutch land area has been reclaimed from water. Most notably the entire province of Flevoland was reclaimed in the 1950s and 1960s, mainly for agricultural use.

Most areas with limited river, delta and coastal influences are marked by Pleistocene aeolian sand and glacial deposits. These marginal and typically infertile, acidic (low pH) podzol soils were utilized for agriculture in the Middle Ages, some of which over time formed into plaggen soils (Blume & Leinweber, 2004). During the early Holocene, fen peat filled the slightly undulating hills and brook valleys, while raised peat bogs were forming on the plateaus (Kempen et al., 2009, 2012a, 2015). These raised bogs were drained and excavated between the early 17th century and the mid-20th century, transforming them into intensive agricultural lands for producing mainly grain, potato, sugar beet and dairy products (Kempen et al., 2009, 2012a, 2015). The few remaining raised peat bog layers are typically less than 1.5 m thick (Kempen et al., 2015). In contrast, the aeolian loess deposits found in the southernmost and highest regions of the Netherlands contributed to the development of fertile soils predominantly composed of silt.

1.3.2 Soil surveys and maps in the Netherlands

Historically, the Netherlands was at the forefront of soil mapping. Scientific soil investigations in the Netherlands were started by Winand C.H. Staring in the mid-1800s followed by Jan van Baren and David J. Hissink in the early 1900s (Bouma & Hartemink, 2003). The first publication of the spatial distribution of soil properties in the Netherlands dates back to the 19th century (Felix, 1995). Systematic soil mapping became institutionalized with the establishment of the Dutch Soil Survey institute, or "Stichting voor Bodemkartering" (StiBoKa) in 1945 (Hartemink & Sonneveld, 2013). From 1950 to 1995, StiBoKa conducted conventional soil surveys (Buringh et al., 1962; de Bakker & Schelling, 1989; ten Cate et al., 1995) and produced regional maps (1:10 000 and 1:25 000 scale) and a national map (1:50 000 scale) of soil classes (Fig. 1.3b; de Vries et al., 2003).

After the development of DSM as a research field, various studies used (geo-)statistical methods to update several regions of the national soil map (Kempen et al., 2009, 2011, 2012a, 2014; de Vries et al., 2014, 2017, 2018; Brouwer et al., 2018; Brouwer & Walvoort, 2019, 2020; Brouwer et al., 2021, 2023). In addition, a variety of thematic maps were derived, such as a map of re-worked soils (Brouwer & van der Werff, 2012), a peat thickness map (Brouwer et al., 2018), a map of soil landscapes (van Delft & Maas, 2022, 2023) and the soil physical units map of the Netherlands (BOFEK; Heinen et al., 2022).

1.4 Knowledge gaps and research challenges

1.4.1 Mapping soil properties at high-resolution using machine learning

While categorical maps of soil type and derived thematic maps are important, there is a critical gap because many users require information on specific, numerical soil properties. As one of Europe's most densely populated countries, multi-functional land use decisions made at a national or regional level need to be implemented at the field scale. This involves a broad range of diverse stakeholders from multiple sectors, such as agriculture, forestry, land and water management, spatial planning, infrastructure, transport and nature conservation. For example, spatial information of soil properties can be used to evaluate soil health on Dutch agricultural fields using tools such as the Open Soil Index (OSI; Ros et al., 2022; Ros, 2023) and Soil Indicators for Agriculture (BLN 2.0; Ros et al., 2023). In addition, data of basic soil properties serve as inputs for modelling agricultural suitability (Mulder et al., 2022), crop precision agriculture (Been et al., 2023) and Soil-Water-Atmosphere-Plant interactions (SWAP; van Dam et al., 1997; Kroes et al., 2017), the latter of which is the underlying model for the Watervision Agriculture and Nature initiative (Hack-ten Broeke et al., 2019). Furthermore, spatial soil property data is also helpful for the Hydrological Instrumentations of the Netherlands (NHI, 2023) and the Delta Program 2024 (Delta Programme, 2023).

The few existing national-scale soil property maps in the Netherlands are at a coarse resolution and do not use machine learning, which typically improves mapping accuracy when large amounts of data are available. The most recent map of SOM was at a resolution of 250 m (van den Berg et al., 2017), while the prototypes developed by Brus et al. (2009) were at 1 km resolution. The maps of Schoumans & Chardon (2015) and Knotters et al. (2022) were not in raster format, instead providing results at the scale of 94 strata based on soil type and groundwater classes (Finke et al., 2001), or for general combinations of a few soil types and land use categories. As covariates such as remote sensing products and national maps of land use (Hazeu et al., 2023) and digital elevation models (AHN, 2023) are nowadays available at 5 - 25 m resolution, useful information for modelling complex relationships between soil-forming factors and soil properties is provided at these scales. However, it is crucial to emphasize that resolution is not an indicator of accuracy (Sect. 1.4.2) and should not be used solely to determine a map's fitness for use (de Bruin et al., 2001; Malone et al., 2013; Knotters & Walvoort, 2020; Szatmári et al., 2021). Furthermore, soil mapping approaches in the Netherlands up to date used no or only a small number of covariates. For example, van den Berg et al. (2017) only used latitude, soil physical unit, soil horizon, land use category and topsoil texture in the trend component of the model. For study areas where a large amount of soil point data and high-quality covariates are available, which is the case in the Netherlands, machine learning has shown to increase prediction accuracy (e.g. Hengl et al., 2015, 2017b; Nussbaum et al., 2018; Keskin et al., 2019; Khaledian & Miller, 2020).

One pivotal soil property lacking high-resolution information at present is soil pH. Soil pH provides information on soil acidity and alkalinity, nutrient availability and affects biological activity, decomposition, metal dissolution and soil physical structure. Hence, soil pH supports land management evaluation such as liming on croplands, acidity of forest soils and the capacity to filter, bind and decompose pollutants (Thomas, 1996). However, the currently available maps of soil pH do not meet many contemporary user demands and map quality can be substantially improved. Brus et al. (2009) created a national-scale soil pH map, but only for 0 - 25 cm at 1 km resolution, whereas many users also require information below 25 cm depth and at higher resolution. In addition, Brus et al. (2009) used co-kriging and only a soil type map as a covariate, but many other high-quality covariates are now available which combined with machine learning can improve map quality. Wamelink et al. (2019) predicted soil pH based on plant species occurrence and three covariates and only in Dutch nature areas. Today, pedometricians can make use of increasing amounts of available spatial data, an extensive toolkit of geostatistical and machine learning approaches and a powerful computational infrastructure, but considerably less effort has been invested in providing appropriate measures of the soil map accuracy.

1.4.2 Assessing map accuracy

Regardless of the study area, one major research challenge in pedometrics is how to properly quantify and communicate the accuracy of soil maps. In response to an initiative within the pedometrics research community, Wadoux et al. (2021b) proposed ten key challenges in the research field of pedometrics. Two of the proposed ten challenges, challenges five and nine, relate to improving ways of quantifying and communicating soil map uncertainty.

Every map is erroneous because it is always a simplified representation of reality that suffers from multiple error sources (Burrough et al., 2015; Heuvelink, 1998, 2014). In DSM, soil maps are created through predictions from statistical or mathematical models and inevitably, these predictions are not error-free, implying that there is associated uncertainty. The uncertainty of maps can be assessed using external (model-free) and internal (model-based) accuracy measures.

Assessing map accuracy using an external approach involves comparing map predictions to independent observations, which is also referred to as statistical validation. Statistical validation usually involves data-splitting, either using a single split involving a calibration and validation set, or repeating this multiple times during n-fold cross-validation. In this way, independent and separate observations not used in model calibration are compared to the predictions at the observation locations (Chatfield, 1995).

Map uncertainty can also be quantified through internal accuracy measures. Although based on the structure and assumptions of a model, the main advantage of internal accuracy measures is that they can provide information about the accuracy at each predicted location. One way of internal accuracy assessment is by representing prediction uncertainty as probability distributions. A probability distribution has the advantage that it relies on statistical theory and enables quantitative characterization of uncertainty in great detail, such as providing information on the shape, width and magnitude of the uncertainty (Heuvelink, 2014). In geostatistical interpolation, prediction uncertainty is easily quantified in the form of the kriging variance (Goovaerts, 2001). For non-geostatistical methods, a probabilistic prediction model can be chosen, which estimates the entire conditional probability distribution of a response variable, i.e. the target soil property. An example of a commonly used probabilistic machine learning model is quantile regression forest (QRF; Meinshausen, 2006), which was first used in DSM by Vaysse & Lagacherie (2017).

Although external and internal accuracy assessment methods are available, a persisting problem is that established methods are often not used, which means that the uncertainty of soil maps is not quantified. Recent reviews showed that only 30 - 56% of studies mapping continuous soil properties estimated prediction uncertainty (Wadoux et al., 2020; Piikki et al., 2021; Chen et al., 2022). Moreover, assessing map accuracy is not straight-

forward and involves many demanding pre-requites, for example the sampling design of the locations used for statistical validation. According to Piikki et al. (2021), only 13% of studies used probability sampling for map validation, which according to sampling theory (Cochran, 1977; de Gruijter et al., 2006; Gregoire & Valentine, 2007; Brus, 2022) is the best approach for assessing map accuracy because it is unbiased (Brus et al., 2011). When using a soil map in a model or analysis, the uncertainty may be so substantial that it compromises the quality of the outputs, posing risks of erroneous conclusions and decisions for end users (Knotters & Vroon, 2015; Knotters et al., 2015a,b; Heuvelink, 2018). The efficacy of uncertainty propagation analysis relies on quantifying input uncertainty realistically, emphasizing the consistent need to quantify uncertainty in soil maps. Without providing the uncertainty of a map, users cannot determine its fitness for use.

Another challenge is to develop spatially explicit, standardized and simple ratings for end-users to compare DSM products at different scales and across the globe. In order to coordinate and guarantee a minimum quality for soil maps, specifications were made regarding the uncertainty of GSM products (Arrouays et al., 2014a; Hempel et al., 2014a). Increasing specifications must be fulfilled with increasing quality of a DSM product, which were organized into so-called *Tiers* (Arrouays et al., 2015). The simple rating system specified in these tiered GSM specifications may provide a powerful measure to communicate the quality of a soil map to end users for a specific purpose and region and make different DSM products more comparable. Although proposed, such ratings using accuracy thresholds were to my knowledge never implemented.

1.4.3 Mapping in 3D

Another challenge in DSM is developing models that can accurately predict in 3D space. Based on end-user needs, the GSM specifications stated that information is necessary at various depth layers and defined six standard prediction depths: 0 - 5 cm, 5 - 15 cm, 15 - 30 cm, 30 - 60 cm, 60 - 100 cm and 100 - 200 cm. However, half of the studies reviewed by Chen et al. (2022) focused on soil properties at less than 30 cm depth only.

With regards to dealing with soil variation over depth, DSM models can be loosely grouped into 2D, 2.5D and 3D models (Nauman & Duniway, 2019; Roudier et al., 2020; Ma et al., 2021; Chen et al., 2022). A 2D model does not account for vertical soil variation, meaning that only one soil layer is mapped. So-called 2.5D models typically first harmonize soil properties for designated depth layers by using methods such as weighted averaging or equal-area splines (e.g. Adhikari et al., 2013; Mulder et al., 2016b; Viscarra Rossel et al., 2015). Next, a separate prediction model is calibrated independently for each designated depth layer. In contrast, 3D models either use a geostatistical approach (e.g. Poggio & Gimona, 2014; Orton et al., 2016, 2020), a regression kriging approach wherein residuals are interpolated using 3D kriging (e.g. Hengl et al., 2014, 2015), predict the parameters of a depth function spatially (e.g. Meersmans et al., 2009a,b; Kempen et al., 2011; Liu et al.,

2016; Ottoy et al., 2017; Rentschler et al., 2019) or use depth as a covariate in modelling (e.g. Akpa et al., 2014; Filippi et al., 2019, 2020; Hengl et al., 2017b; Ramcharan et al., 2018; Zhang et al., 2020). In a recent review, Chen et al. (2022) found that 64% of DSM studies used 2D models and only 8% used 3D models. On a national scale in the Netherlands, most studies mapped in 2D at a single depth layer alone (Brus et al., 2009; Schoumans & Chardon, 2015; Wamelink et al., 2019).

In all modelling approaches discussed above, including 3D models, the covariates themselves were by definition 2D. Kempen et al. (2011) noted that soil type maps can be considered 3D models of soil properties, since soil types and individual pedogenetic horizons (indirectly) contain information about the vertical variation of soil properties. However, using soil type maps for 3D mapping with parametric depth functions is cumbersome, the number of depth function parameters should remain within practical limits and it requires many additional steps, full profile descriptions, expert knowledge, various assumptions and calibrating many separate models (Kempen et al., 2011). Instead, the simpler and more intuitive approach of deriving covariates that are variable in 3D space and incorporating them in a regression matrix in which 3D covariates vary depending on the sampled soil depth, has largely been ignored. To my knowledge, only Gasch et al. (2015) used 3D covariates of soil parameters to predict soil water, temperature and electrical conductivity on a field scale (37 ha). Machine learning has proven advantageous to predict complex, non-linear relationships between soil-forming factors and soil properties in 2D space (Wadoux et al., 2020), so including 3D covariates in machine learning models may improve mapping accuracy in 3D space. Furthermore, although soil properties vary continuously with depth in many areas, in regions under strong anthropogenic influence, such as the Netherlands, sharp discontinuities in the depth distribution of soil properties frequently occur (Fig. 1.4; Kempen et al., 2011). Hence, using covariates variable in 3D space may improve predictions at locations where sharp continuities occur compared to using depth as a covariate, as the latter tends to produce predictions that vary continuously over depth.

1.4.4 Space-time mapping in 3D+T

Dynamic soil properties not only vary over depth but also in space and time, i.e. in 3D+T. However, the overwhelming majority of DSM studies ignore time. This is because space-time mapping of soil properties remains a major challenge, with relatively few studies having mapped temporal changes and mostly only at the field scale (Meersmans et al., 2011, 2016; Sun et al., 2012; Stockmann et al., 2015; Gasch et al., 2015; Minasny et al., 2016; Gray & Bishop, 2016, 2019; Yigini & Panagos, 2016; Schillaci et al., 2017; Hengl et al., 2017a; Sanderman et al., 2017; Song et al., 2018; Stumpf et al., 2018; Reyes Rojas et al., 2018; Adhikari et al., 2019; Huang et al., 2019; Szatmári et al., 2019; Zhou et al., 2019; Heuvelink et al., 2020; Sun et al., 2021; OpenGeoHub et al., 2022; Yang et al., 2022). These studies have used various space-time modelling approaches. One approach

Figure 1.4: Soils under anthropogenic influence: a) deep plowing on arable land (Brouwer & van der Werff, 2012, Photo 2), b) soil profile of a plow layer on top of podzol remains (Kempen et al., 2011, Fig. 1) and c) soil profile of a deep-plowed and drained peat (organic) soil under pea cultivation (Brändli et al., 2016).

is to develop a baseline map for a reference year and modifying that map based on dynamic soil-forming factors such as land use or climate change. Another approach is to map changes between two years by developing separate DSM models for both years (Meersmans et al., 2011; Szatmári & Pásztor, 2019). Such an approach requires a sampling design that is revisited, i.e. monitoring data, which most areas do not have. Another approach is space-for-time-substitution (e.g. Adhikari et al., 2019; Yang et al., 2022), although differences in a dynamic soil property due to temporal changes are not necessarily comparable to differences due to spatial variation. Commonly, soil property changes are mapped using a static model, by simply replacing covariates such as land use or climate during prediction with a historic map or future scenario (Yigini & Panagos, 2016; Meersmans et al., 2016; Gray & Bishop, 2016, 2019; Adhikari et al., 2019; Huang et al., 2019; Reyes Rojas et al., 2018). A promising possibility is to derive dynamic covariates that cover the time period in which the soil point data were collected, thereby developing a model that is explicit in 2D+T during both calibration and prediction (Gasch et al., 2015; Heuvelink et al., 2020; OpenGeoHub et al., 2022). Another approach is to use (semi-) mechanistic models such as CENTURY (Parton et al., 1987), RothC (Coleman & Jenkinson, 1996) and Millenial (Abramoff et al., 2018, 2022) for soil organic carbon dynamics. A recent promising improvement is integrating the advantages of mechanistic models for modelling temporal dynamics and the advantages of machine learning models for spatial prediction (Zhang et al., 2024). However, this also requires point monitoring data and applications for different soil and land use types are limited.

Besides one study on the field scale (Gasch et al., 2015), to the best of my knowledge, DSM models have so far never been explicit in 3D+T. Moreover, although Gasch et al.

(2015) predicted in 3D+T, covariates were variable in 2D+T and 3D, but no covariates variable in 3D+T were used. Despite much demand for accurate soil information in 3D+T, deriving covariates that are explicit and variable in 3D+T for DSM models using machine learning has not been explored. Hence it is no surprise that one of the ten challenges in the pedometrics research community relates to developing models in 3D+T (Wadoux et al., 2021b).

In the Netherlands, with its far-reaching land surface changes in peatlands, reclaimed land and highly anthropogenic landscapes (Sect. 1.3.1), one crucial soil property subject to temporal changes is SOM. SOM enhances the availability of plant nutrients, improves moisture retention, stabilizes soil structure, increases permeability and chemical buffering and influences the biodegradability of organic molecules such as pesticides (Blume et al., 2010; Noellemeyer & Six, 2015). In fact, the most recent map of SOM in the Netherlands was specifically developed for its use in GeoPEARL, a model that assesses the leaching potential of plant protection products (van den Berg et al., 2017). However, besides other limitations (Sect. 1.4.1), the SOM map developed by van den Berg et al. (2017) was static and the SOM content and carbon stock estimates provided by Knotters et al. (2022) were at the scale of general soil type and land use classes and thus not spatially explicit. SOM has also received considerable attention due to its relevance in the global carbon cycle and climate change because it acts as a major sink and source of soil organic carbon (Amelung et al., 2020; Moinet et al., 2023; Paul et al., 2023). Most soil organic carbon is located in peatlands, or organic soils, which cover a substantially higher proportion of the Netherlands than they do in most other countries (Sect 1.3.1). Based on the Climate Agreement of the Netherlands, emissions must be reduced by 1 Mton CO_2-eq from peat soils (Nol et al., 2010; van Beek et al., 2011) and by 0.4 - 0.6 Mton CO_2-eq from mineral agricultural soils before 2030 (Government of the Netherlands, 2019). Initiatives like the "Smart Land Use" project aim to sequester an additional 0.5 Mton CO_2-eq per year to Dutch mineral agricultural soils (Slier et al., 2023). Hence, there is high demand for accurate SOM information in 3D+T.

Space-time mapping of future scenarios

Modeling future scenarios is useful for strategic planning, risk management, innovation, resource allocation, policy development, and stakeholder engagement, helping society anticipate and adapt to future uncertainties effectively. However, few future scenarios map soils, even though soil-related ecosystem services are an essential part of sustainable transitions. When coupled with future scenarios, e.g. of land use or climate change, space-time DSM models provide the tools necessary to bridge this gap. Nonetheless, few studies have used DSM for predicting dynamic soil properties in the future (Meersmans et al., 2016; Yigini & Panagos, 2016; Gray & Bishop, 2016, 2019; Reyes Rojas et al., 2018; Adhikari et al., 2019). These studies used various future land use and/or climate scenarios. However, I am not aware of studies that have used the concept of nature-based

solutions or nature-inclusive scenarios to predict soil properties in the future. This may provide a powerful tool to spatially model soil health and related ecosystem services because envisioning nature-inclusive scenarios for the future helps resolve challenges we are facing today (Keesstra et al., 2018; Sowińska-Świerkosz & García, 2022). Furthermore, for transformative change to address the challenges that negatively affect our planet, such as climate change and loss of biodiversity (IPBES, 2019; Pörtner et al., 2021), we need approaches that address the interdependent challenges in an integrated way to avoid negative trade-offs and feedbacks (Larrosa et al., 2016).

In the Netherlands, several future visions of nature-based solutions and nature-inclusive scenarios have been developed (Baptist et al., 2019; Breman et al., 2022). Baptist et al. (2019) developed a new vision for the Netherlands in 2120 through nature-based solutions in the domains of water management, energy, agriculture, circular economy, urbanisation and biodiversity. Although an image of the Netherlands was drawn, no georeferenced, spatially explicit map was developed, which is a prerequisite for using it in space-time DSM modelling. Instead, Breman et al. (2022) developed a national scenario of a nature-inclusive society for 2050 within the framework of a geographic information system. The goal of this vision was to tackle several urgent societal challenges, such as nature conservation and biodiversity, climate change, quality of living, farming transition, energy transition and water quality (Breman et al., 2022). Coupling space-time DSM models with future nature-inclusive scenarios can underscore the importance of soils, envision potential futures and prompt pertinent questions and discussions on soil health and sustainable transitions.

1.4.5 Operational mapping platforms for delivering soil information

Approximately 78 % of articles reviewed by Chen et al. (2022) with study areas across the globe mapped only three, closely related soil properties: SOM, carbon content or carbon stocks. Many other key soil properties defined by GSM (Arrouays et al., 2014a; Hempel et al., 2014a) are less frequently mapped. For example, besides soil pH and SOM, additional information such as soil texture and cation exchange capacity is also crucial for regulating plant nutrient availability and leaching for both agricultural and environmental applications. This is particularly relevant in the Netherlands (Stokstad, 2019; Erisman, 2021; Aarts & Leeuwis, 2023), as it has the highest livestock density in the EU (Eurostat, 2022, p. 32) and ranks as the world's second-largest agricultural exporter (Jukema et al., 2023). Therefore, there are many businesses specialized in optimizing fertilizer and manure applications for crop production, but also for environmental accounting. For example, an estimated 1 300 000 ha are phosphate saturated soils, where phosphate loss due to leaching exceeds ecologically tolerable limits (Römkens & Oenema, 2004). Hence, providing spatially explicit information of a broad range of soil properties is crucial to adhere to Targets 4.2 and 4.3 of the Soil Deal for Europe, which aim to reduce fertilizer use by at least 20% and reduce nutrient losses by at least 50% by 2030 (European Commission,

2021).

To fill the gap of a large variety of different soil properties in the Netherlands and elsewhere, another challenge is to develop operational, reproducible, standardised, largely automated and efficient DSM workflows. Although various countries have provided soil property maps using DSM (Sect. 1.2), the emphasis and evaluation criteria of these GSM products typically revolves around the prediction maps and not on the DSM workflow, even though the modelling workflow is one of the main advantages of DSM compared to conventional soil mapping (Sect. 1.1). Heuvelink et al. (2010) proposed storing models instead of maps, allowing for more flexibility according to user requirements and saving storage capacity, among many other advantages. However, an operational soil information system was never implemented in the Netherlands and, regardless of the area, is typically not the focus of published scientific studies. Openly accessible tutorials elucidating standard DSM workflows (e.g. Malone et al., 2017; Hengl & MacMillan, 2019; Brus et al., 2017; Brus, 2019, 2022) have undoubtedly had a big impact on the increasing use of DSM. While these are important for education and training, they are not intended for routine use in a given area. Yet standardized and reproducible DSM workflows are not only beneficial for research institutions and the DSM community as a whole. By creating reproducible, standardised, largely automated and efficient workflow, DSM can contribute to open science. Open science promotes transparency, collaboration, reproducibility, innovation, public engagement, efficiency and lead to more robust, impactful, and equitable scientific research (Lowndes et al., 2017; Ferguson et al., 2023). Thereby, open science is also linked to targets 17.6 and 17.8 of the Sustainable Development Goals, which aim to improve global cooperation and access to science, technology, and innovation, while enhancing the capacity of low- and middle-income countries through knowledge sharing. Pedometrics and DSM are not exempt from these challenges and also have a role to play.

1.5 Problem definition

To summarize the knowledge gap and research challenges previously discussed, firstly, there is a need for high resolution maps of key soil properties using machine learning methods that exploit the abundance of currently available data on a national-scale in the Netherlands. Secondly, there are many remaining challenges in DSM related to assessing map accuracy. Thirdly, predictions are needed in 3D, and for dynamic soil properties, in the past, present and future in 3D+T, but doing so using the predictive power of machine learning is largely unexplored. Lastly, a generic framework is needed to develop operational, reproducible, standardised, largely automated and efficient DSM workflows to facilitate the inclusion of soil spatial variability as a routine, on-demand and integral part of decision support systems.

1.6 Aim, objectives and research questions

The aim of this thesis is to develop a high-resolution soil modelling and mapping platform for the Netherlands in 3D+T called "BIS-4D". I have named it after the "Bodemkundig Informatie Systeem", the Dutch soil database, and since it predicts soil properties and their uncertainties in four dimensions (3D+T). I aim to fill the gap of high-resolution information of soil pH, SOM and other key soil properties on a national scale in the Netherlands, thereby also contributing to the GSM project. I further defined the following objectives and research questions related to specific research challenges:

1. Develop a high-resolution, national-scale DSM model of soil pH in 3D with improved ways of assessing map quality
   (a) How can we assess map quality using various statistical validation strategies, including design-based inference of a probability sample in 3D space?
   (b) How can we quantify and visualize prediction uncertainty in a simple, standardized and appealing way?
2. Develop a high-resolution, national-scale DSM model of SOM explicit in 3D+T
   (a) How can we derive covariates that are variable in 2D+T and 3D+T as drivers of spatio-temporal SOM dynamics?
   (b) To what extent can mapping SOM in 3D+T serve as a new paradigm for monitoring soil health?
3. Apply the high-resolution, national-scale DSM model explicit in 3D+T to map SOM in 2050 based on a nature-inclusive land use scenario (Breman et al., 2022)
   (a) To what extent is a nature-inclusive scenario for 2050 conducive to enhancing SOM-related soil health in the future?
   (b) What is the added value and what are potential applications of using DSM for modelling future scenarios?
4. Develop an operational, reproducible, standardised, largely automated and efficient workflow for modelling and mapping a broad range of soil properties and their uncertainties at high-resolution on a national scale
   (a) What are the strengths, limitations and potential applications of an operational DSM modelling platform and produced maps?
   (b) How can we contribute to open science and make DSM models reproducible and easy to update?

Using BIS-4D, I expect to address the aim, objectives and research questions using various methods outlined in Fig. 1.5. In the case of a dynamic soil property, the steps can be

summarized as follows: 1) gather and prepare soil point data, where Y is a target soil property that varies in 2D space (s), depth (d) and in time (t); 2) gather and prepare covariates (X) representing the soil forming factors and derive dynamic covariates in 2D+T and 3D+T; 3) overlay soil point observations explicit in 3D+T with covariates in 3D+T to obtain a regression matrix; 4) use machine learning for model calibration and prediction in 3D+T, creating a soil property prediction map at a given depth layer and for a given year; and 5) assess map quality using statistical validation and prediction uncertainty. Time t is represented using different colors. In the case of a soil property considered to be static, the same steps are used but in 3D, not 3D+T. These methods and steps are investigated and discussed further in Chapters 2 - 5 and provided here only as a schematic overview.

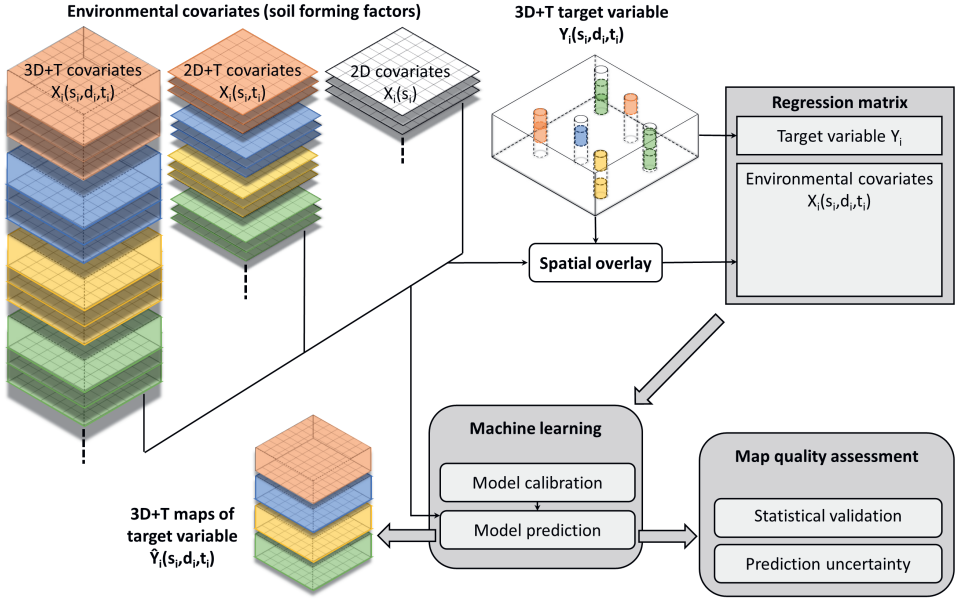


Figure 1.5: Graphical abstract of the BIS-4D soil modelling and mapping platform, where Y is a target soil property and X are covariates that vary in 2D space (s), depth (d) and in time (t). Time t is represented using different colors.

1.7 Thesis structure

This thesis is organized in six chapters, including the general introduction. Chapters 2 - 5 address the four objectives and related research questions described above (Sect. 1.6) and are peer-reviewed publications (Chapter 2), have been peer-reviewed and accepted for

publication (Chapter 3) or are currently undergoing peer-review (Chapters 4 - 5). Hence, Chapters 2 - 5 can also be read separately.

In Chapter 2, I develop a 3D model to predict soil pH at 25 m resolution between 0 - 2 m depth in the Netherlands. I assess mapping accuracy using various statistical validation approaches and introduce spatially explicit accuracy thresholds based on GSM specifications to visualize prediction uncertainty.

In Chapter 3, I extend the 3D mapping methodology to 3D+T by developing a model that can predict SOM for any year between 1953 - 2022, also at 25 m resolution and between 0 - 2 m depth. To do so, I develop novel dynamic covariates explicit in 2D+T and 3D+T and revisit soil sampling locations from the past throughout the Netherlands to statistically validate SOM changes.

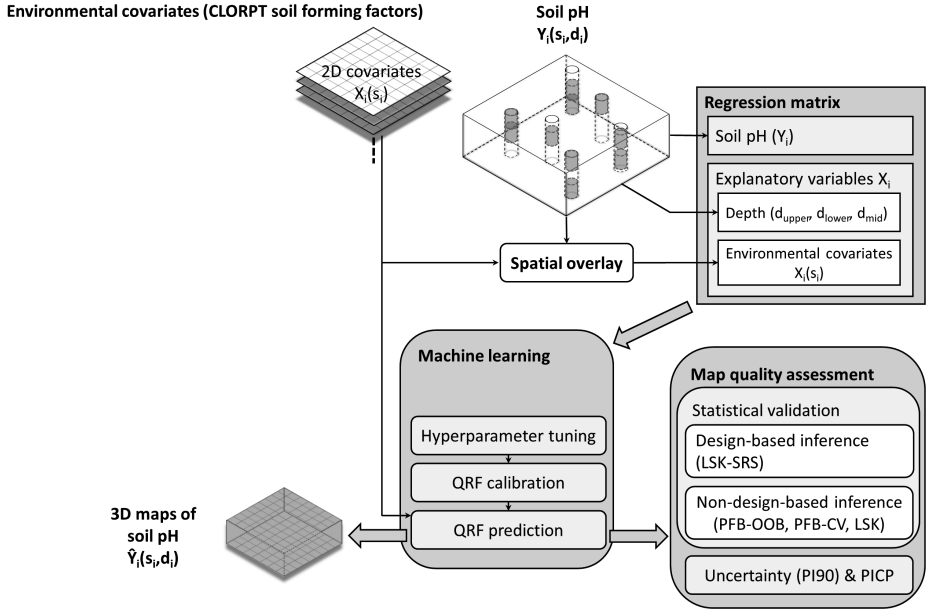
Next, the 3D+T model from Chapter 3 is applied to predict SOM in 2050 based on a nature-inclusive land use scenario (Chapter 4). Here, I investigate whether SOM-related soil health and nature-inclusive transitions are mutually beneficial and discuss the potentials and limitations of combining the 3D+T SOM model with the 2050 scenario model.

In Chapter 5, I further build on previous chapters by developing BIS-4D into an operational, reproducible, standardised, largely automated and efficient modelling and mapping platform. I predict soil texture (clay, silt and sand content), bulk density, SOM, pH, total nitrogen, oxalate-extractable phosphorus, cation exchange capacity and their associated uncertainties at 25 m resolution between 0 - 2 m depth on a national scale.

Finally, in Chapter 6, I discuss the general findings and implications of this thesis, recommend future research and conclude. All references of the literature are combined in a bibliography at the end of the thesis.

Chapter 2

Tier 4 maps of soil pH at 25 m resolution for the Netherlands



This chapter is based on:

Helfenstein, A., Mulder, V.L., Heuvelink, G.B.M., Okx, J.P., 2022. Tier 4 maps of soil pH at 25 m resolution for the Netherlands. *Geoderma* 410, 115659. https://doi.org/10.1016/j.geoderma.2021.115659

Abstract

Accurate and high resolution spatial soil information is essential for efficient and sustainable land use, management and conservation. Since the establishment of digital soil mapping (DSM) and the goals set by the GlobalSoilMap (GSM) working group, great advances have been made to attain spatial soil information worldwide. Highly populated areas such as the Netherlands demand multi-functional land use, for which information of key soil properties such as pH is essential to make decisions. We a) provide soil pH prediction maps at six standard depth layers between 0 m to 2 m for the Netherlands at 25 m resolution, whereby the calibrated Quantile Regression Forest (QRF) model allows for prediction at any desired depth, and b) determine map accuracy using various statistical validation strategies and evaluation of prediction uncertainty. This study is unique among GSM products by including design-based inference of a probability sample as an external accuracy assessment and providing Tier 4 maps with spatially explicit accuracy thresholds for end-users based on GSM specifications.

QRF models were tuned and calibrated using 15 338 soil observations between 0 m and 2 m depth from 4230 locations and 195 covariates representing the soil-forming factors. The following statistical validation strategies were used for external accuracy assessment of map quality: out-of-bag, location-grouped 10-fold cross-validation, an independent validation set (5677 observations, 1367 locations) and a stratified random sample of the independent validation set separated by depth layer. Mean error (ME), root mean squared error (RMSE), model efficiency coefficient (MEC) and the prediction interval coverage probability (PICP) were calculated in all four strategies. In addition, the 90th prediction intervals were used to categorize each 25 m pixel into "none", A, AA or AAA quality as a measure of the internal accuracy assessment.

We obtained large differences depending on the four external accuracy assessment strategies and depth layer (ME = -0.08 - 0.20, RMSE = 0.41 - 0.83, MEC = 0.64 - 0.90, PICP of PI90 = 0.80 - 0.94). Design-based inference (LSK-SRS) was most indicative of map accuracy based on sampling theory (ME = 0.09 - 0.17, RMSE = 0.7 - 0.79, MEC = 0.73 - 0.82). QRF prediction uncertainty was slightly overestimated. Less than 10 % of pixels were designated with AA and AAA and therefore we recommend future studies to also test the achievability of high quality thresholds for Tier 4 GSM maps. We believe these 3D soil pH maps at 25 m resolution are useful for a variety of end users and that our workflow can be applied elsewhere and for other soil properties to further diminish the gap of missing spatial soil information.

2.1 Introduction

Soil is a vital part of the natural environment and essential for global ecosystem services, including production of food and fiber, water infiltration, climate regulation, and maintaining biodiversity. Decision makers therefore require accurate spatial soil information to ensure that the soil and land are being used, managed and conserved in an efficient and sustainable way. Digital soil mapping (DSM) is often used to attain spatially explicit soil information. DSM is the computer-assisted production of soil type and soil property maps, using statistical models to infer the relationship between a response, the soil type or soil property, and the predictors, the spatially exhaustive environmental explanatory variables (McBratney et al., 2003; Scull et al., 2003). Usually, the predictors, also termed covariates, are directly or indirectly related to the main soil forming factors: climate, organisms, relief or topography, parent material and time (Dokuchaev, 1899; Jenny, 1941).

The GlobalSoilMap (GSM) working group of the International Union of Soil Sciences (IUSS) was formed across eight geographic "nodes" around the world to strive for a common goal: a high-resolution spatial soil information system of selected soil properties and their uncertainties at six standard depths for the entire world (Arrouays et al., 2014b). In order to achieve this common goal, both top-down and bottom-up approaches have been implemented. SoilGrids, perhaps the most prominent top-down approach, provides global maps of key GSM soil properties developed by ISRIC – World Soil Information, which have since its initial release (Hengl et al., 2014) been improved and updated twice (Hengl et al., 2017b; Poggio et al., 2021). Other examples of global soil maps include Global Gridded Surfaces of Selected Soil Characteristics (IGBP-DIS; Global Soil Data Task Group, 2000), WISE30sec (Batjes, 2016), S-World (Stoorvogel et al., 2017) and SoilKsatDB for soil saturated hydraulic conductivity (Gupta et al., 2021).

In parallel, bottom-up approaches to create soil information systems (SIS) at regional, national and continental scales have also been implemented. A few examples of countries with SIS using DSM techniques include Denmark (Adhikari et al., 2014a), the United States (Hempel et al., 2014b), Nigeria (Akpa et al., 2014), Australia (Viscarra Rossel et al., 2015), France (Mulder et al., 2016a,b), Scotland (Poggio & Gimona, 2017a) and more recently Brazil (Gomes et al., 2019), China (Liang et al., 2019; Liu et al., 2020) and India (Dharumarajan et al., 2019, 2020). A global soil organic carbon map (GSOC) was also created based on national soil organic carbon (SOC) maps from 110 countries (Brus et al., 2017; FAO, 2018), which are maintained by the Food and Agriculture Organization of the United Nations (FAO, 2017).

The first publication of the spatial distribution of soil properties in the Netherlands dates back to the 19th century (Felix, 1995). Systematic soil mapping became institutionalized with the establishment of the Dutch Soil Survey institute (StiBoKa) in 1945 (Hartemink &

Sonneveld, 2013). From 1950 to 1995, StiBoKa conducted conventional soil surveys (Buringh et al., 1962) and produced national maps of soil types at a 1:50 000 scale. A review of the history of soil mapping in the Netherlands and its different phases including the first decade of the 21st century was conducted by Hartemink & Sonneveld (2013). Various studies compared different (geo)statistical methods and developed prototypes of qualitative and quantitative soil property maps for the Netherlands using the data collected by StiBoKa (Brus & Heuvelink, 2007; Brus et al., 2009; Kempen et al., 2014). In addition, a variety of DSM techniques were used to update the Dutch soil maps, with a focus on soil organic matter (SOM) and peatland regions (Kempen et al., 2009, 2011, 2012a). More recently, SOM was estimated at a national scale using a soil type and binary land use map (arable land or grassland), at a resolution of 250 m at four fixed depths (van den Berg et al., 2017). On the basis of this SOM map, a Dutch contribution to the Global Soil Organic Carbon (GSOC) map was also delivered for the topsoil (0 cm to 30 cm), which was spatially aggregated to 1 km resolution (Walvoort & Hoogland, 2017). Nevertheless, there is an increasing demand for accurate, 3D, and high resolution information of key soil properties for the Netherlands. This is especially important for highly populated and relatively small countries such as the Netherlands (land area = 33 481 km^2) because land use decisions are often made on a field scale, e.g. per agricultural parcel.

Since the establishment of DSM as a research field, the main focus has been on implementing new methods to improve the predictive performance of soil maps. Today, pedometricians can make use of increasing amounts of available spatial data as well as an extensive toolkit of geostatistical and machine learning approaches combined with a powerful computational infrastructure. However, considerably less effort has been invested in providing appropriate measures of the quality of soil maps. This is essential for DSM products to be adopted by a broader community, for future research guidance and most importantly, to ensure that the quantified accuracy is suitable to fulfill the map's purpose (Arrouays et al., 2020).

The quality of maps can be evaluated using internal and/or external accuracy assessment measures. One way to quantify internal, or model-based accuracy assessment is using the prediction uncertainty. In this regard, Quantile Regression Forest (QRF; Meinshausen, 2006) models are advantageous within the DSM toolkit not only due to their predictive performance, but also for their ability to quantify prediction uncertainty. Ensemble decision tree models such as Random Forest (RF; Breiman, 2001) and QRF have repeatedly outperformed other machine learning and non-machine learning approaches in DSM applications (e.g. Hengl et al., 2015; Nussbaum et al., 2017; Keskin et al., 2019). In addition, QRF delivers a probability distribution of the soil property at each prediction location, rather than a single (mean) prediction as with RF (Meinshausen, 2006). To the best of our knowledge, this makes it unique among other machine learning approaches in that the algorithm inherently also gives an indication of the prediction uncertainty. This may be a reason for the increasing use of QRF in DSM in recent years (e.g. Vaysse & Lagacherie,

2017; Lagacherie et al., 2019, 2020; Dharumarajan et al., 2020; Poggio et al., 2021).

Another advantage of QRF prediction uncertainty is that it can be incorporated in the concept of accuracy thresholds for Tier 4 GSM products. In order to coordinate and guarantee a minimum quality for soil maps, specifications were made regarding the spatial entity, soil properties, date, uncertainty, validation, documentation and reproducibility of GSM products (Arrouays et al., 2014a; Hempel et al., 2014a). Increasing specifications have to be fulfilled with increasing quality of a DSM product, which are organized into so-called *Tiers* (Arrouays et al., 2015). Tier 4 products, which have the strictest requirements, specify three levels of accuracy thresholds (A, AA and AAA) depending on the soil property and depth layer (Appendix 2.6, Table 2.5), although note that also none of these three levels can be met. These levels specify that prediction uncertainties should be within certain ranges, with an increasingly narrow and therefore accurate range from A to AAA. This may provide a powerful measure to communicate the quality of a soil map to end users for a specific purpose and region. Expressing the uncertainty of predictions in a meaningful way for end users was described as one of the ten major challenges for pedometricians (Wadoux et al., 2021b). However, these accuracy thresholds have to our knowledge not yet been used in DSM studies.

Internal accuracy assessment using QRF has the advantage that prediction uncertainty and their respective GSM accuracy thresholds are spatially explicit. However, the disadvantage is that these are based on the model structure and model assumptions. Therefore, there is also a need for model-free evaluation of the map's accuracy, i.e. external accuracy assessment.

For assessing the external accuracy, the vast majority of DSM studies use statistical validation methods (Wadoux et al., 2020; Piikki et al., 2021). This usually involves data-splitting, either using a single split involving a calibration and validation set, or repeating this multiple times during n-fold cross-validation (CV). These validation methods are a form of external accuracy assessment because the independent and separate observations not used in model calibration are compared to the predictions at the observation locations (Chatfield, 1995). However, if the validation locations are not selected using a probability sampling design, then the accuracy assessment may be biased (Brus et al., 2011; Brus, 2014, 2019). In summary, external accuracy assessment without a probability sample gives an indication of the accuracy at independent locations, but these locations may not be indicative of the map itself. Therefore, Brus et al. (2011) conclude that, when evaluating map quality, a probability sample and associated design-based statistical inference should be used for the external accuracy assessment whenever possible. This classical sampling theory method is statistically sound and has been extensively described in statistics (Cochran, 1977) and environmental science (de Gruijter et al., 2006; Gregoire & Valentine, 2007). However, in two recent systematic reviews, Wadoux et al. (2020) reported that only two out of 150 studies used an additional probability sample for vali-

dation (Subburayalu & Slater, 2013; Lacoste et al., 2014) and Piikki et al. (2021) reported that only 13 % of 188 studies used probability sampling. This is most likely because probability samples are often not available due to time and cost restraints (Domburg et al., 1997; Hartemink & McBratney, 2008; Hartemink et al., 2010).

Given the strong demand for high-resolution 3D soil information in the Netherlands and the need to properly assess map quality, this study has two main objectives. Firstly, we aim to contribute to the GSM project by providing soil pH prediction maps for the Netherlands at 25 m resolution, at any desired depth between 0 m and 2 m, using QRF. We chose to focus on soil pH because it is an indispensable soil property to assess soil processes and fertility: it not only provides information on acidity and alkalinity, but is also an indication of nutrient availability, metal dissolution and (micro-)biological activity (Miller & Kissel, 2010; Weil & Brady, 2017). Secondly, we aim to quantify map accuracy using a) spatially explicit QRF prediction uncertainty and respective GSM Tier 4 accuracy thresholds and b) statistical validation strategies.

2.2 Materials and methods

2.2.1 Soil point datasets

We used 21 015 pH measurements, or observations, from 5597 locations between 0 m to 2 m depth excluding the O horizon or humus layer (Table 2.1, Fig. 2.1). Excluding built-up (urban) and water surface area, this approximately yields an average density of 1 soil sampling location per 5 km^2. All observations were retrieved from the Dutch soil database, or "Bodemkundig Informatie Systeem" (BIS; IenM & TNO, 2017; TNO, 2020). We chose to use pH measurements conducted in KCl suspension (pH [KCl]) as opposed to the internationally more frequently used H_2O or $CaCl_2$ suspension methods because KCl suspension was the preferred measurement method in the Netherlands from 1950 to 2000 (Supplement S1, Fig. S1[1]). There were less than 1000 measurements available using the other methods and we refrained from converting between methods since this may introduce substantial uncertainty.

Table 2.1: Descriptive statistics of soil pH [KCl] for calibration (PFB) and validation (LSK) data.

Dataset	Locations	Observations	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Skewness
Calibration (PFB)	4230	15338	0.90	4.20	4.80	5.20	6.10	9.00	0.52
Validation (LSK)	1367	5677	1.90	4.40	5.20	5.53	7.00	8.20	0.26

For model calibration, we used 15 338 pH [KCl] measurements from 4230 locations (Table 2.1, Fig. 2.1). At these locations, profile descriptions, or "Profielbeschrijving" (PFB),

[1]Supplements of Chapter 2 are available at `https://doi.org/10.1016/j.geoderma.2021.115659` under "Supplementary data".

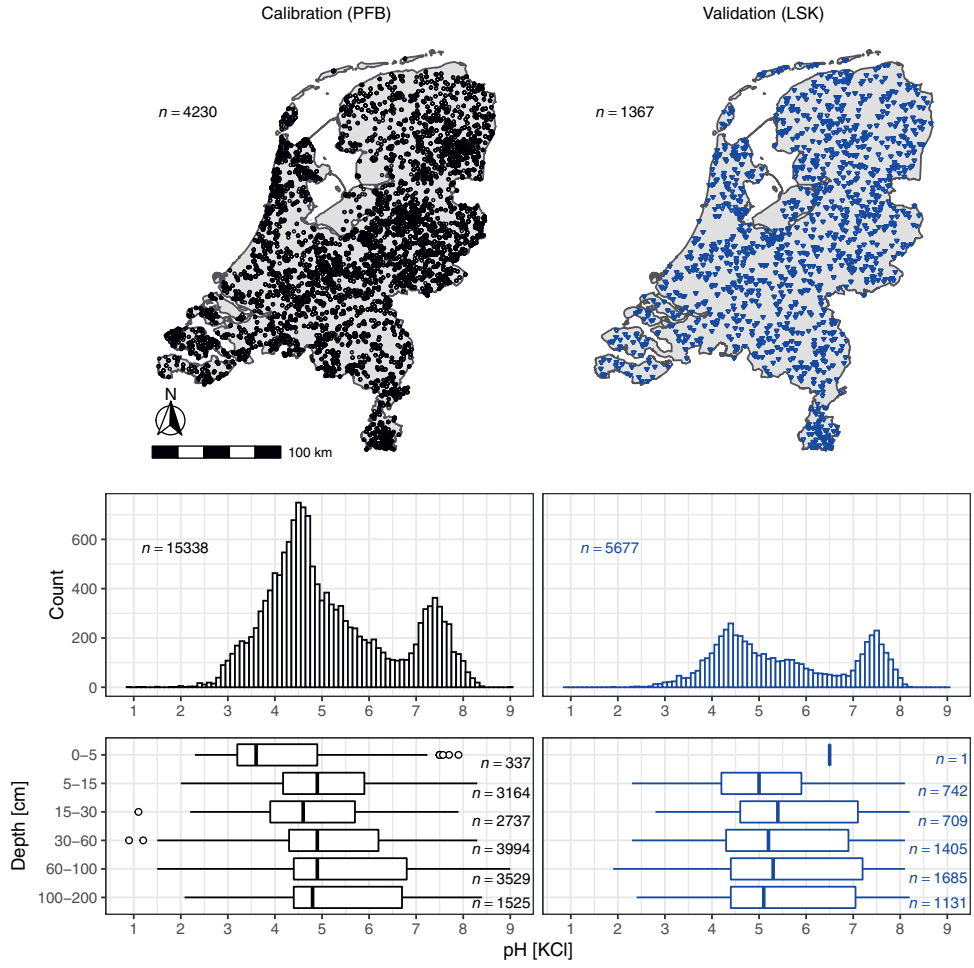


Figure 2.1: Soil pH [KCl] sampling locations, histograms and boxplots grouped by depth layer of calibration (PFB; left in black) and validation (LSK; right in blue) data. Observations were grouped into depth layers using the midpoint of each sampled soil horizon.

were made, soil samples were collected from each horizon between 1953 and 2012 and measured in the lab (Supplement S1, Fig. S1). This dataset was specifically chosen for model calibration because it constitutes the majority of soil pH data in the Netherlands. The somewhat clustered locations cover all regions of the Netherlands with the exception of southwestern Flevoland (Fig. 2.1). The pH calibration data follow a bimodal distribution with the majority of values between 4 and 5 and a smaller peak around 7.5 (Fig. 2.1). Bimodal distributions for soil pH are common and in the case of the Netherlands can be attributed to the dominating Pleistocene sandy soils vs. Holocene clay soils. Grouped

into the GSM depth layers by the respective midpoints of the sampled layers, the median pH values are around 4.5 to 5 across all depth layers. Three unexpectedly low values were measured between 15 cm and 60 cm depth (Fig. 2.1), but there was insufficient evidence for them to be classified as outliers and removed.

The separate and independent validation data were gathered during the "Landelijke Steekproef Kaarteenheden" (LSK) between 1993 and 2000 (5677 measurements from 1367 locations; Supplement S1, Fig. S1). Soil sampling locations were determined in the LSK campaign using a probability sample, more specifically a stratified simple random sample (SRS), wherein 94 strata were defined based on soil type and groundwater class (Finke et al., 2001; Visschers et al., 2007). As with the calibration (PFB) data, observations were made for each soil horizon, which indicates that it is only a SRS in 2D space. This has implications for the statistical validation (Sect. 2.2.6). The validation set also has a bimodal distribution, although the relative difference between the two peaks is much smaller than for the calibration set (Fig. 2.1). Consequently and in contrast to the calibration data, the overall median as well as the median of each grouped depth layer is above 5 (Table 2.1, Fig. 2.1). Note that there are considerably fewer observations ($n = 251$) in the validation set with a midpoint between 15 cm to 30 cm compared to the other depth layers.

2.2.2 Covariate selection

The covariates (total number $P = 195$) were chosen specifically to represent the soil forming factors. The covariates are summarized in Table 2.2 and a complete list is included in Supplement S2, Table S1.

As indicators of the soil forming factor climate, we used the long-term mean, minimum and maximum temperature and precipitation between 1981 and 2010 from the Royal Netherlands Meteorological Institute (Table 2.2; KNMI, 2020).

The majority of covariates used in this study are historical, agricultural or natural land use and vegetation maps relating to the soil forming factor "organism" (Table 2.2). We specifically chose a large number of these maps because there is profound anthropogenic influence and high land use intensity in the Netherlands. Approximately 82 % of the land surface in the Netherlands is agricultural, urban or infrastructure (Hazeu et al., 2020). Multiple versions covering different time spans were included.

We used the national digital elevation model (DEM) of the Netherlands, or "Actueel Hoogtebestand Nederland" (AHN), and commonly used DEM derivatives for the soil forming factor topography (Table 2.2). The standard deviation of the systematic as well as random error of AHN 2 and 3 is ± 5 cm (AHN, 2023). With such high accuracy, we considered topographic covariates to be informative not only for hilly regions but also for the large majority of the Netherlands that is relatively flat. We computed the following commonly used DEM derivatives: slope, profile curvature, deviation from the mean value

Table 2.2: Summary of the main covariates used grouped by soil forming factor. For the complete list of covariates, see Supplement S2, Table S1.

Soil forming factor	Description	Timespan/ version	Source
Climate	Long-term mean, minimum & maximum temperature	1981-2010	KNMI (2020)
Climate	Long-term mean precipitation	1981-2010	KNMI (2020)
Organism	Land use (historical): "Historisch Grondgebruik Nederland" (HGN)	1900, 1960, 1970, 1980	Alterra (2004)
Organism	Land use (recent): "Landelijk Grondgebruiksbestand Nederland" (LGN)	1986-2019	WENR (2020); Hazeu et al. (2020)
Organism	Land use (recent): "Bestand Bodemgebruik" (BBG), Top10NL	1993, 1996-2019	CBS (2015); BRT (2020)
Organism	Copernicus land monitoring: CORINE Land Cover (CLC), Riparian land cover, water and wetness index, grassland vs. non-grassland, tree cover density	1986, 2000, 2006, 2012, 2018	EEA (2018); Thunnissen & Middelaar (1995); Hazeu & Wit (2004); EEA (2007)
Organism	Nature land cover maps	1988, 1990, 2003, 2004, 2013	Bakker et al. (1989); Kramer & Clement (2015); Sanders & Prins (2001)
Organism	Agricultural crop parcels ("BRP Gewaspercelen")	2005-2019	EZK (2019)
Organism	Agricultural management type, ammonia & nitrogen emissions, manure application	1993, 2018, 2019	BIJ12 (2019); RIVM (2020)
Organism	Water drainage classes, areas behind dikes or not		Maas et al. (2019)
Organism	Vegetation maps: forest classified by age, recreational use, tree species, tree height		de Vries & Al (1992); Clement (2001)
Topography	DEM: "Actueel Hoogtebestand Nederland" (AHN) & hillshade	AHN1 (1997-2004), AHN2 (2007-2012), AHN3 (2014-2019)	AHN (2023)
Topography	AHN2 derivatives: slope, profile curvature, deviation from mean, openness, Topographic Wetness Index (TWI), Multiresolution Valley Bottom Flatness (VBF), valley depth	AHN2 (2007-2012)	AHN (2023)
Parent material	Geomorphology based on geomorphological classes, genesis, form, formation begin and end and relief	2004, 2008, 2019	Koomen & Maas (2004); Maas et al. (2019)
Parent material	Physical geography and groundwater maps	2013, 2004	EZK (2013); KRW (2004)
Parent material	(Paleo-)geographical maps	9000-250 B.C. 100-1850 A.D.	Vos (2015); Vos et al. (2020)

within a local neighborhood, positive and negative openness, Topographic Wetness Index (TWI), Multiresolution Valley Bottom Flatness (MrVBF; Gallant & Dowling, 2003) and valley depth (Wood, 1996, 2009). The deviation from the mean value was computed within a radius of 11 cells (275 m) to account for local changes in topography. AHN 2 was used to obtain these derivatives because it has a higher accuracy than AHN 1 and because AHN 3 has not been thoroughly validated yet (AHN, 2023). A hillshade from the AHN 2 was also downloaded and used.

We used geomorphological, paleo- and physical geography maps as indicators of parent material (Table 2.2). The parent material for soils in the Netherlands consists almost exclusively of geologically young material from fluvial and coastal lowlands of the Holocene age as part of the Rhine-Meuse delta (60 %) as well as Pleistocene sand (van der Meulen et al., 2013). In this sense for the Netherlands there are no lithology or bedrock maps commonly used in DSM studies in other parts of the world.

For many covariates, multiple versions from different years were included to account for changes in soil forming factors over time. In addition, several of the covariates were based on each other. For example, "Landelijk Grondgebruiksbestand Nederland" (LGN) uses "Bestand Bodemgebruik" (BBG) and "Top10NL" data. This indicates that many of the covariates are highly correlated. Ensemble decision tree models are robust against highly correlated data; it does not cause an overfit or decrease prediction accuracy. However, it is important to note that the higher the number of correlated covariates, the lower the relative importance of each will become, which leads to a distorted variable importance measure (Strobl et al., 2007; Kuhn & Johnson, 2013). For this study, we did not refrain from using many highly correlated covariates because we deemed prediction accuracy more important than model interpretability based on variable importance measures.

2.2.3 Covariate preprocessing

All covariates were first visually explored for inconsistencies. Rasters were exported at a target resolution of 25 m because this matches the resolution of the LGN land use maps (WENR, 2020; Hazeu et al., 2020) and allows for land use decisions at a fine resolution, e.g. within a small agricultural parcel.

The first step of covariate preparation and preprocessing was to project all covariates to the Amersfoort or RD New coordinate reference system (EPSG:28992). Next, all covariates were resampled to a common origin, extent and resolution. In this step, continuous covariates were resampled using the cubic spline method whereas categorical covariates were resampled using the nearest neighbor method. During reprojection and resampling, the AHN2 was used as reference and the AHN2 "no-data" layer was used as a mask (water and buildings).

Many of the categorical covariates were reclassified because some classes did not occur at

observation locations. For example, the detailed classes ($n = 15$) of different cereals in crop rotation covariates ("BRP Gewaspercelen") were aggregated into one general cereals class (Supplement S2, Table S1).

We stacked the covariates and extracted values at all calibration locations by overlaying them with the covariate stack, resulting in a regression matrix used for model tuning and calibration. Sampling depth information was also included as a predictor in the regression matrix. Including depth along with spatial covariates in a so-called "3D" modelling approach has been used before (Akpa et al., 2014; Filippi et al., 2019, 2020; Hengl et al., 2017b; Ramcharan et al., 2018; Zhang et al., 2020) and is compared in detail to so-called 2D and 2.5D approaches in Ma et al. (2021). More specifically, we included the midpoint of each sampled layer or horizon, as well as the upper and lower boundary to also account for horizon thickness. In summary, we chose to include depth information so that predictions can easily be made at any chosen depth (user specific) and as a means to account for changes in soil pH over depth.

2.2.4 Model tuning and calibration

For model tuning, calibration, and prediction, it is important to differentiate between mean and median predictions when using QRF. During calibration, for each node in each tree, RF keeps only the mean of the observations that fall into each node. In contrast, QRF keeps the value of all observations in each node (Meinshausen, 2006). Based on this, the fitted QRF can then be used to yield a cumulative probability distribution (i.e., quantiles of the distribution) of the response pH at every sampled location and depth during prediction. In predictive modelling, users are generally interested in the best possible predictions that are closest to the "truth". It thus makes sense to go for the expected value, i.e. the conditional mean. If the median is used, as e.g. retrieved from the 0.50 quantile in QRF, predictions may be biased if the response is not symmetrically distributed. In addition, median predictions are not additive, if e.g. soil organic carbon stocks need to be calculated from a soil organic carbon map. However, the advantage of using the median is that it is more robust to outliers. For model tuning, we grew RF (not QRF) models with the goal of optimizing hyper-parameters for mean predictions; therefore, there was no need to keep the value of all observations in each node as in QRF (Meinshausen, 2006), which greatly decreased computation time and did not change the tuning results. However, for the final model calibration, a QRF was fitted so that predictions could thereafter be made for both the mean and quantiles (including median).

Model tuning for RF was performed using a location-grouped 10-fold CV wherein all PFB observations from the same location were grouped, abbreviated hereafter as PFB-CV. This means that all observations from the same soil profile were either part of the hold-in or hold-out fold across each of the 10 folds. We tested all combinations (full cartesian grid search) of the following hyper-parameters (Boehmke & Greenwell, 2020):

- **Number of trees in the forest (*ntree*):** 100, 150, 200, 250, 500 (`ranger` default), 750, 1000
- **Number of covariates to consider at any given split (*mtry*):** $\sqrt{P}$ (`ranger` default) and 25 %, 33.3 % (`randomForest` default) and 40 % of P, i.e. 14, 49, 65, 78
- **Complexity of each tree (*minimal nodesize*):** 1, 3 and 5
- **Sampling with replacement (*replace*)**: TRUE (sample with replacement) and FALSE (sample without replacement)
- **Fraction of observations to sample (*sample.fraction*):** 0.5, 0.63 and 0.8 (based on recommendations from Boehmke & Greenwell (2020); this only applies if *replace* = FALSE)

The final set of hyper-parameters was chosen based on the lowest root mean squared error (RMSE) across the 10-fold CV. When the increase in RMSE was below 0.1 %, the model with fewer trees was chosen to reduce computation time. Besides the commonly tuned *ntree*, *mtry* and *nodesize* hyper-parameters, we also tested different values related to the sampling scheme. Sampling with replacement can lead to biased variable split selection when there are many categorical covariates with varying numbers of levels (Janitza et al., 2016; Strobl et al., 2007). Hence, we tested sampling without replacement because we had many categories that were not balanced, hoping to achieve a less biased use of all levels across the trees in the forest. In addition, decreasing the sample fraction size of observations leads to more diverse trees and thus lowers between-tree correlation, which can increase the prediction accuracy, especially if there are a few dominating covariates (Boehmke & Greenwell, 2020). The splitting rule used during tree construction (*splitrule*) was held constant at the default value of selecting the split at each node that minimizes the variance of the response.

The final QRF used for model predictions was fitted using all soil observations in the calibration set (n = 15 338), covariates including depth indications (P = 195) and the final set of optimized hyper-parameters. Permutation was used to assess relative variable importance during model fitting. In this method, the mean squared error (MSE) is compared to the MSE after permuting the values of a covariate, yielding a difference in MSE per covariate. These MSE differences are normalized by the standard deviation of the MSE differences over all covariates (Breiman, 2001).

2.2.5 Maps of predicted soil pH, uncertainty and accuracy thresholds

The calibrated QRF were used to derive the mean, median (0.50 quantile; $q_{0.50}$), 0.05 quantile ($q_{0.05}$) and 0.95 quantile ($q_{0.95}$) at every 25 m pixel and each standard depth layer specified by GSM (0 - 5 cm, 5 - 15 cm, 15 - 30 cm, 30 - 60 cm, 60 - 100 cm and 100 - 200 cm) over the Netherlands. Predictions were made at the same support as the observations, i.e. at point support at the center of each pixel and the specified depth increment. Support is

defined as the area or volume over which a measurement or prediction is made (Webster & Oliver, 2007, Sect. 4.8).

In addition, spatially explicit 90 % prediction intervals (PI90) were obtained at every 25 m pixel as a measure of prediction uncertainty as follows:

$$PI90 = q_{0.95} - q_{0.05} \tag{2.1}$$

As an additional measure of map quality using internal accuracy assessment, we used the PI90 to designate every 25 m pixel at every predicted depth layer into one of four thresholds: none, A, AA and AAA (Table 2.3). These accuracy thresholds are specified by GSM Tier 4 products (Arrouays et al., 2015) and do not vary over depth in the case of soil pH (Appendix 2.6, Table 2.5). From "none" to AAA, the PI90 (uncertainty) of a given prediction gradually decreases, indicating that users can be very certain about predictions at AAA locations and least certain about predictions at "none" locations. For example, for a AAA pixel 9 out of 10 times the true value is less than ±0.5 pH units from the mean prediction, and less than ±1.5 pH units from the mean prediction for a A pixel.

2.2.6 Evaluation of map accuracy using statistical validation

Non-design-based inference

We also evaluated map quality using external accuracy assessment in the form of statistical validation strategies (Table 2.3). Firstly, we used the out-of-bag (OOB) observations, in other words the PFB observations not selected during bootstrapping when QRF is calibrated (PFB-OOB; Breiman, 2001). This is commonly used in various disciplines to assess accuracy of RF or other ensemble decision tree models. Secondly, we used location-grouped 10-fold CV (PFB-CV; Sect. 2.2.4). Compared to PFB-OOB, this method was chosen because it prevents observations from the same location in being both in the hold-in and hold-out set, wherein the hold-in samples are used for model calibration and the hold-out for model validation. PFB-OOB and PFB-CV both only make use of the PFB calibration dataset (Sect. 2.2.1). Thirdly, we used the LSK dataset as an independent validation set. The probability sampling design of the LSK cannot easily be utilized when considering all depths because there are multiple observations from different depth layers at the same locations. This means that in 3D, it cannot be considered a SRS. We nevertheless included this strategy because we wanted to investigate whether there are substantial differences between a non-design-based vs. design-based (see below) inference of LSK.

To obtain commonly used accuracy metrics, mean predictions at all depths were used to calculate residuals and estimate from them the mean error (ME or bias), the RMSE and

Table 2.3: Five strategies to evaluate map quality, based on an internal or external (i.e. statistical validation) accuracy assessment. ME, RMSE, MEC and their respective CI95s in LSK-SRS were calculated using probability sampling theory for SRS (Eq. 2.5 - 2.11). *PI90 and PICP of LSK and LSK-SRS are identical, respectively, since the same observations are compared to the respective PIs. **Accuracy metrics in LSK-SRS can only be computed for separate layers in order to adhere to the probability sampling design, whereas they can also be computed using observations at all depths for the other statistical validation strategies.

Accuracy assessment	Acronym	Description	Statistical validation	Dataset	Accuracy metrics	2D space	Depth**
Internal	–	Tier 4 GSM accuracy thresholds	–	PI90 of predictions	**None** (PI90 > 3.0 pH units) **A** (PI90 ≤ 3.0 pH units) **AA** (PI90 ≤ 2.0 pH units) **AAA** (PI90 ≤ 1.0 pH units)	explicit (25 m pixels)	User specific
External	PFB-OOB	Out-of-bag	non-design-based	PFB	ME, RMSE, MEC, PI90, PICP	point locations	All, layers
External	PFB-CV	Location-grouped 10-fold CV	non-design-based	PFB	ME, RMSE, MEC, PI90, PICP	point locations	All, layers
External	LSK	Independent validation	non-design-based	LSK	ME, RMSE, MEC, PI90*, PICP*	point locations	All, layers
External	LSK-SRS	SRS of independent validation	design-based	LSK	ME, RMSE, MEC, CI95, PI90*, PICP*	strata weighed	Layers

the model efficiency coefficient (MEC):

$$\widehat{ME} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right) \tag{2.2}$$

$$\widehat{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2} \tag{2.3}$$

$$\widehat{MEC} = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2} \tag{2.4}$$

where n is the number of observations, Y_i and $\widehat{Y}_i$ are the i^{th} observation and prediction, respectively, at a certain location and depth, and $\overline{Y}$ is the mean of all observations. The MEC was originally used in hydrological modelling (Nash & Sutcliffe, 1970) and is also referred to as the mean squared error skill score in other disciplines such as meteorology (Wilks, 2011). In addition, all quantiles from 0 to 1 were predicted at all depths at all observation locations for statistical validation to obtain the PI90 as well as the prediction interval coverage probability (PICP) of all PIs. The PICP is the proportion of observations that fall into the corresponding prediction interval (Papadopoulos et al., 2001). If the model is able to accurately quantify the uncertainty, then the percentage of observations within a PI should be close to the PICP.

Design-based inference

In order to conduct a design-based inference of map accuracy using the SRS probability sample, we grouped the LSK observations into the GSM depth layers so that there was at most one observation at each location (LSK-SRS) for every depth layer. Observations were grouped into GSM depth layers by allocating each observation to the layer in which the midpoint of the sampled soil horizon lies. This means that some locations had no observations for that particular depth layer while other layers had more than one. For locations where there were more than one observation per depth layer, the observation was chosen whose midpoint was closest to the midpoint of the GSM depth layer. This meant that not every observation was used for LSK-SRS. If the distances were identical, then the observations and predictions were averaged (mean). If there were no observations for an entire stratum for a particular depth layer, then that stratum was removed from the analysis. The number of observations that were left out or averaged as well the left-out strata and the percentage of land they constituted were reported for each depth layer (Table 2.4). The uppermost GSM depth layer cannot be validated using LSK (both design- and non-design based inference) because there is only 1 observation from 0 - 5 cm (Fig. 2.1 and Table 2.4).

Table 2.4: Metadata of the LSK-SRS method per depth layer, including number of removed observations, number of averaged observation pairs, percentage of the strata that were removed from the total ($H = 94$), which strata were removed and the percentage of the Netherlands covered. *We refer to the strata codes from Finke et al. (2001), Appendix 1.

Depth layer	Observations		Strata		% NL coverage
	# removed	# averaged pairs	% removed	Removed*	
0 - 5 cm	-	-	-	-	-
5 - 15 cm	0	0	6.38	1904, 1910, 1915, 2007, 2108, 2114	98.51
15 - 30 cm	0	0	7.45	1913, 1914, 1917, 2102, 2116, 2117, 2901	95.66
30 - 60 cm	73	13	0	-	100
60 - 100 cm	417	56	0	-	100
100 - 200 cm	222	4	9.57	1502, 1503, 1504, 1505, 1915, 2201, 2401, 2601, 2701	97.15

For each depth layer (except 0 - 5 cm), the estimates of ME, RMSE and MEC (Eq. 2.2 - 2.4) were adjusted for LSK-SRS according to probability sampling theory. In addition, the lower and upper 97.5 % confidence limits, which together give the 95 % confidence intervals (CI95) of these metrics were also computed according to sampling theory (de Gruijter et al., 2006, Sect. 7.2.4). The estimated mean error ($\widehat{ME}$), the associated estimation error variance and the lower and upper confidence limits were computed as follows:

$$\widehat{ME} = \sum_{h=1}^{H} \left(w_h \cdot \frac{1}{n_h} \sum_{i=1}^{n_h} \left(Y_{hi} - \widehat{Y}_{hi} \right) \right) \tag{2.5}$$

$$Var(\widehat{ME} - ME) = \sum_{h=1}^{H} \left(w_h^2 \cdot \frac{1}{n_h(n_h-1)} \cdot \sum_{i=1}^{n_h} \left(Y_{hi} - \widehat{Y}_{hi} - \left(\frac{1}{n_h} \sum_{i=1}^{n_h} \left(Y_{hi} - \widehat{Y}_{hi} \right) \right) \right)^2 \right) \tag{2.6}$$

$$lower\ \&\ upper\ CL\ =\ \widehat{ME}\ \pm\ qt\,(0.95,\ n-H)\ \cdot\ \sqrt{Var(\widehat{ME} - ME)} \tag{2.7}$$

where H is the total number of strata, n_h is the number of observations in stratum h ($h = 1, \ldots, H$), w_h is the stratum weight, which equals the stratum area A_h divided by the total area A, Y_{hi} and $\widehat{Y}_{hi}$ are the i^{th} observation and prediction in stratum h, respectively, and $qt(0.95, n-H)$ is the 0.95 quantile with $n-H$ degrees of freedom.

The estimated mean squared error ($\widehat{MSE}$), its estimation error variance and respective lower and upper confidence limits were computed in a similar manner:

$$\widehat{MSE}\ =\ \sum_{h=1}^{H} \left(w_h\ \cdot\ \frac{1}{n_h}\ \sum_{i=1}^{n_h} \left(Y_{hi} - \widehat{Y}_{hi} \right)^2 \right) \tag{2.8}$$

$$Var(MSE - \widehat{MSE}) = \sum_{h=1}^{H} \left(w_h^2 \cdot \frac{1}{n_h(n_h-1)} \cdot \sum_{i=1}^{n_h} \left(\left(Y_{hi} - \widehat{Y}_{hi} \right)^2 - \left(\frac{1}{n_h} \sum_{i=1}^{n_h} \left(Y_{hi} - \widehat{Y}_{hi} \right)^2 \right) \right)^2 \right) \tag{2.9}$$

$$lower\ \&\ upper\ CL\ =\ \widehat{MSE}\ \pm\ qt\,(0.95,\ n-H)\ \cdot\ \sqrt{Var(MSE - \widehat{MSE})} \tag{2.10}$$

The RMSE and its respective CI95 were obtained by simply taking the square root of the $\widehat{MSE}$ and its lower and upper confidence limits. ME and RMSE and respective CI95 metrics are in units of the response variable (pH [KCl]).

The estimate of the model efficiency coefficient ($\widehat{MEC}$) and its CI95 for LSK-SRS were calculated as follows:

$$\widehat{MEC}\ =\ 1 - \frac{\widehat{MSE}}{\widehat{Var(Y)}} \tag{2.11}$$

where $\widehat{Var(Y)}$ is defined in Eq. 7.16 in de Gruijter et al. (2006) as:

$$\widehat{Var(Y)}\ =\ \widehat{\bar{Y}_{st}^2}\ -\ (\widehat{\bar{Y}}_{st})^2\ +\ \widehat{V}(\widehat{\bar{Y}}_{st}) \tag{2.12}$$

where

$$\widehat{\bar{Y}_{st}^2}\ =\ \sum_{h=1}^{H} \left(w_h\ \cdot\ \frac{1}{n_h}\ \sum_{i=1}^{n_h}\ Y_{hi}^2 \right) \tag{2.13}$$

$$\widehat{\widehat{Y}}_{st} = \sum_{h=1}^{H} \left(w_h \cdot \frac{1}{n_h} \sum_{i=1}^{n_h} Y_{hi} \right) \tag{2.14}$$

$$\widehat{V}(\widehat{\widehat{Y}}_{st}) = \sum_{h=1}^{H} \left(w_h^2 \cdot \frac{1}{n_h(n_h-1)} \sum_{i=1}^{n_h} \left(Y_{hi} - \frac{1}{n_h} \sum_{i=1}^{n_h} Y_{hi} \right)^2 \right) \tag{2.15}$$

The CI95 of the MEC was computed by taking a bootstrap sample from all observations per stratum 1000 times and then retrieving the 0.025 and 0.975 quantile of the distribution. For strata with just one observation per depth layer ($n_h = 1$), a within-stratum variance cannot be calculated in Eq. 2.6, 2.9 and 2.15 because a minimum of two observations are required. For these strata, we took the average of the within-stratum variances of all strata with two or more observations.

2.2.7 Software and computational framework

The computational framework was entirely based on open source software and performed on a Ubuntu 20.04.1 operating system (OS) with 48 cores and 126 GB working memory (RAM). QGIS (version 3.16.3) was used for covariate and soil prediction map visualization (QGIS Development Team, 2023). All scripts, metadata, reclassification tables (the original covariate values, a description of each class, the reclassified value and description of the reclassified class) of the categorical covariates and model outputs (soil pH and their associated uncertainty and accuracy threshold maps) are openly accessible (see code and data availability below).

Resampling, reclassification of categorical covariates and masking covariates for buildings and water bodies was done using the GDAL (version 3.1.3) functions `gdalwarp`, `gdal_calc` and `gdal_translate`, respectively (GDAL/OGR contributors, 2023). Reclassification of categorical covariates was automated as much as possible. First, a table was exported from R with all the values within a raster. A short description of the original value (e.g. 3) as well as a reclassified value (e.g. 2) were manually added where necessary (e.g. "barley" and "cereals"). Lastly, each reclassification table was imported back into R, converted into a string format necessary for GDAL's `gdal_calc` function and reclassified accordingly. DEM derivatives were calculated using SAGA-GIS (version 7.3.0; Conrad et al., 2015). Covariate preprocessing steps using GDAL and SAGA-GIS were run on the OS as suggested in Hengl & MacMillan (2019) but parallelized in R (version 4.0.3; R Core Team, 2023) using the `doParallel` (Wallig et al., 2022b) and `foreach` packages (Wallig et al., 2022a). GDAL and SAGA-GIS were specifically chosen for these steps because it massively decreased computation time compared to using similar functions in R using the `raster` (Hijmans, 2020) or `terra` packages (Hijmans, 2023). All other covariate preprocessing steps including extracting covariate values at calibration locations were done in R using the `raster` or `terra` packages.

All model tuning, calibration and evaluation using statistical analysis was done in R. The indices necessary for the location-grouped 10-fold CV were made using the `CAST` package (Meyer, 2023). The remaining model tuning and selection of hyper-parameters were done using the `caret` package (Kuhn, 2019, 2022). We used the `ranger` package (Wright & Ziegler, 2017) with the option "quantreg" to grow a QRF and without it to grow a RF (for tuning). For predictions, the option "quantiles" was used to predict quantiles while the option "response" was used to predict the mean. A combination of the `ranger` and `terra` packages was used for predicting at all locations and depths. Finally, prediction maps were visualized using the `rasterVis` package (Lamigueiro & Hijmans, 2023). The complete computational workflow for 24 pH maps (mean and three quantiles for six depth layers) took approximately 688 CPU-hours and included covariate preprocessing (96 CPU-hours), model tuning and calibration (232 CPU-hours) and prediction (360 CPU-hours).

2.3 Results

2.3.1 Model tuning, calibration and variable importance

Out of the 336 possibilities of hyper-parameter combinations tested, we chose $ntrees = 500$, $mtry = 49$, minimal $nodesize = 1$ and a sampling scheme without replacement with a sample fraction of 0.8. This set of hyper-parameters resulted in the lowest RMSE (0.713) out of all combinations across the location-grouped 10-fold CV. Increasing the number of trees above 500 decreased the RMSE by less than 0.1 %. The range of RMSE values obtained from all hyper-parameter combinations was 0.712 - 0.732. Slightly improved performance with higher numbers of trees and 25 % of the total covariates to consider at any given split ($mtry$) align with the general recommendations of using ensemble decision tree models. Sampling without replacement with sample fractions of 0.8 or lower generally led to lower RMSE values. These results can be explained by the high number of categorical covariates with large differences in the number of classes.

The most important variables of the final model calibration based on permutation were physical geographical maps, followed by geomorphological maps, the AHN (DEM), forest type, land use and temperature maps (Supplement S3, Fig. S2). However, these variable importance measures are not reliable due to high correlation between a large proportion of the covariates (Strobl et al., 2007; Kuhn & Johnson, 2013).

2.3.2 Soil pH maps: mean predictions

Mean prediction maps of soil pH at 25 m resolution varied across the different GSM depth layers (Fig. 2.2). High pH values indicating alkaline soils were found in the marine clay regions, for example in the Southwest (Zeeland) and the regions where land was reclaimed, or "polders" (e.g. Flevoland). Low pH indicating acidic soils were found in sandy areas, e.g. the glacial moraines of the Saalien ice age such as the Utrechtse Heuvelrug and Veluwe regions. In the very south of the Netherlands (Limburg), the model predicted neutral or slightly alkaline soils. This is the only region in the Netherlands that contains calcareous sediments. There was also a distinct pattern along rivers, such as along the Rhine, Maas and IJssel River valleys. Here, the pH was also neutral, reflecting the riverine clays and sediments being deposited along the river banks.

The spatial patterns of the mean predictions over all depth layers suggested that with increasing depth, vegetation and land use played a smaller role (Fig. 2.2). QRF was able to detect the large effect that forested and peatland regions had on soil pH for the uppermost soil layers, revealing acidic conditions in general. With increasing depth, vegetation and land use appeared to become less important and the spatial patterns at these depths resemble geomorphological and parent material indicators. There were a few exceptions to this general pattern in the deepest soil layer (100 - 200 cm) that showed distinct local patterns of low pH values, which might be attributed to regions with thick peat layers.

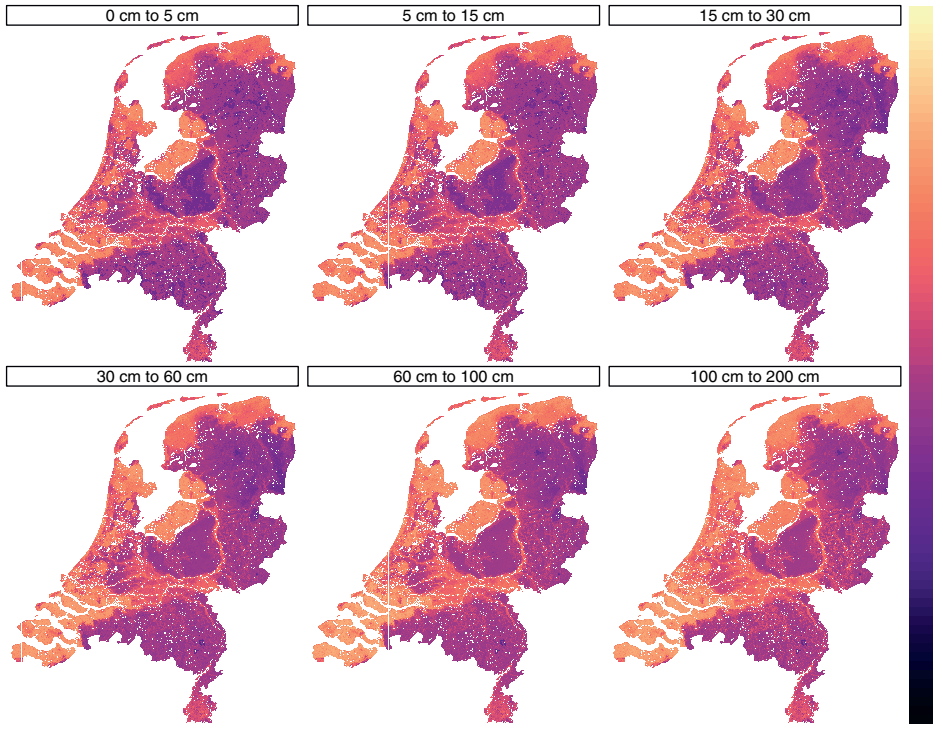


Figure 2.2: Soil pH [KCl] predictions (mean) for every 25 m pixel over the Netherlands for the six depth layers specified by GSM.

Predictions at a high resolution revealed differences in soil pH between and within small agricultural parcels (Fig. 2.3, left).

2.3.3 Soil pH maps: quantiles, PI90 and accuracy thresholds

The maps of quantiles, PI90 and corresponding accuracy thresholds were regarded as a spatially explicit internal accuracy assessment since it quantifies the prediction uncertainty of the calibrated QRF model. For example, between 15 - 30 cm, $q_{0.05}$ showed low (acidic) values almost throughout the Netherlands except for the marine clay regions (Fig. 2.4). $q_{0.95}$ revealed high (alkaline) values of soil pH except for Pleistocene "coversand" regions, which showed values around 6. This pattern was evident across all depth layers (Supplement S4, Fig. S3). In contrast, the loamy riverine and peat areas in the Netherlands were contrasted by much higher uncertainty (Fig. 2.5 and Supplement S4, Fig. S3).

The PI90 and accuracy thresholds thereof indicated large uncertainty of QRF predictions (PI90 $>$ 2.0, i.e quality "none" or "A") for the majority of the Netherlands (Fig. 2.5 & 2.6).

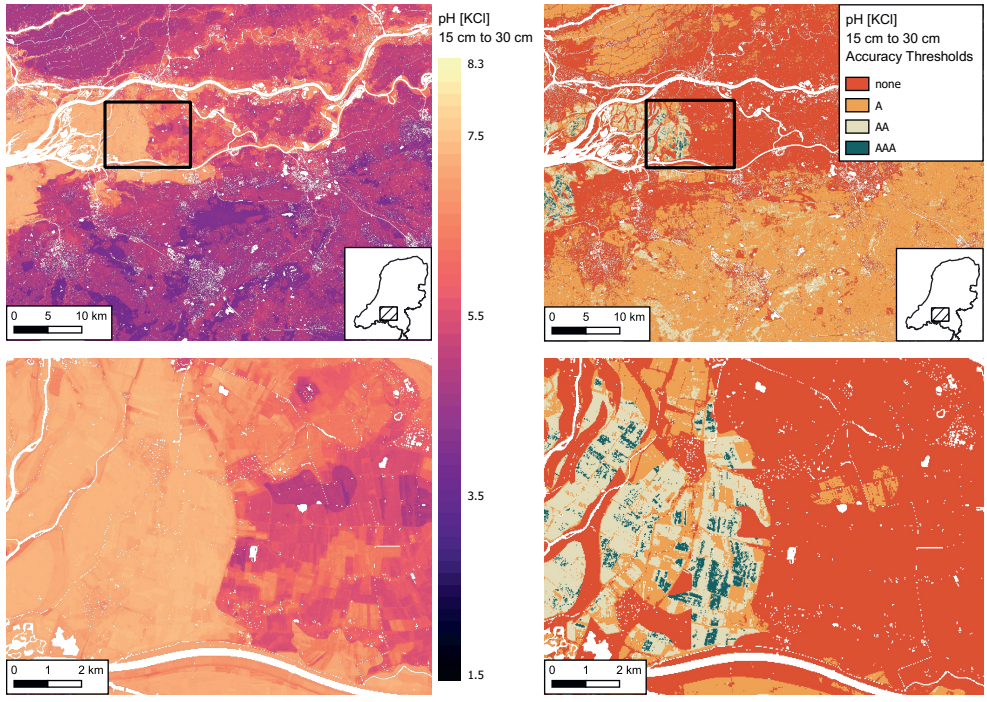


Figure 2.3: Two-stage zoom-in of mean predictions of pH [KCl] (left) and corresponding accuracy thresholds (right) for the depth layer 15 - 30 cm in the southern part of the Netherlands.

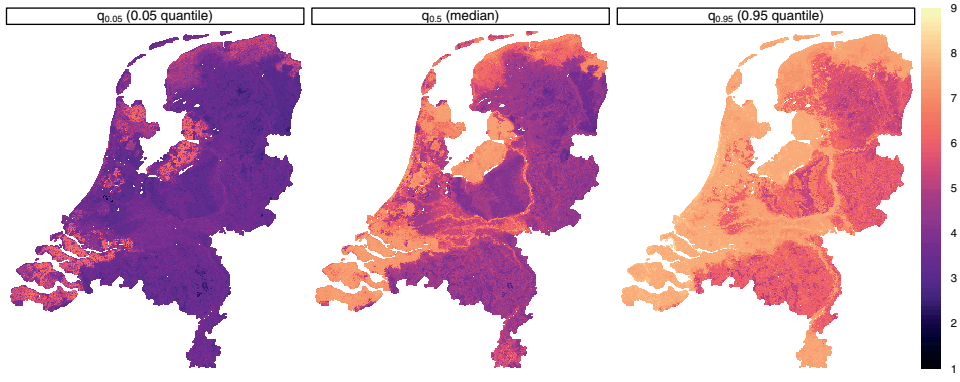


Figure 2.4: $q_{0.05}$ (left), median (middle) and $q_{0.95}$ (right) pH [KCl] for every 25 m pixel over the Netherlands for the depth layer 15 - 30 cm.

Depending on the depth layer, the percentage of pixels for each accuracy threshold ranged between 39.5 % to 64.7 % for "none", 26.5 % to 53 % for A, 2.0 % to 9.0 % for AA and less than 0.2 % for AAA. In general, areas with marine clay soils or sandy soils showed smaller PI90 and better accuracy thresholds. AAA quality was only achieved for a few pixels in the marine clay soils in the depth layers 0 - 5 cm, 5 - 15 cm and especially 15 - 30 cm. With increasing depth, a larger part of the Netherlands did not even achieve the lowest threshold A. However, many sandy soil regions improved from A to AA quality with increasing depth (Fig. 2.6). When zooming in, the maps of accuracy thresholds also revealed differences between and within agricultural parcels (Fig. 2.3, right). Areas with high variation in accuracy thresholds were often not the same areas as areas with high variation in pH predictions.

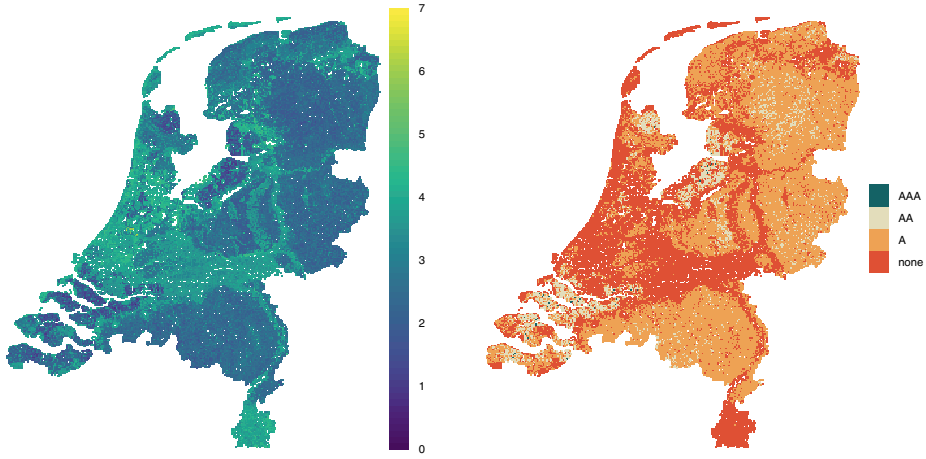


Figure 2.5: The PI90 (left) and corresponding accuracy thresholds (none, A, AA and AAA; right) for the depth layer 15 - 30 cm.

2.3.4 Evaluation of map accuracy using statistical validation

Non-design-based inference

The external accuracy assessment of soil pH maps using non-design-based statistical validation techniques revealed different results between PFB-OOB, PFB-CV and LSK (Fig. 2.7, 2.8 and 2.9). The accuracy plots and metrics of mean predictions over all depth layers combined indicated the best performance using PFB-OOB (ME = 0.01 pH, RMSE = 0.47 pH, MEC = 0.88), followed by PFB-CV (ME = -0.01 pH, RMSE = 0.71 pH, MEC = 0.72) and the worst performance using LSK (ME = 0.11 pH, RMSE = 0.79 pH, MEC = 0.68; Fig. 2.7). Statistical validation using PFB-OOB suggested much higher map accuracy than both PFB-CV and LSK, where residuals were larger, RMSE higher

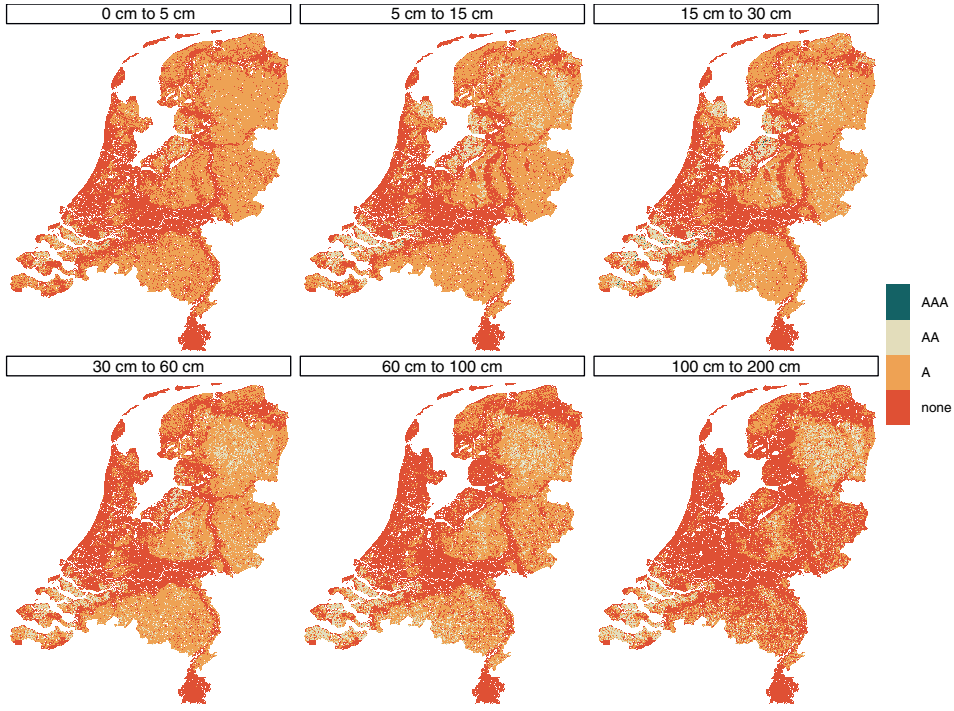


Figure 2.6: pH [KCl] accuracy thresholds (none, A, AA and AAA) for the six depth layers specified by GSM.

and MEC lower. However, we do not recommend the reader to choose the statistical validation method based on apparent performance, as the metrics may not necessarily indicate the "true" map accuracy (Sect. 2.4.1). Both strategies that used the PFB dataset had a ME around zero, indicating an unbiased map, whereas the strategy using the independent validation set (LSK) indicated that QRF systematically under-predicted pH by 0.11 units. For PFB-CV and LSK, observations in the low pH ranges were generally predicted too high, while observations in the high pH ranges were generally predicted too low.

We also found a clear discrepancy between PFB-OOB, PFB-CV and LSK based on accuracy metrics of mean predictions over depth (Fig. 2.8). At PFB locations, the model slightly overpredicted soil pH for the first 5 cm and slightly underpredicted soil pH from 60 cm to 200 cm but was unbiased for the depth layers in between. Statistical validation using LSK resulted in positive ME values for all depth layers and the bias was highest for the depth layers 15 to 30 cm (ME = 0.20), followed by 30 - 60 cm (ME = 0.14) and 60 - 100 cm (ME = 0.11). In contrast to ME results, RMSE values of PFB-OOB and PFB-CV indicated different results for all depth layers. PFB-OOB RMSE results indicated the

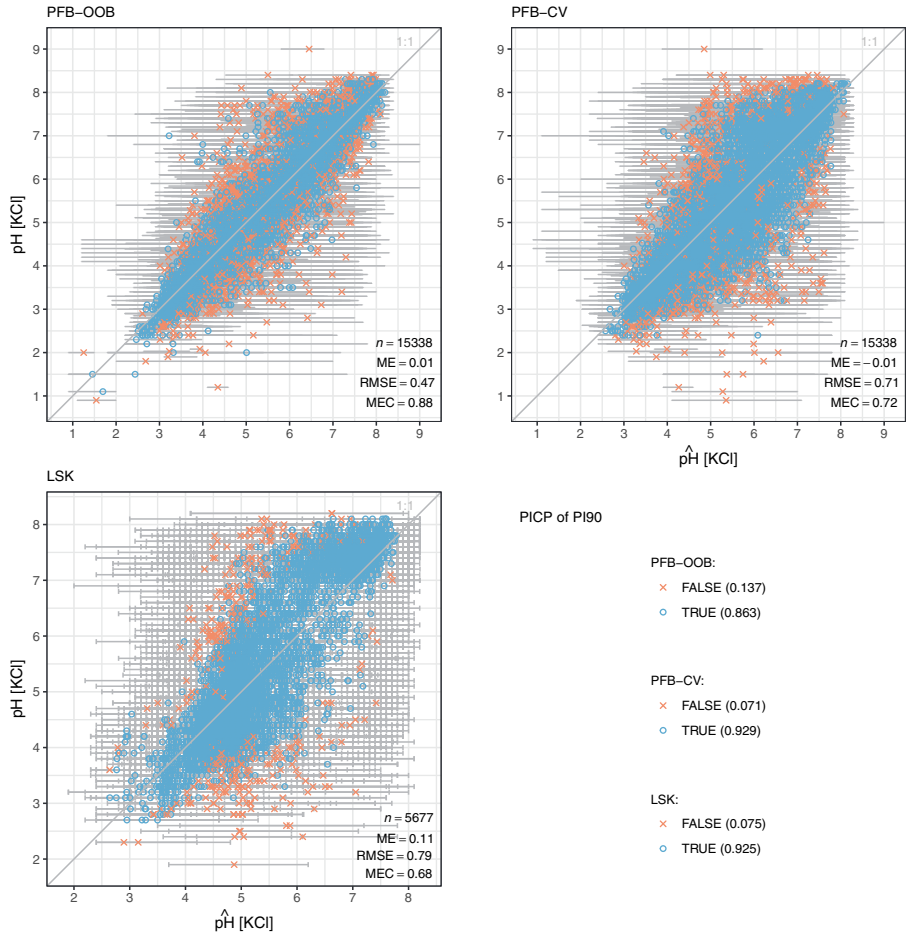


Figure 2.7: Accuracy plot (predicted vs. observed) and metrics of soil pH [KCl] at all depths for PFB-OOB, PFB-CV and LSK. The horizontal grey error bars are the PI90. Blue circles indicate observations within the PI90 (i.e. error bars cross the 1:1 line) and orange crosses indicate observations outside the PI90 (i.e. error bars do not cross the 1:1 line). The PICP of PI90 specifies the percentage of the observations inside (TRUE) or outside (FALSE) the PI90. Notice the different scale of the axes for LSK compared to PFB-OOB and PFB-CV.

best model performance at depths 15 cm to 60 cm (RMSE $\approx$ 0.4). PFB-CV also indicated lower accuracy in depth layers below 60 cm (RMSE = 0.77), but even for the upper 5 cm revealed relatively poor results compared to PFB-OOB (RMSE = 0.62 vs. RMSE = 0.49). In comparison to PFB-OOB and PFB-CV, RMSE values using the LSK were higher for all depth layers but only varied slightly over depth between 0.76 and 0.83. MEC values of PFB-OOB indicated best model fit for predicting at 15 cm to 60 cm (0.90) and lowest

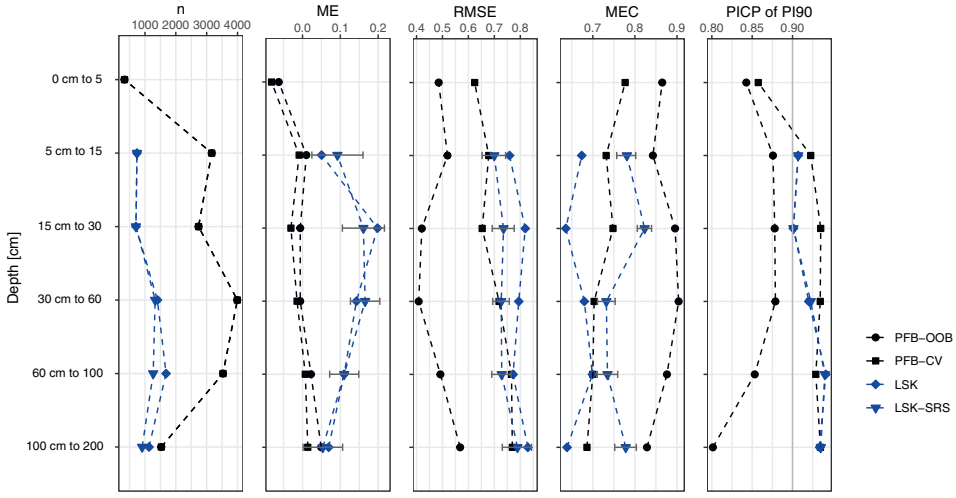


Figure 2.8: Total number of observations (n), ME, RMSE, MEC and PICP of PI90 of the different strategies (PFB-OOB, PFB-CV, LSK and LSK-SRS) over depth. Dashed lines do not represent actual data and are only for visual guidance. Gray error bars indicated the CI95 of LSK-SRS accuracy metrics.

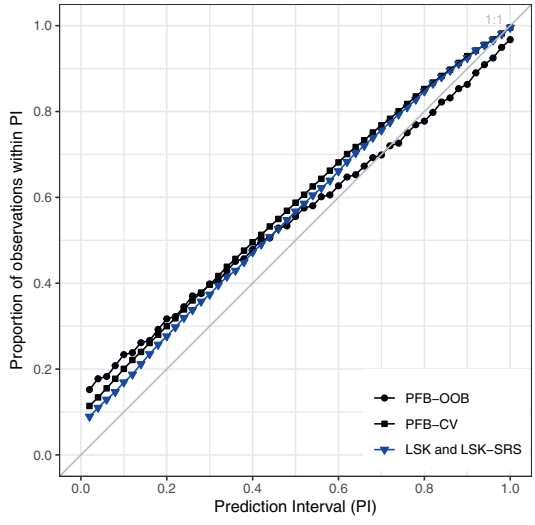


Figure 2.9: PICP of all observations for the different strategies (PFB-OOB, PFB-CV, LSK and LSK-SRS). The PICP of LSK and LSK-SRS are identical because the observations and predictions are the same. Lines connecting the points do not represent actual data and are only for visual guidance.

for 100 - 200 cm (0.83). MEC values using PFB-CV were highest for 0 - 5 cm (0.78) and gradually decreased to 0.69 in the deepest layer. As with ME and RMSE, LSK also indicated relatively poor results of map accuracy over depth based on lower MEC values at all depths between 0.64 and 0.70. Accuracy evaluation using the LSK dataset could not be made for 0 - 5 cm, since there was only one observation with a midpoint within this depth layer (Fig. 2.8).

The PFB-CV and LSK results of the PICP of PI90 (Fig. 2.7 and 2.8) and all other PIs (Fig. 2.9) indicated that QRF prediction uncertainty was slightly overestimated. The PICP of PI90 for PFB-CV and LSK were between 0.925 and 0.929 over all depths (Fig. 2.7) and above 0.90 for most depth layers (Fig. 2.8). Only the uncertainty of the depth layer 0 - 5 cm for PFB-CV was underestimated. The prediction uncertainty based on PFB-CV and LSK was also overestimated for all remaining PIs (Fig. 2.9). In contrast, the PICP of PI90 for PFB-OOB revealed a clear underestimation of the prediction uncertainty (0.863; Fig. 2.7), especially at depth layers 0 - 5 cm and 60 cm to 200 cm (Fig. 2.8). Based on the PFB-OOB evaluation, the 0.70 PI of QRF was the most accurate, wherein lower PIs overestimated the prediction uncertainty and higher PIs underestimated the prediction uncertainty (Fig. 2.9).

Design- vs. non-design-based inference

The statistical validation of the design-based inference (LSK-SRS) resulted in ME values between 0.09 and 0.17, RMSE values between 0.70 and 0.79 and MEC values between 0.73 and 0.82, depending on the depth layer (Fig. 2.8). LSK-SRS indicated higher map accuracy compared to LSK (ME = 0.05 - 0.14, RMSE = 0.76 - 0.83 and MEC = 0.64 - 0.70). The small differences in the number of observations (n) used in the statistical validation for LSK and LSK-SRS were because only one observation was used per depth layer for LSK-SRS (Sect. 2.2.6, Table 2.4). Similarly to LSK, evaluation of the map using LSK-SRS also showed biased results, following a similar pattern over depth. The ME values of LSK were within the CI95 of the ME of LSK-SRS. RMSE values also followed the pattern of LSK over depth. However, the values indicated a higher map accuracy and were closer to the PFB-CV metrics. The RMSE values of LSK were outside the CI95 range except for the depth interval from 100 - 200 cm, but the RMSE metrics of PFB-CV were mostly within this range. The MEC metrics and their respective CI95s of LSK-SRS indicated better mapping accuracy results than PFB-CV and LSK. The CI95 did not overlap with MEC metrics from other approaches. Overall, the CI95 were narrow for both RMSE and MEC metrics, indicating a high certainty of RMSE and MEC values. The PICP metrics of LSK and LSK-SRS were identical because the observations and quantile predictions were the same (Fig. 2.8 and 2.9). In summary, metrics across all four strategies (PFB-OOB, PFB-CV, LSK and LSK-SRS) varied by respective minimum and maximum values over all depths between ME = -0.03 and 0.20, RMSE = 0.42 and 0.82, MEC = 0.64 and 0.90 and PICP = 0.80 and 0.93.

2.4 Discussion

2.4.1 Map accuracy using statistical validation strategies

The large differences depending on the external accuracy assessment strategy used (PFB-OOB, PFB-CV, LSK or LSK-SRS; Table 2.3) emphasize that map accuracy depends largely on the statistical validation approach. Different approaches can yield substantially different indications of the same map's quality. Hence, the statistical validation strategy needs to be chosen carefully. Based on sampling theory (Cochran, 1977; de Gruijter et al., 2006; Gregoire & Valentine, 2007), maps should be validated with a design-based approach using a probability sample whenever possible (Brus et al., 2011). Therefore, LSK-SRS may be regarded as the best estimate of the "true" map accuracy in our study and is further advantageous because the CI95 also quantifies the accuracy of the estimated metrics (Fig. 2.8). Nonetheless, we acknowledge that even the uncertainty of design-based metrics (CI95) are themselves imperfect and prone to uncertainty, as numerical experiments using pseudo values have shown (Lagacherie et al., 2019).

However, a slight disadvantage of LSK-SRS in our study is that some observations had to be removed or averaged and some depth layers had no observations for specific strata (Table 2.4). The depth layer 0 - 5 cm was not evaluated and the metrics of depth layers 5 - 15 cm, 15 - 30 cm and 100 - 200 cm only pertain to between 95.66 % and 98.51 % of the Netherlands. This may be avoided in other studies by considering the soil in 3D space when planning the sampling design and deciding which target depth layers to map beforehand. Such an approach may be planned in a similar way as for design-based inference of spatio-temporal models, where the probability sample needs to include both locations and time (Brus, 2014). In the case of the Netherlands, the LSK sampling campaign was planned before the standard GSM depth layers were defined (Finke et al., 2001; Visschers et al., 2007).

Both LSK and LSK-SRS strategies suggest that soil pH maps are positively biased, i.e. systematically under-predicted pH, with ME = 0.05 to 0.17 (Fig. 2.8). Depth layers between 15 cm to 60 cm showed the largest bias. This may be due to the difference in distribution of the PFB (calibration) vs. LSK (validation) data (Fig. 2.1). The relatively large peak in observations around 4.5 pH in PFB in comparison to the peaks for LSK results in a lower average in the calibration data. Ensemble decision trees tend to predict the average well while performance decreases towards the tails of the distribution. This is most likely due to averaging of trees in the forest (e.g. Hengl et al., 2018). Hence, the calibrated QRF using PFB data possibly led to overall biased predictions at LSK locations. The values of ME for LSK-SRS according to sampling theory also indicate that the predictions are not only biased at validation locations, but for all of the Netherlands. Such a bias may be avoided by using a representative dataset of all of the Netherlands (e.g. probability sample) also for calibration, not only for validation. On one hand, an-

other possible reason for the bias may be that PFB and LSK data originate from different time periods (Supplement S1, Fig. S1). On the other hand, field and lab methods and protocols remained the same for both PFB and LSK data, so a bias solely due to the year of soil sampling and analysis is unlikely.

We found that QRF was able to detect changes in pH over depth, as indicated by the qualitative evaluation of the spatial patterns (Fig. 2.2 and Sect. 2.3.2), and that accuracy slightly decreased over depth, as indicated by the design-based inference (Fig. 2.8). RMSE increased from 0.70 (5 - 15 cm) to 0.73 - 0.74 (15 cm to 100 cm) to 0.79 (100 - 200 cm). However, the MEC and ME did not indicate lower accuracy over depth. Performance often decreases with depth in DSM studies due to fewer observations and fewer covariates available indicative of soil conditions at lower depths (Keskin & Grunwald, 2018).

Using LSK-SRS as a reference, PFB-CV was most indicative of map quality out of the non-design-based inference strategies. RMSE, MEC and PICP values of PFB-CV were closest to those of LSK-SRS (Fig. 2.8 and 2.9). OOB validation for ensemble decision tree models without grouping soil profile locations overestimated map accuracy. This is supported by Meyer et al. (2018), who show that without leaving out all observations from entire locations, model accuracy metrics are overly optimistic. In contrast, validation using the independent dataset (LSK) without design-based inference was too pessimistic over all depth layers based on RMSE and MEC values (Fig. 2.8). This may be because when ignoring the probability sample, observations from small (niche) strata, where the predictive performance is likely worse, are oversampled and frequently occurring strata, where predictive performance is likely better, are undersampled in comparison to their relative occurrence in the study area. In summary, using either PFB-OOB or not accounting for the LSK sample design both resulted in misleading map accuracy metrics.

For other studies that do not have the resources to validate using a probability sample, the location-grouped k-fold CV used here (PFB-CV) may be further refined to estimate map accuracy. Random k-fold CV, even when grouped by location, potentially still leads to biased estimates since the data are often clustered or unevenly distributed (Brenning, 2012; Schratz, 2019). When data are clustered or unevenly distributed, data dense areas are weighed more than sparsely sampled areas in random k-fold CV. Thus, the CV indicates how accurate the model is at predicting the sampled data, but not necessarily the area. To overcome these challenges, we recommend to use weighted CV: a form of random k-fold CV where the dataset is resampled into multiple datasets based on point density (van Ebbenhorst Tengbergen, 2021). Weighted CV was not tested here due to the availability of a probability sample to perform design-based inference (LSK-SRS). We do not recommend spatial partitioning, i.e. spatial CV, or the use of buffers in CV (e.g. Brenning, 2005, 2012; Le Rest et al., 2014; Pohjankukka et al., 2017; Roberts et al., 2017; Ploton et al., 2020; Hengl et al., 2021), as these are not theoretically sound and are likely systematically and potentially severely over-pessimistic (Wadoux et al., 2021a).

2.4.2 Comparison to other soil pH maps

The main improvements of our soil pH maps for the Netherlands compared to other GSM products are a better estimation of the "true" map accuracy using design-based inference and the high spatial resolution (25 m). Here, we compare our design-based LSK-SRS accuracy metrics with those of other studies. However, it is important to note that other studies did not use probability samples and design-based inference for validation and so results are not directly comparable. Accuracy metrics of studies using a non-probability sample may be considered overly optimistic in the case of clustered data (Wadoux et al., 2021a).

Our maps have similar patterns as previous soil pH maps of the Netherlands. Although an initial map for soil pH from 0 cm to 25 cm was made for the Netherlands using a soil type map and co-kriging, mean predictions are difficult to compare because no statistical validation was done (Brus & Heuvelink, 2007; Brus et al., 2009). Nevertheless, qualitative evaluation of the spatial patterns reveals strong resemblance of mean predictions (Fig. 2.4; Brus et al., 2009, p. 19).

For Denmark, another Northern European country similar in both size and soil variability to the Netherlands, Adhikari et al. (2014b) used a hybrid model consisting of Cubist followed by local point kriging of the residuals and validated maps using 25 % of samples. Results were only reported for the depth layer with the best performance (5 - 15 cm), for which an R^2 of 0.46 and RMSE of 0.61 were achieved. Our MEC value, which is comparable to R^2, is better (0.78) while our RMSE is higher (0.70) for the same depth layer using design-based inference.

On average, our results for map accuracy using design-based inference are also comparable to other recent GSM products that used ensemble decision trees (RF or QRF) to model soil pH. For example, Chen et al. (2019) predicted topsoil (0 cm to 20 cm) pH for China using RF and assessed the accuracy with a random 10-fold CV (not grouped by location). We attained comparable results for depths 5 cm to 30 cm in terms of RMSE (0.70 to 0.74) compared to their study (0.72), but our MEC values were higher than their R^2 values (MEC = 0.78 - 0.82 vs. 0.71).

We also compared our results to the recent SoilGrids version 2.0 (Poggio et al., 2021), for which we compared global metrics as well as prediction performance in the Netherlands. SoilGrids 2.0 also used QRF but assessed map accuracy using a CV procedure based on spatial stratification. When comparing global metrics, we attained better results for depths 5 cm to 100 cm (RMSE = 0.70 to 0.74; MEC = 0.73 - 0.82) compared to their values over all depths (RMSE = 0.77 pH (water), MEC = 0.66 - 0.69; Poggio et al., 2021), although for the depth layer 100 - 200 cm, our RMSE was higher (0.79). We attained identical PICP of PI90 results as SoilGrids 2.0 for depths 5 - 15 cm and 15 - 30 cm (0.91 and 0.90, respectively), but slightly poorer results for 30 cm - 200 cm (0.92 - 0.94 vs. 0.89 -

0.91). However, we achieved much higher performance, for example for the depth layer 5 cm to 15 cm (ME = 0.09, RMSE = 0.70, MEC = 0.78, PICP of PI90 = 0.91) compared to SoilGrids (ME = -0.74, RMSE = 1.23, MEC = 0.34, PICP of PI90 = 0.71) when evaluating prediction performance in the Netherlands using LSK-SRS design-based inference. These magnitudes of differences in predictive performance between our maps and SoilGrids for the Netherlands were consistent when comparing all depth layers. These results align with Mulder et al. (2016b) and Chen et al. (2019), who compared their national products with the older version of SoilGrids (Hengl et al., 2017b) and also achieved higher accuracy.

2.4.3 Uncertainty using QRF

The PICP obtained from the QRF quantiles reveal that PI's are generally too large and hence prediction uncertainty was overestimated. For applications, this means that users can be slightly more certain than indicated because the PI90 in reality appears to cover more than 90 % of the observations (Fig. 2.8 and 2.9). Our maps of $q_{0.05}$ and $q_{0.95}$ for depths 0 cm to 30 cm (e.g. Fig. 2.4) match those of Brus et al. (2009, p. 20), who used co-kriging based on a soil type map of the Netherlands. This suggests that in this case, the uncertainty quantification using QRF was similar to the kriging variance. However, other studies have often found large differences in the spatial distribution of the PI when comparing QRF and kriging (Vaysse & Lagacherie, 2017; Baake, 2018; Szatmári & Pásztor, 2019).

A modelling approach using QRF can make use of the flexibility and predictive performance of machine learning while still attaining an estimate of prediction uncertainty in a general context. However, the limitations of prediction methods such as QRF are that they do not deliver knowledge of the different sources of this uncertainty. Recent studies have developed approaches to either quantify uncertainty of data used as model inputs in DSM, such as the measurement errors of soil observations (van Leeuwen et al., 2021) or covariates, or how these errors can be incorporated in machine learning algorithms such as RF (van der Westhuizen et al., 2022).

There are many possible sources that may have contributed to the QRF prediction uncertainty, such as the inability of the covariates to explain all soil pH variation, lab measurement errors and the temporal variation of soil pH over time. We used legacy data from 1953-2012 (Supplement S1, Fig. S1), but ignored time even though soil pH may have changed over the decades. Even within one year, soil pH varies with season and soil moisture content, with higher pH values associated with wetter soils and winter conditions and lower pH values with drier soils and summer conditions (Miller & Kissel, 2010; Robinson et al., 2017). However, differences that might be expected due to soil pH temporal variation (e.g. 0.5 pH units) are only a major source of uncertainty in areas for which the PI90 is very low (e.g. AAA pixels). Therefore, such temporal variation is smaller than the vast

majority of uncertainty quantified here. In agreement with Arrouays et al. (2017), there is a need to address soil measurement age and time in future DSM studies. Accounting for time may potentially improve prediction accuracy by removing this source of uncertainty. Moreover, it potentially estimates how pH has changed over time for different parts of the Netherlands.

2.4.4 Accuracy thresholds for Tier 4 GSM maps and user applications

Using the GSM accuracy thresholds for Tier 4 products for the PI90, the large majority of the Netherlands was designated A or "none" quality. We believe that there are several reasons for this. Firstly, as indicated by the PICP of PI90 (Fig. 2.7, 2.8 and 2.9), the uncertainty quantification using QRF was overestimated, meaning that overall, slightly better accuracy threshold designation can be expected (e.g. less "none" and more AAA). Secondly, the accuracy thresholds are also dependent on the spatial support. Uncertainty of predictions at block support is typically smaller than for point predictions because within-block variation is averaged out. The degree of uncertainty reduction depends on the degree of within-block spatial variation and therefore uncertainty reduction by spatial aggregation can only be computed if the spatial correlation is included in the model, e.g. as in Szatmári et al. (2021). Thus, we expect more AAA areas with increasing spatial support. Thirdly and most importantly, the achievability of AA or AAA accuracy thresholds are largely dependent on the size and variability of soil observations in the study area. Higher accuracy thresholds can generally be achieved in study areas where there is less variation while mostly lower accuracy thresholds are achieved in areas where there is a large variation. If users require AA or AAA accuracy for their intended use, we recommend to conduct a local or regional mapping study where there is less variation (e.g. sandy soils) and to increase the sampling density.

Based on our results, AA and AAA thresholds are difficult to achieve for national maps and we are curious whether other countries will obtain similar results. We think it is important that accuracy thresholds remain ambitious because thresholds below "none" would imply such a high uncertainty that it would most likely be meaningless for user applications. We hope that our results will lead to a discussion that includes end-users about the uncertainty ranges of the GSM accuracy thresholds.

We believe that accuracy thresholds as used here have several advantages. Unlike statistical validation, which can only make use of observations at sampled locations, they are spatially explicit and can be designated to each pixel, as is also the case for other uncertainty measures (Heuvelink, 2014, 2018). Moreover, we believe accuracy thresholds are easier to communicate with end-users than other widely used uncertainty metrics, e.g. PI90. A user merely has to know the quality required for their specific application and then look at the map of four possible thresholds. Note that maps are not only useful where there are high quality pixels; many users may only require e.g. A quality.

National-scale measures, legislation or projects that are based on soil information can easily be applied specifically to areas above a certain threshold. For agricultural applications for example, our maps may potentially be useful in order to consider uncertainty for liming recommendations (Libohova et al., 2019). Lark et al. (2014) used a similar idea to communicate to users where critical trace element values might approach agronomically important thresholds.

Our modelling framework is convenient for a variety of users that require spatially explicit soil pH information and associated uncertainty with quantified accuracy at 25 m resolution for any desired depth anywhere in the Netherlands. If users are interested in the overall pH map accuracy for the Netherlands, we recommend to use the LSK-SRS (and PFB-CV if between 0 - 5 cm) metrics. If users are interested in a small target area within the Netherlands or require spatially explicit accuracy measures, then we recommend the use of PI90 and accuracy threshold maps. Given that the accuracy is within an acceptable range (for a given target area), the high resolution maps may be used for local and small-scale land use planning and management. In this regard, we hope that these soil pH maps are useful for the Dutch Ministries for Agriculture, Nature and Food Quality, Economic Affairs and Climate Policy, Infrastructure and Water Management, governmental water-boards, as well as farmers, researchers from different fields and non-profit organisations. The reproducible, efficient and flexible computational workflow may also make it attractive to generate future maps of other target soil properties for the European Joint Program (*EJP*) on agricultural soil management of the European Union (Keesstra et al., 2021) or the Global Soil Partnership (*GSP*) of the Food and Agriculture Organization of the United Nations (FAO).

2.5 Conclusion

This study contributed to the GSM project by providing soil pH prediction maps for the Netherlands at 25 m resolution, at six standard depth layers (0 - 5 cm, 5 - 15 cm, 15 - 30 cm, 30 - 60 cm, 60 - 100 cm and 100 - 200 cm), yet the calibrated model allows prediction at any user-required depth. We compared non-design-based to design-based external accuracy assessment strategies using ME, RMSE, MEC, PI90 and PICP metrics. Among these statistical validation methods, the probability sample available in the Netherlands presented a unique opportunity for accuracy assessment using design-based inference (LSK-SRS). Consequently, we were able to provide unbiased estimates of the "true" map quality and quantify the accuracy of these estimates with confidence intervals. We used a robust, reproducible and data-driven DSM workflow that uses QRF to quantify spatially explicit uncertainty as an internal accuracy assessment. In addition, these are to our knowledge the first Tier 4 GSM maps, since they also provide spatially explicit accuracy thresholds as quality rankings. We attained A and "none" quality accuracy thresholds for the large majority of the Netherlands and therefore, we call upon future studies to also test

whether the highest Tier 4 GSM quality rankings are difficult to achieve for national scale soil maps. We hope that our soil pH maps are useful for national agencies and initiatives and expect that with modest modification our workflow can be applied to other soil properties and other areas in the world to meet the increasing demand for spatial soil information.

2.6 Appendix A. GSM accuracy thresholds

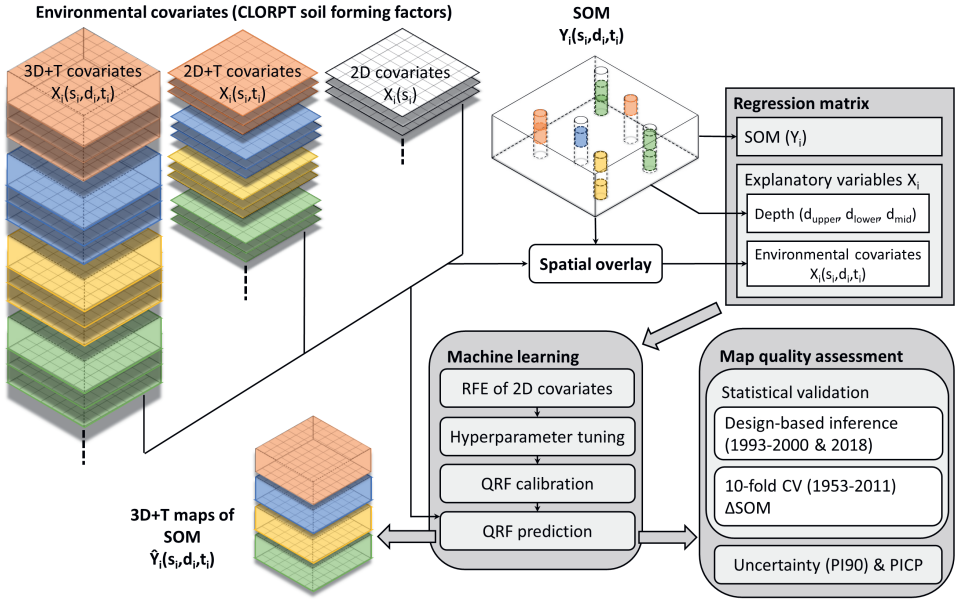
Table 2.5: These tabular data are referred to in Arrouays et al. (2015) but to the best of our knowledge has not been published. The specifications for the response of this study, soil pH (multiplied by 10), is shown in bold.

Soil property	**Unit**	**A**	**AA**	**AAA**
		Depth $0-5$		
Depth to rock	cm	Mean $\pm$ 50% (Mean)	Mean $\pm$ 30% (Mean)	Mean $\pm$ 15% (Mean)
Plant Exploitable (Effective) Depth	cm	Mean $\pm$ 50% (Mean)	Mean $\pm$ 30% (Mean)	Mean $\pm$ 15% (Mean)
SOC	g/kg	Mean $\pm$ 60% (Mean)	Mean $\pm$ 35% (Mean)	Mean $\pm$ 15% (Mean)
pH x 10		**($\pm$) 15**	**($\pm$) 10**	**($\pm$) 5**
Clay	g/kg	Mean $\pm$ 40% (Mean)	Mean $\pm$ 25% (Mean)	Mean $\pm$ 15% (Mean)
Silt	g/kg	Mean $\pm$ 40% (Mean)	Mean $\pm$ 25% (Mean)	Mean $\pm$ 15% (Mean)
Sand	g/kg	Mean $\pm$ 40% (Mean)	Mean $\pm$ 25% (Mean)	Mean $\pm$ 15% (Mean)
Coarse fragments	m^3/m^3	Mean $\pm$ 35% (Mean)	Mean $\pm$ 25% (Mean)	Mean $\pm$ 15% (Mean)
ECEC	$mmol_c/kg$	Mean $\pm$ 40% (Mean)	Mean $\pm$ 30% (Mean)	Mean $\pm$ 15% (Mean)
		Depth $5-15$		
Depth to rock	cm	Mean $\pm$ 50% (Mean)	Mean $\pm$ 30% (Mean)	Mean $\pm$ 15% (Mean)
Plant Exploitable (Effective) Depth	cm	Mean $\pm$ 50% (Mean)	Mean $\pm$ 30% (Mean)	Mean $\pm$ 15% (Mean)
SOC	g/kg	Mean $\pm$ 50% (Mean)	Mean $\pm$ 30% (Mean)	Mean $\pm$ 15% (Mean)
pH x 10		**($\pm$) 15**	**($\pm$) 10**	**($\pm$) 5**
Clay	g/kg	Mean $\pm$ 40% (Mean)	Mean $\pm$ 25% (Mean)	Mean $\pm$ 15% (Mean)
Silt	g/kg	Mean $\pm$ 40% (Mean)	Mean $\pm$ 25% (Mean)	Mean $\pm$ 15% (Mean)
Sand	g/kg	Mean $\pm$ 40% (Mean)	Mean $\pm$ 25% (Mean)	Mean $\pm$ 15% (Mean)
Coarse fragments	m^3/m^3	Mean $\pm$ 40% (Mean)	Mean $\pm$ 30% (Mean)	Mean $\pm$ 15% (Mean)
ECEC	$mmol_c/kg$	Mean $\pm$ 40% (Mean)	Mean $\pm$ 30% (Mean)	Mean $\pm$ 15% (Mean)
		Depth $15-30$		
Depth to rock	cm	Mean $\pm$ 50% (Mean)	Mean $\pm$ 30% (Mean)	Mean $\pm$ 15% (Mean)
Plant Exploitable (Effective) Depth	cm	Mean $\pm$ 50% (Mean)	Mean $\pm$ 30% (Mean)	Mean $\pm$ 15% (Mean)
SOC	g/kg	Mean $\pm$ 50% (Mean)	Mean $\pm$ 30% (Mean)	Mean $\pm$ 15% (Mean)
pH x 10		**($\pm$) 15**	**($\pm$) 10**	**($\pm$) 5**
Clay	g/kg	Mean $\pm$ 40% (Mean)	Mean $\pm$ 25% (Mean)	Mean $\pm$ 15% (Mean)
Silt	g/kg	Mean $\pm$ 40% (Mean)	Mean $\pm$ 25% (Mean)	Mean $\pm$ 15% (Mean)
Sand	g/kg	Mean $\pm$ 40% (Mean)	Mean $\pm$ 25% (Mean)	Mean $\pm$ 15% (Mean)
Coarse fragments	m^3/m^3	Mean $\pm$ 40% (Mean)	Mean $\pm$ 30% (Mean)	Mean $\pm$ 15% (Mean)
ECEC	$mmol_c/kg$	Mean $\pm$ 40% (Mean)	Mean $\pm$ 30% (Mean)	Mean $\pm$ 15% (Mean)

Depth 30 − 60				
Depth to rock	cm	Mean ± 50% (Mean)	Mean ± 30% (Mean)	Mean ± 15% (Mean)
Plant Exploitable (Effective) Depth	cm	Mean ± 50% (Mean)	Mean ± 30% (Mean)	Mean ± 15% (Mean)
SOC	g/kg	Mean ± 40% (Mean)	Mean ± 25% (Mean)	Mean ± 15% (Mean)
pH x 10		**(±) 15**	**(±) 10**	**(±) 5**
Clay	g/kg	Mean ± 40% (Mean)	Mean ± 25% (Mean)	Mean ± 15% (Mean)
Silt	g/kg	Mean ± 40% (Mean)	Mean ± 25% (Mean)	Mean ± 15% (Mean)
Sand	g/kg	Mean ± 40% (Mean)	Mean ± 25% (Mean)	Mean ± 15% (Mean)
Coarse fragments	m^3/m^3	Mean ± 40% (Mean)	Mean ± 30% (Mean)	Mean ± 15% (Mean)
ECEC	$mmol_c/kg$	Mean ± 40% (Mean)	Mean ± 30% (Mean)	Mean ± 15% (Mean)
Depth 60 − 100				
Depth to rock	cm	Mean ± 50% (Mean)	Mean ± 30% (Mean)	Mean ± 15% (Mean)
Plant Exploitable (Effective) Depth	cm	Mean ± 50% (Mean)	Mean ± 30% (Mean)	Mean ± 15% (Mean)
SOC	g/kg	Mean ± 40% (Mean)	Mean ± 25% (Mean)	Mean ± 15% (Mean)
pH x 10		**(±) 15**	**(±) 10**	**(±) 5**
Clay	g/kg	Mean ± 40% (Mean)	Mean ± 25% (Mean)	Mean ± 15% (Mean)
Silt	g/kg	Mean ± 40% (Mean)	Mean ± 25% (Mean)	Mean ± 15% (Mean)
Sand	g/kg	Mean ± 40% (Mean)	Mean ± 25% (Mean)	Mean ± 15% (Mean)
Coarse fragments	m^3/m^3	Mean ± 50% (Mean)	Mean ± 40% (Mean)	Mean ± 20% (Mean)
ECEC	$mmol_c/kg$	Mean ± 40% (Mean)	Mean ± 30% (Mean)	Mean ± 15% (Mean)
Depth 100 − 200				
Depth to rock	cm	Mean ± 50% (Mean)	Mean ± 30% (Mean)	Mean ± 15% (Mean)
Plant Exploitable (Effective) Depth	cm	Mean ± 50% (Mean)	Mean ± 30% (Mean)	Mean ± 15% (Mean)
SOC	g/kg	Mean ± 40% (Mean)	Mean ± 25% (Mean)	Mean ± 15% (Mean)
pH x 10		**(±) 15**	**(±) 10**	**(±) 5**
Clay	g/kg	Mean ± 40% (Mean)	Mean ± 25% (Mean)	Mean ± 15% (Mean)
Silt	g/kg	Mean ± 40% (Mean)	Mean ± 25% (Mean)	Mean ± 15% (Mean)
Sand	g/kg	Mean ± 40% (Mean)	Mean ± 25% (Mean)	Mean ± 15% (Mean)
Coarse fragments	m^3/m^3	Mean ± 50% (Mean)	Mean ± 40% (Mean)	Mean ± 20% (Mean)
ECEC	$mmol_c/kg$	Mean ± 40% (Mean)	Mean ± 30% (Mean)	Mean ± 15% (Mean)

Chapter 3

Three-dimensional space and time mapping reveals soil organic matter decreases across anthropogenic landscapes in the Netherlands



This chapter is based on:

Helfenstein, A., Mulder, V.L., Heuvelink, G.B.M., Hack-ten Broeke, M.J.D., 2024. Three-dimensional space and time mapping reveals soil organic matter decreases across anthropogenic landscapes in the Netherlands. *Communications Earth & Environment* 5, 1–16. `https://doi.org/10.1038/s43247-024-01293-y`

Abstract

For restoring soil health and mitigating climate change, information of soil organic matter is needed across space, depth and time. Here we developed a statistical modelling platform in three-dimensional space and time as a new paradigm for soil organic matter monitoring. Based on 869 094 soil organic matter observations from 339 231 point locations and the novel use of environmental covariates variable in three-dimensional space and time, we predicted soil organic matter and its uncertainty annually at 25 m resolution between 0 - 2 m depth from 1953 - 2022 in the Netherlands. We predicted soil organic matter decreases of more than 25 % in peatlands and 0.1 - 0.3 % in cropland mineral soils, but increases between 10 - 25 % on reclaimed land due to land subsidence. Our analysis quantifies the substantial variations of soil organic matter in space, depth, and time, highlighting the inadequacy of evaluating soil organic matter dynamics at point scale or static mapping at a single depth for policymaking.

3.1 Introduction

Soil organic matter (SOM) plays a crucial role in achieving multiple Sustainable Development Goals (SDGs) (United Nations, 2015; Keesstra et al., 2016; Jónsson et al., 2016; Lorenz et al., 2019; Smith et al., 2021), in particular SDG target 2.4 on sustainable food production and resilient agricultural practices and target 15.3 on land degradation neutrality by 2030. Furthermore, SOM is linked to six of the eight mission objectives of the Soil Deal for Europe (European Commission, 2021), which aims to fulfill European and international commitments to the SDGs. In 2023, the European Commission underscored its commitments through the Directive on Soil Monitoring and Resilience (European Commission, 2023b), a legislative proposal where soil health is defined as "the continued capacity of soils to support ecosystem services" (European Commission, 2021).

Besides being essential for soil health, SOM offers an opportunity for climate change mitigation through carbon sequestration (Chabbi et al., 2017; Amelung et al., 2020). Studies have shown that SOM increase is feasible under best management practices, and consequentially, the "4 per mille Soils for Food Security and Climate" was launched with the aspiration to increase global SOM stocks by 4 per 1000 (or 0.4 %) per year (Minasny et al., 2017). Coordinated efforts are underway to develop best practices for measuring, reporting, and verifying SOM changes (Smith et al., 2020), while simultaneously adapting agricultural systems to facilitate carbon farming as a means of mitigating greenhouse gas emissions (Sharma et al., 2021).

However, while it is perhaps too early to determine the effectiveness of international commitments, SOM continues to decrease in European croplands (Fernández-Ugalde et al., 2020) and peatlands (Leifeld & Menichetti, 2018). It is therefore no surprise that the increase of SOM and conservation of peat soils remains the main challenge related to soil health (Vanino et al., 2023).

With its intensive agriculture, degraded peatlands, and highly anthropogenic landscapes, the Netherlands is an ideal case for examining SOM changes to address the priority soil health challenge, "4 per mille", the Soil Deal and SDG targets. Situated in Europe's largest delta, the Rhine-Meuse-Scheldt delta, agriculture in the Netherlands is widely regarded as the most intensive in Europe (Debonne et al., 2022). Before the start of agriculture in the low lying regions less than 2000 years ago, more than 50 % of what is now the Netherlands was covered in peat (Vos et al., 2020; Erkens et al., 2016). Through drainage, excavation and/or agricultural use of peatlands, this has now been reduced to 15 % (Fig. 3.1d & e; SI1[1]). The Netherlands is largely composed of anthropogenic

[1]Supplementary information (SI) of Chapter 3 is available at `https://doi.org/10.1038/s43247-024-01293-y` under "Supplementary information".

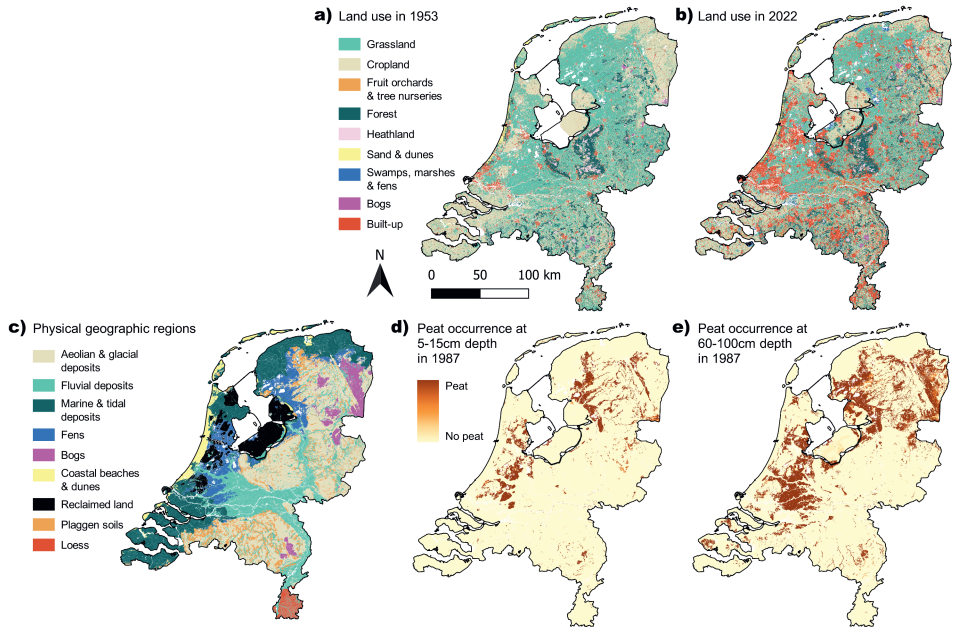


Figure 3.1: Land use map of the Netherlands in 1953 (a) and 2022 (b), main physical geographic regions (c) and peat occurrence in 1987 at 5 - 15 cm (d) and 60 - 100 cm depth (e). See Sect. 3.3.2 for how maps of dynamic covariates were derived for all years between 1953 and 2022 and SI1 for information on plaggen soils shown in c.

landscapes: historic land reclamation (17 %[2]; Fig. 3.1c), conversion to urban areas (15 %; Fig. 3.1a & b) and re-landscaping of new nature and recreational areas have had a tremendous impact on the soils.

The challenge of increasing SOM for restoring soil health and fulfilling international commitments requires high resolution, spatio-temporally explicit SOM assessment to facilitate management practices and land use decisions tailored to local soil conditions. To address this challenge, we developed a modelling platform in 3D space and time (3D+T) as a new paradigm for SOM monitoring and mapping. It provides annual predictions of SOM and its uncertainty in the Netherlands, at 25 m resolution at point support between 0 m and 2 m depth from 1953 (first measurements; Fig. 3.2) to 2022. We used machine-learning, 869 094 SOM observations from 339 231 point locations (approximately 10 locations per km^2; Fig. 3.2 & Table 3.1) and spatially explicit environmental covariates. Using quantile regression forest (Meinshausen, 2006), the median of the predicted probability distribution was taken as the predicted SOM value while the 90th prediction interval (PI90) indicates prediction uncertainty (Sect. 3.3.8). The covariates were either static (2D), variable in

[2]Percentages of present day surface area

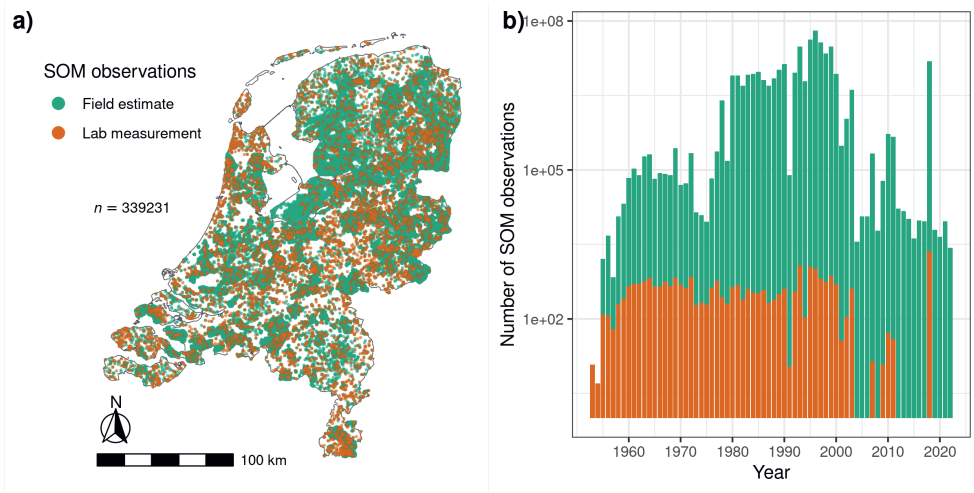


Figure 3.2: Map showing the 339 231 locations with SOM observations (a) and their temporal distribution (b). Laboratory measurements are shown in orange and field estimates are shown green. Note log-scale of y-axis.

time (2D+T) (Heuvelink et al., 2020; OpenGeoHub et al., 2022) or variable in 3D+T, and serve as proxies for soil-forming factors. While climate, relief, and parent material were considered static, land use (Fig. 3.1a & b) and the occurrence of peat (Fig. 3.1d & e) were considered dynamic due to their greater propensity to anthropogenic influence over 70 years compared to the other soil-forming factors (Borrelli et al., 2017). Besides prediction uncertainty, we used statistical validation to assess mapping accuracy. Space-time mapping of soil properties remains a major challenge, with relatively few studies having mapped temporal changes on a regional to global scale (Stockmann et al., 2015; Gray & Bishop, 2016; Yigini & Panagos, 2016; Hengl et al., 2017a; Sanderman et al., 2017; Stumpf et al., 2018; Huang et al., 2019; Szatmári et al., 2019; Heuvelink et al., 2020; OpenGeoHub et al., 2022). To the best of our knowledge, our approach is the first to use a 3D+T dynamic covariate (Fig. 3.1d & e). Furthermore, these are the first SOM maps in 3D+T on a national scale. As a convention of this paper, SOM and absolute changes in SOM between two years (ΔSOM) are expressed as mass percentages.

3.2 Results and Discussion

3.2.1 SOM decrease in peatlands

Our findings indicate that between 1953 and 2022 there was a decrease of more than 1 % in SOM on 14 % of the land surface area of the Netherlands, which is equivalent to 4750 km^2. Furthermore, there was a decrease of over 10 % in SOM on 4.5 % of the land surface area,

Table 3.1: Number of SOM laboratory measurements and field estimates used for model tuning and calibration separated by depth layer.

Depth [cm]	SOM lab measurement	SOM field estimate
0 - 5	1049	18 873
5 - 15	5538	230 710
15 - 30	2500	117 800
30 - 60	5329	209 918
60 - 100	6170	138 122
100 - 200	2249	130 836

which amounts to 1520 km^2, predominantly occurring in peatlands. In former peat layers now classified as mineral soil layers, average SOM decreases at 0 - 30 cm depth ranged between 9 - 21 % (Table 3.2). For soils still classified as peat, average SOM decreases were mostly greater than 2 %. Regardless of peat oxidized or not, average SOM decreases in croplands were substantially higher compared to grasslands and forests. A recent study conducted in the Netherlands found that there were no substantial changes in SOM in organic soils at 0 - 30 cm depth (Knotters et al., 2022). However, at 30 - 100 cm depth, the study showed a decline of 7 - 9.5 % between 1998 and 2018, which is consistent with our research findings in organic and former organic soils across all depths. The prediction maps also align with the spatial patterns of SOM in other studies (van den Berg et al., 2017; Knotters et al., 2022).

Table 3.2: Predicted average changes in SOM [%] for 0 - 30 cm depth between 1953 (left) and 2022 (top) for combinations of soil type (peat vs. mineral) and land use (grass = grassland, crop = cropland, forest).

		2022					
		Peat, grass	Peat, crop	Peat, forest	Mineral, grass	Mineral, crop	Mineral, forest
1953	Peat, grass	−0.6	−2.1	−0.5	−10.0	−13.6	−10.9
	Peat, crop	−2.1	−3.5	−2.2	−15.1	−21.4	−17.2
	Peat, forest	−2.3	−2.8	−1.1	−9.4	−12.3	−10.1
	Mineral, grass	5.2	11.7	10.6	0.0	−0.3	−0.1
	Mineral, crop	16.2	23.3	18.3	0.3	−0.1	0.2
	Mineral, forest	6.5	22.3	9.1	0.1	−0.2	0.0

3D spatial predictions of SOM changes (Fig. 3.3) were different for bog and brook-valley peatlands, located mainly in the Northeast of the country, and fen peatlands, located mainly in the low-lying West and Northwest of the country (Fig. 3.1c). In bogs and brook-valleys, peat layers were often thinner than 1 m (Fig. 3.3a, b & e, Supplementary videos[3]) and SOM decreased by more than 10 % or even more than 25 % between 1953 and 2022 (Fig. 3.3c, d & f). Time-lapse maps spanning the entire 70-year period provide a

[3]Supplementary videos of Chapter 3 are available at `https://doi.org/10.1038/s43247-024-01293-y` under "Supplementary information".

visualization of these gradual changes over time at different depth layers (Supplementary videos). SOM decreases exceeding 25 % were primarily predicted at depths below 20 cm (Fig. 3f). This can be attributed to the fact that in peatlands, SOM predictions were mainly confined to a range of 5 - 20 % within the uppermost 20 cm, while deeper depths had predictions surpassing 25 % (Fig. 3.4a, b & e, Supplementary videos). Fig. 3.4c shows a typical brook-valley region on a 1:25 000 map (left), where SOM decreased by 10 - 25 % at 5 - 15 cm depth due to peat oxidation. The center of this map depicts a location under intensive grassland visited as part of the ΔSOM validation dataset (Sect. 3.3.1). Here, yearly SOM predictions between 1953 and 2022 (lines) and measurements in 1965 and 2022 (points) from 9 - 18 cm and 23-30 cm depth also underline this trend of decreasing SOM over time (Fig. 3.4c). Overall, our findings corroborate previous surveys in the northeastern province of Drenthe, which indicated that peat layers less than 40 cm thick are found in roughly 47 % of the area previously classified as having thick peat soils, while mineral soils cover approximately 55 % of the area that was initially identified as shallow peat soils (de Vries et al., 2009; van Kekem et al., 2005; Kempen et al., 2012a).

In fens, peat layers exceeding 1 m within the predicted 0 - 2 m depth range were found to be less susceptible to SOM decrease compared to thinner layers (Fig. 3.3). Upon comparing Fig. 3.3e & f, a noticeable trend emerges between 110 000 and 130 000 Easting: a decrease in SOM is predicted in the thinner peat layers that are located adjacent to the thicker ones. Typical for these fen meadow regions, Fig. 3.4d shows little to no SOM changes between 1953 and 2022 at 30 - 60 cm depth. Here, yearly SOM predictions between 1953 and 2022 and measurements in 1971 and 2022 at 23 - 50 cm depth were above 40 % and did not change. SOM measurements increased from 40 % to 49 % in the upper 6 cm, perhaps because the area was turned into a nature conservation area in the 1980s.

Although our predictions indicate limited SOM changes in peat layers exceeding 1 m, it is important to note that such areas may still experience a decline in carbon stocks or net CO_2 emissions. As our modelling was restricted to the top 2 m, SOM changes beyond this depth were not captured in our analysis. Therefore, it is crucial to exercise caution when interpreting our findings with respect to carbon loss or gain in soils with peat layers thicker than 2 m. In fact, previous studies reported carbon losses up to 1 g/kg/year (Reijneveld et al., 2009) and peat oxidation rates up to 1 cm/year in the fen meadow regions of the Netherlands (van den Akker et al., 2008; Hoogland et al., 2012).

3.2.2 SOM changes in reclaimed lands due to land subsidence

Model predictions reveal SOM increases of more than 10 % in large areas of reclaimed land (Fig. 3.1c) below 80 cm depth between 145 000 - 170 000 Easting (Fig. 3.3d & f). Time-lapse maps spanning 70 years visualize these gradual changes at 100 - 200 cm depth (Supplementary video 6). Land subsidence due to clay ripening, peat oxidation and soil compaction caused peat layers below 80 cm depth to shift upwards in terms of relative

depth (Zuur, 1958; van den Akker et al., 2007, 2008; Brouwer et al., 2018) (Fig. 3.5), leading to SOM increase. As a result of land subsidence, SOM also decreased by 10 - 25 % directly below thinner peat layers of approximately 50 cm (Fig. 3.3f between 155 000 - 170 000 Easting). Even at 0 - 30 cm depth, some soil layers changed from mineral to peat as a result of land subsidence, which explains predicted increases of 5 - 23 % SOM (Table 3.2). If peat layers closer to the surface rise above groundwater levels, oxidation and carbon emissions will increase in the next decades. Carbon emissions as a result of land subsidence have been confirmed not only in the Netherlands (Erkens et al., 2016; van Asselen et al., 2018), but also other coastal plains and deltas worldwide (Törnqvist et al., 2008; Syvitski et al., 2009).

3.2.3 Little SOM change in mineral soils

The model showed no substantial changes in SOM in the top 30 cm of mineral soils in grasslands or forests between 1953 and 2022 (Table 3.2). However, an average decrease of 0.1 % was predicted in croplands. When grasslands or forests were converted into croplands, we predicted an average decrease of 0.1 - 0.3 % in SOM, while the reverse scenario resulted in an increase of a similar amount. The model's predictions for 2022 indicated that SOM levels in mineral soils in the top 30 cm were between 2.5 - 5 % (Fig. 3.4a & e). Below 30 cm, temporal changes in SOM were almost non-existent in mineral soils (Fig. 3.3d & f, Supplementary videos 4 - 6). This outcome was expected because land use, which was considered a dynamic 2D+T covariate in the model (Fig. 3.1a & b, Table 3.6, Sect. 3.3.2), had little effect on the subsoil. Below 30 cm depth, the model's predictions ranged from 1 - 2.5 % for clay and loamy soils along the rivers, delta, and loess regions and less than 1 % in the Pleistocene sandy areas (Fig. 3.4b & e, Supplementary videos 4 - 6). These findings are reasonable because subsoils typically have lower SOM than topsoils.

Despite its overall effectiveness, the model occasionally did not detect subtle temporal changes in SOM, as demonstrated in Fig 3.4e (right). In this case, measurements indicated a 1 % decrease in SOM between 1968 and 2022 in a sandy soil currently used for maize cultivation. However, the model's predictions remained relatively constant. Analysis of a 1:25,000 map of the surrounding area revealed few predicted changes in SOM above 1 % at a depth of 5 - 15 cm, with the exception of a nearby brook valley located southeast of the measurement site (Fig 3.4e; left). Small changes in SOM as a consequence of land use changes can also be assumed based on the model's variable importance, where dynamic land use covariates were among the least important (Fig. S9).

While the general trends we found in mineral soil across different land uses are consistent with other findings for the Netherlands and Europe, the average rate of SOM changes were lower than in previous research. When converted to SOM changes (Sect. 3.3.7), Reijneveld et al. (2009) found increases of 0.40 % for grasslands (0 - 5 cm) and 0.32 % for

croplands (0 - 25 cm) in mineral soils in the Netherlands between 1984 and 2004. Chardon et al. (2009) confirmed constant or slight increases in Dutch agricultural soils with SOM less than 14 %. However, Conijn & Lesschen (2015) found an overall increase in SOM for permanent grassland but an overall decrease in SOM in croplands using a dynamic soil-crop model (Wolf et al., 2003; Groenendijk et al., 2013). More recently, Knotters et al. (2022) found a decrease in SOM of 0.38 % and 0.86 % in Dutch croplands in mineral soils between 1998 and 2018 at 0 - 30 cm and 30 - 100 cm depth, respectively. However, for Dutch grasslands, the same study found a decrease of 0.48 % SOM for 30 - 100 cm, while no substantial change was found for 0 - 30 cm depth. Between 2009 and 2015, converted SOM changes (Sect. 3.3.7) on a European scale were approximately -0.09 % on croplands, 0.24 % on grasslands and -0.51 % for grasslands converted to cropland (Fernández-Ugalde et al., 2020).

To summarize, our study and most recent research conducted in the Netherlands and Europe indicate that there is a decrease in SOM in croplands, particularly when converted from grassland, and an increase in SOM in grasslands. When accounting for the differences in time periods across the compared studies, the predicted changes in SOM we found were about 10-fold lower compared to actual measurement analyses at point or field-scale in other studies. Future studies should investigate this further, but one reason for this difference might be that random forest and other regression models smoothen predictions (Hengl et al., 2018), possibly leading to smaller predicted SOM changes. In addition, our model included both mineral and peat soils with SOM values up to 100 % (Table 3.5), potentially decreasing the model's sensitivity to detect changes of smaller magnitude in mineral soils. To improve the accuracy of 3D+T mapping specifically for mineral soils, we suggest mapping them separately from organic soils or choosing a hierarchical approach (Nussbaum et al., 2023), investing in repeated measurements at the same locations for model calibration, and deriving covariates related to agricultural management practices. Future studies should investigate whether performance of our 3D+T modelling approach improves when monitoring data (more repeated measurements) are used during model calibration.

3.2.4 Model accuracy assessment

Model accuracy was assessed using a 10-fold cross-validation with data from 1953 - 2011 (MEC = 0.64; Fig. 3.4a) and design-based inference from 1993 - 2000 (MEC = 0.50) and 2018 (Table 3.3; Sect. 3.3.8). The relatively high proportion of peat samples with SOM values up to 100 % (Table 3.5) may explain relatively high RMSE values around 10 %. The predicted SOM content at 0 - 30 cm depth was underestimated (ME $>$ 1 for all methods; Table 3.3), which probably also relates to smaller predicted SOM changes in mineral topsoils compared to other studies (see above). Inaccuracies in the 30 - 100 cm layer in 2018 may be due to positional errors, differences in sampling support, or changes in laboratory methods between the calibration and validation data (Knotters et al., 2022).

Furthermore, the use of the same data for model calibration and the national soil map generation, from which dynamic peat covariates were derived (Fig. 3.5), may have biased our predictions. The overall spatial patterns of predicted SOM align with previous SOM mapping studies in the Netherlands (van den Berg et al., 2017; Knotters et al., 2022).

Table 3.3: Model accuracy metrics of SOM [%] predictions using 10-fold cross-validation with laboratory measurements from 1953 - 2011 and design-based inference of an independent probability sample using measurements from 1993 - 2000 and 2018, respectively (Sect. 3.3.8). The lower and upper 97.5 % confidence limits of the accuracy metrics were computed using design-based inference according to de Gruijter et al. (2006), which together give the 95 % confidence intervals (CI95; Sect. 3.3.8).

Years	Depth (cm)	n	**ME**	CI95 of ME	**RMSE**	CI95 of RMSE	**MEC**	CI95 of MEC	**PICP**
1953-2011	0 - 30	6264	**1.97**	-	**9.04**	-	**0.65**	-	**0.88**
1953-2011	30 - 100	7526	**0.38**	-	**10.02**	-	**0.65**	-	**0.87**
1953-2011	100 - 200	1509	**0.00**	-	**10.33**	-	**0.29**	-	**0.88**
1993-2000	0 - 30	1185	**1.29**	1.05, 1.54	**4.87**	4.23, 5.44	**0.49**	0.36, 0.59	**0.76**
1993-2000	30 - 100	1172	**0.20**	-0.34, 0.74	**9.79**	8.36, 11.03	**0.50**	0.18, 0.65	**0.91**
1993-2000	100 - 200	808	**0.82**	0.04, 1.60	**9.63**	6.82, 11.79	**0.52**	0.39, 0.67	**0.96**
2018	0 - 30	1143	**1.15**	0.89, 1.42	**5.58**	4.85, 6.22	**0.44**	0.27, 0.56	**0.85**
2018	30 - 100	1139	**-4.28**	-4.89, -3.67	**15.28**	14.05, 16.43	**-0.93**	-1.39, -0.53	**0.96**

Our estimates of prediction uncertainty (PI90) in 3D space (SI2) and time (Fig. 3.4c-e) were reliable based on the evaluation of prediction interval coverage probability (PICP; SI2; Sect. 3.3.8). However, when divided by depth, design-based inference from 1993 - 2000 and 2018 revealed that prediction uncertainty was overly-optimistic at 0 - 30 cm depth and slightly pessimistic below 30 cm depth (Table 3.3). Areas with high SOM values (peatlands) and urban areas had the highest prediction uncertainty, which was overall greater at lower depths (SI2). Fig. 3.4c confirms that uncertainty decreased as predictions decreased over time. Mineral soils had lower uncertainty (SI2 & Fig. 3.4e).

An important limitation of our modelling approach is that it does not quantify the uncertainty of differences in SOM (ΔSOM) and of spatial aggregates (e.g. Table 3.2). These uncertainties can only be obtained if cross- and spatial correlation in prediction errors are quantified. For instance, the prediction error variance of ΔSOM at some location and depth is given by the sum of the variances of the SOM prediction errors at the two points in time for that location and depth, minus twice their covariance. Computation of the covariance requires the correlation between the two prediction errors. In case of spatial aggregation one must first quantify the spatial correlation of the SOM prediction errors, for instance by semivariograms, after which the uncertainty of the spatial aggregate can be computed using a block kriging of the residuals. Wadoux and Heuvelink (Wadoux & Heuvelink, 2023) did this in 2D space, but it is unclear how this should be efficiently done in 3D space and time (four dimensions). Semivariogram fitting in 3D+T is extremely challenging given that space-time and lateral-vertical anisotropies would have to be accounted for, while also the conventional geostatistical assumptions on multivariate normality and

second-order stationarity would have to be questioned. There are also considerable computational challenges when conducting block kriging in a 3D+T context at high spatial resolution. We know of only one study that estimated 3D+T semivariograms and applied 3D+T kriging (Gasch et al., 2015), but this study assumed normality and used a simplified metric space-time semivariogram, which might not be appropriate for SOM in the Netherlands. Uncertainty quantification of SOM changes and spatial aggregates in 3D+T at scales relevant to management and policy is a critical task to investigate in future research, since the uncertainty related to soil monitoring has prompted widespread doubts about the feasibility of measuring and verifying SOM and soil organic carbon changes (Moinet et al., 2023; Paul et al., 2023). Thus, we advocate that future research should investigate the uncertainty quantification of SOM changes and SOM spatial aggregates, but given the challenges and complexity of such analysis this was beyond the scope of this research.

Temporal SOM changes (ΔSOM) were difficult to predict (Fig. 3.4b). The model sometimes failed to detect ΔSOM (Fig. 3.4b-e). However, prediction errors at point support tend to average out when increasing spatial and temporal support (Webster & Oliver, 2007; Szatmári et al., 2021; Wadoux & Heuvelink, 2023), e.g. by averaging over larger areas, depth layers and years, making the 3D+T maps useful for many applications. For instance, predictions of samples c1 and c2 (lines in Fig. 3.4c) showed an unrealistic decrease of more than 10 % SOM between 1972 and 1973, but when averaged over several years, the trend of decreasing SOM was confirmed by measurements in 1965 and 2022 (points in Fig. 3.4c). In general, temporal SOM variation was often lower than 3D spatial variation. However, the limited ΔSOM validation data (127 measurements from 63 locations; Fig. 3.3c) highlight the need for long-term soil monitoring using consistent methodologies in the field and laboratory (Smith et al., 2020). For additional validation of ΔSOM predictions beyond the dataset used here, future studies could compare our predicted SOM changes with measured changes at individual locations of long-term field experiments (Kooistra & Kuikman, 2002) for different soils and land uses. This would allow evaluation of the model's ability to predict temporal changes, while currently the accuracy assessment was based on changes both in space and time.

3.2.5 3D+T mapping: a new paradigm for SOM monitoring

Our study demonstrated that SOM is highly variable over depth and time, which is insufficiently captured by non-spatially explicit (Schrumpf et al., 2011; Fernández-Ugalde et al., 2020; Knotters et al., 2022) or 2D mapping methods (Heuvelink et al., 2020; Poggio et al., 2021) that are currently used for reporting SOM changes for soil health and climate mitigation commitments. In the Netherlands, approaches to assess SOM changes encompass long-term field experiments conducted at specific locations, summarized by Kooistra and Kuikman (Kooistra & Kuikman, 2002), some of which served as the basis for process-based models (Wolf et al., 2003; Chardon et al., 2009; Groenendijk et al., 2013; Conijn

& Lesschen, 2015). Reijneveld et al. (2009) assessed changes of agricultural parcels using farm data. Knotters et al. (2022) assessed changes on the scale of four selected domains of interest, categorized by mineral soil, organic soils, grasslands and croplands. Unlike our research, none of the above studies explicitly account for spatial variation when modelling SOM dynamics. Furthermore, process-based models are often constrained by soil type (e.g. mineral) or land use (Coleman & Jenkinson, 1996). In contrast, our 3D+T model accounts for SOM variation in space, depth and time and can predict for mineral and organic soils under any land use. In addition, the 3D+T model does not require repeated measurements from the same location, although prediction accuracy is likely to increase with better monitoring data. This is a major advantage because most areas in the world do not have monitoring data beyond individual field trials.

Previous digital soil mapping studies used spline depth functions (Adhikari et al., 2014a), geostatistical methods (Poggio & Gimona, 2014; Orton et al., 2016, 2020), parametric depth functions (Kempen et al., 2011) or depth as a covariate (Chapter 2; Poggio et al., 2021) to map a soil property at different depths. Gasch et al. (2015) used 3D covariates (soil parameters) and a 2D+T covariate (crop type) to predict soil water, temperature and electrical conductivity in 3D+T on a field scale (37 ha). However, to our knowledge this is the first study to use a 3D+T covariate. Machine learning has proven advantageous to predict complex, non-linear relationships between soil-forming factors and soil properties in 2D space (Wadoux et al., 2020). Our study represents a next step in extending the predictive power of machine learning to 3D+T. In doing so, the 3D+T model was able to detect complex relationships between SOM and peat occurrence, which varied considerably in space, depth and time. The 3D+T covariate was the most important covariate in the model (Fig. S9). As a result, we found major changes in SOM not only in space but also over depth and time, especially in peatlands and reclaimed land. Therefore, we recommend that 3D+T modelling approaches be incorporated in SOM reporting alongside point monitoring networks to provide spatially explicit information. This is key for facilitating multi-functional land use policies and management practices based on local soil conditions. The 3D+T SOM maps with quantified uncertainty can aid decision-making. It can support decision-making on where to implement measures to increase SOM stocks to address the 4 per mille initiative and incorporate carbon farming in agricultural practices. They are also an important first step towards providing spatially explicit changes in soil carbon stocks and CO_2 emissions from soils. As stated in the Climate Agreement of the Netherlands, the emissions of peat soils must be reduced with 1 Mton CO_2-eq and of mineral agricultural soils by 0.4 - 0.6 Mton CO_2-eq before 2030 (Government of the Netherlands, 2019). Moreover, these maps identify where SOM-related soil health has declined most in the past 70 years and where restoration potential is highest and most urgent. Finally, the 3D+T SOM maps can serve as visual tools to raise awareness of the importance of soils for society, which is the Soil Deal's eighth mission objective.

This research goes beyond the mere mapping of SOM between 1953 and 2022, as it has

far-reaching implications for the future. With advancements in the explainable machine learning research domain, models using algorithms such as quantile regression forest are no longer considered black-box models. Recently, these methods have been applied in soil science to gain new insights into the complex relationship between covariates and soil properties (Wadoux & Molnar, 2022; Wadoux et al., 2023). Future studies could identify potential local drivers of SOM dynamics by using explainable machine learning methods (Wadoux & Molnar, 2022; Wadoux et al., 2023) to study the relationship between covariates and changes in SOM over time using the 3D+T methodology. Furthermore, this will allow the prediction of potential future changes in SOM. For example, by considering various scenarios involving groundwater levels, land use changes, or climatic indicators (van Dorland et al., 2023), we can forecast changes in SOM (Yigini & Panagos, 2016), providing crucial insights into the measures necessary to restore soil health.

The implications of the findings reported in this study also extend beyond the Netherlands. Peatland conversion (Murdiyarso et al., 2010; Dohong et al., 2017), land reclamation (Martín-Antón et al., 2016), and agricultural intensification (Kopittke et al., 2019) are ongoing in many parts of the world, but there is a lack of spatio-temporal soil and land use data in many of these regions. Moreover, the mechanisms underlying SOM dynamics apply to other bioclimatic zones as well, suggesting that the changes observed in the Netherlands may be relevant to less data-rich regions across the globe. These findings are particularly relevant to deltas worldwide, which often share similar geographic features and are home to approximately 350 million people (Ericson et al., 2006; Edmonds et al., 2020).

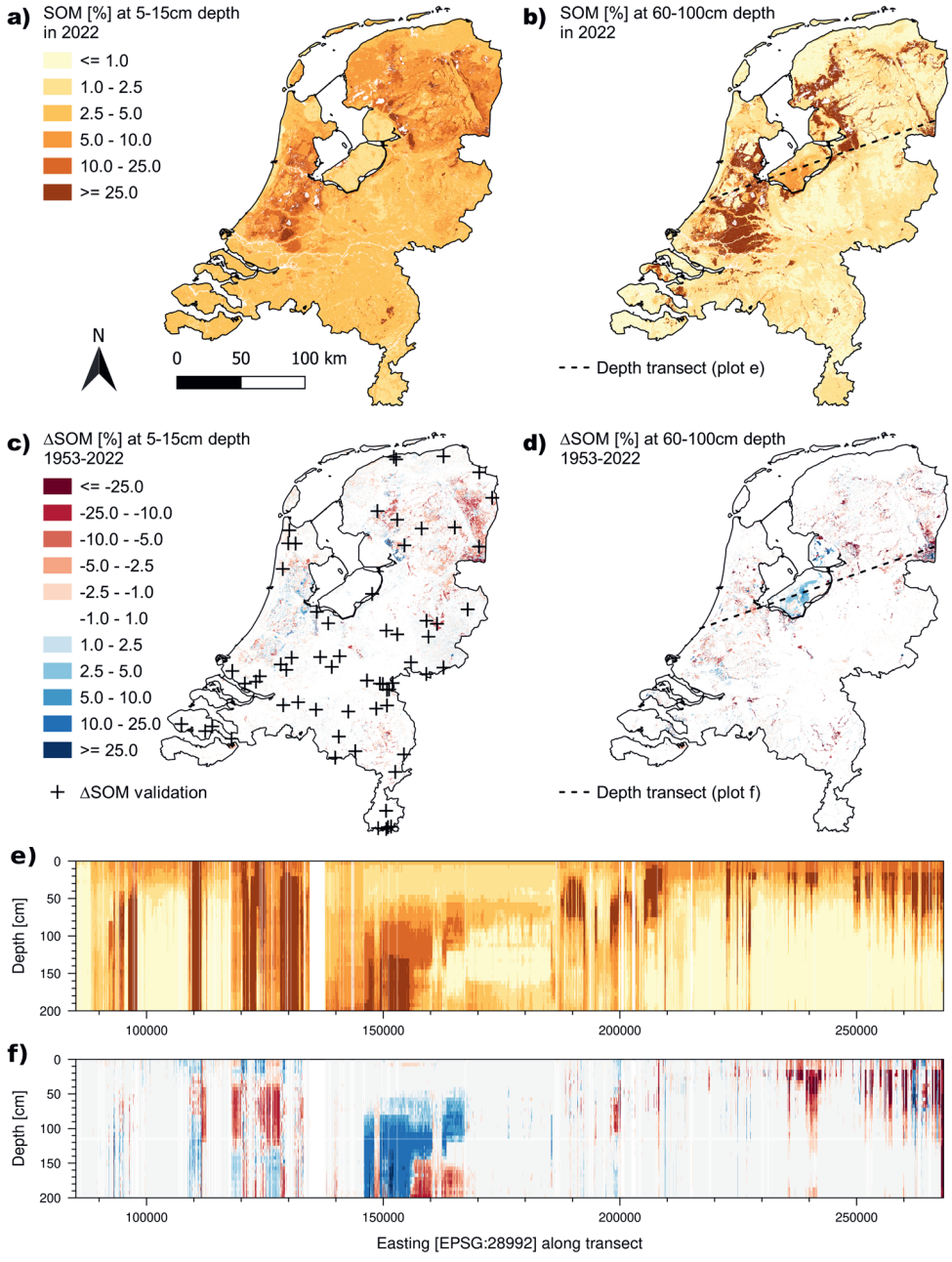


Figure 3.3: a) SOM [%] predictions for 2022 at 5 - 15 cm depth (a) and 60 - 100 cm depth (b); ΔSOM [%] (1953 - 2022) at 5 - 15 cm (c) and 60 - 100 cm depth (d); SOM [%] predictions for 2022 (e) and ΔSOM [%] predictions for 1953 - 2022 (f). Predictions shown in e and f were made from 0 to 200 cm depth with 5 cm depth increments along the transect shown in b and d. The crosses in c show the ΔSOM validation locations (Sect. 3.3.1).

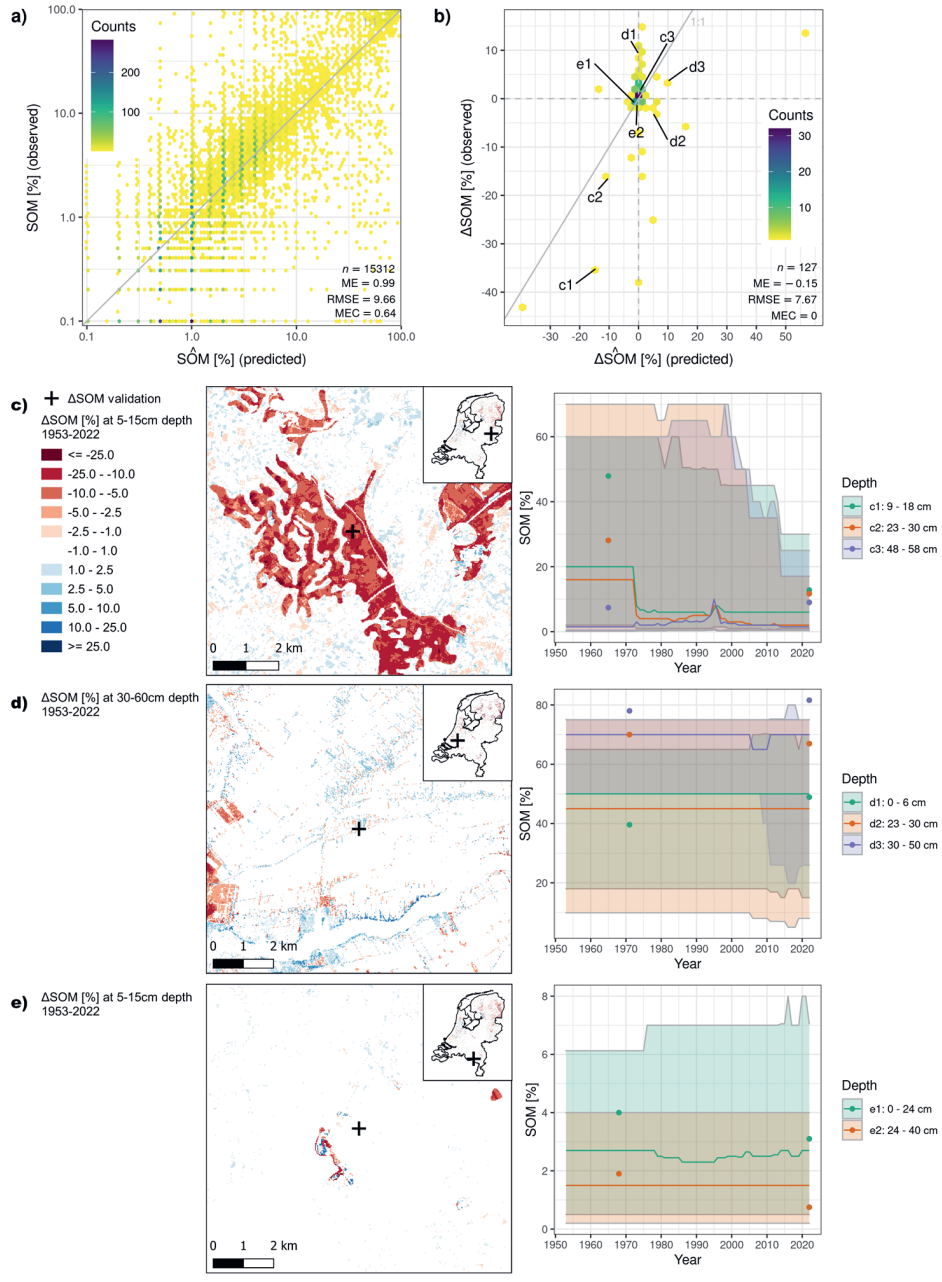


Figure 3.4: Accuracy plot and accuracy metrics using 10-fold cross-validation of SOM [%] laboratory measurements between 1953 - 2011 (a) and validation of SOM [%] temporal changes (ΔSOM) at 63 locations shown in Fig 3.3c (b, Sect. 3.3.8). Zoom-in maps [1:25 000] of three ΔSOM validation locations (c-e; left) and time series plots from these locations sampled at two or three depths (c-e; right). Time series plots depict SOM [%] laboratory measurements (points), model predictions [%] (line) and prediction uncertainty [%] (PI90, background) between 1953-2022. Time series of samples in c-e are also indicated in b.

3.3 Materials and Methods

3.3.1 Soil point data

We obtained 869 094 observations of SOM from 339 231 point locations using different datasets, most of which are part of the Dutch soil database (BIS; Fig. 3.2 & Table 3.4). SOM observations consist of either measurements in the laboratory using loss on ignition at 550 °C (NEN 5753, 2020) or field estimates. Regarding the latter, soil surveyors estimated SOM in the field by looking and touching the soil sample for its color and texture. An expert estimation is then made, also based on their extensive, regional pedological knowledge of the soil forming factors, soil texture, expected soil type and using SOM laboratory measurements at nearby locations. In instances where multiple soil surveyors made estimates at the same location and depth, the individual estimates were merged and the median value recorded (ten Cate et al., 1995; de Bakker & Schelling, 1966, 1989). Based on approximately eight thousand paired laboratory measurements and field estimates, and assuming that the laboratory measurement error is negligible compared to the field estimation error, the mean error, mean absolute error and standard deviation of the field estimation error were 0.23 %, 2.09 % and 4.6 % SOM, respectively. Hence, these errors were within an acceptable range for our purpose, considering that laboratory measurements themselves are also subject to errors (van Leeuwen et al., 2021). Field estimates were discarded from all modelling steps whenever laboratory measurements from the same 3D location were available. Only the PFB and BPK datasets of BIS were used for model calibration because the probability sampling design of the LSK and CCNL datasets was ideal to independently assess mapping accuracy (Chapter 2; Finke et al., 2001; Visschers et al., 2007) (Table 3.4).

Soil point data for model calibration

For model calibration, we used 15 312 laboratory measurements from 4298 locations (PFB) and 840 638 field estimates from 334 668 locations (PFB and BPK; Fig. 3.2a; Table 3.4). The locations with laboratory measurements, arranged in a purposive sampling design, were selected in the past to create the national soil map (1:50 000) (de Vries et al., 2003), meaning that soil variability is well covered. Soil samples measured in the laboratory were collected by genetic soil horizon between 1953 and 2011 (PFB; Fig. 3.2b).

In contrast to the laboratory measurements, the majority of field estimates (BPK) were spatially clustered in specific areas for regional soil mapping purposes. SOM field estimates were removed if there was a laboratory measurement available from the same location and genetic soil horizon. We decided to include field estimates in model calibration due to the additional spatio-temporal coverage (1953-2022), especially in recent years (Fig. 3.2b). However, since field estimates are less accurate than laboratory measurements and due to their clustered spatial distribution, we tested excluding field estimates

Table 3.4: Overview of datasets used for modelling. With the exception of ΔSOM, dataset names are those used in BIS, but in this study we refer to them by their temporal coverage. Pure panel is synonymous to static-synchronous (Brus, 2022). "Horizon" indicates that soil samples were observed by genetic soil horizon, whereas "fixed" indicates that observations were made at fixed depth intervals. See Table 3.1 for the number of laboratory measurements in PFB and field estimates in BPK and Table 3.3 (n) for the number of laboratory measurements for LSK (1993 - 2000) and CCNL (2018) across fixed depth intervals. Obs. = observations; Field = field estimates; CV = cross-validation.

Dataset	Sampling design	Space-time design	Statistical validation	Locations	Obs.	Depth [cm]	Temporal coverage	References
PFB	Purposive	-	10-fold CV	4298	15 312	0 - 200 (horizon)	1953-2011	de Vries et al. (2003); van den Berg et al. (2017); Chapter 2
BPK (field)	Purposive	-	-	334 668	840 638	0 - 200 (horizon)	1953-2022	van den Berg et al. (2017)
LSK	Probability	Supplemented panel (Fuller, 1999; Brus, 2022)	Design-based inference	1185	4952	0 - 200 (horizon)	1993-2000	Finke et al. (2001); Visschers et al. (2007); Knotters et al. (2022); Chapter 2
CCNL	Probability	Supplemented panel (Fuller, 1999; Brus, 2022)	Design-based inference	1144	2295	0 - 30, 30 - 100 (fixed)	2018	van Tol-Leenders et al. (2019); van den Elsen et al. (2020); Teuling et al. (2021); Knotters et al. (2022)
ΔSOM	Purposive	Pure panel (Fuller, 1999; Brus, 2022)	Non-design based	63	127	0 - 200 (horizon)	1953-1999; 2022	SI3

and assigning them lower weights during model tuning (see below).

Soil point data for model accuracy assessment

Four different datasets were used for statistical validation to assess model accuracy (Table 3.4). The first dataset (PFB) consisted of the same 15 312 laboratory measurements used during model calibration. For model tuning and accuracy assessment purposes, this dataset was used for cross-validation (see below).

We further had the LSK and CCNL datasets available specifically collected for validation purposes of the national soil map (1:50 000) (de Vries et al., 2003). The LSK consisted of 4952 SOM laboratory measurements from 1185 locations sampled by horizon between 1993 and 2000. These soil sampling locations were determined using a national probability sample, more specifically a stratified simple random sample. The dataset is described in more detail in Finke et al. (2001) and Visschers et al. (2007) and its use for validating digital soil maps in Sect. 2.2.1.

All LSK locations that were still accessible were re-sampled at two fixed depth increments (0 - 30 cm and 30 - 100 cm) in 2018 (van Tol-Leenders et al., 2019; van den Elsen et al., 2020;

Table 3.5: Descriptive statistics of SOM observation datasets shown in Table 3.4. Min. = minimum; Qu. = quartile; Max. = maximum.

Dataset	Temporal coverage	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Skewness
PFB	1953-2011	0.00	0.60	2.00	8.39	5.10	99.90	3.29
BPK	1953-2022	0.00	1.50	4.00	15.47	10.00	99.00	1.82
LSK	1993-2000	0.00	0.90	2.30	7.70	5.10	95.00	3.47
CCNL	2018	0.50	1.80	3.40	7.51	6.80	78.70	3.09
ΔSOM	1953-1999; 2022	0.00	0.60	1.80	10.07	5.40	96.90	2.84

Teuling et al., 2021; Knotters et al., 2022). This so-called CCNL dataset consists of 2295 laboratory measurements from 1144 locations. In terms of space-time design, the LSK and CCNL datasets are therefore a supplemented panel because only a subset of the sampling locations of the first survey were re-visited approximately 20 years later (Brus, 2022). Despite the supplemented panel design, substantial methodological differences in the LSK and CCNL datasets prevents a temporal assessment at point scale (Cavero Panez, 2021), as described in more detail in Sect. 2.1 and Appendix C of Knotters et al. (2022), who used these data to study temporal SOM changes within domains (not at point scale).

We will refer to the three datasets described above by their measurement years, i.e. 1953 - 2011, 1993 - 2000 and 2018.

In order to also assess changes in SOM over time (ΔSOM) at point locations, we re-sampled the same 1 - 3 uppermost genetic soil horizons from 63 PFB locations in 2022, leading to a total of 127 samples (Fig. 3.3c; Table 3.4). These PFB locations were first sampled between 1953-1999 (depending on the location) and because all locations were re-sampled in 2022, it can be termed a pure panel space-time design (Brus, 2022). In contrast to LSK and CCNL, we sampled identical legacy soil horizons as in the past to the best of our abilities. The purposive sampling design of ΔSOM locations is described in detail in SI3. The 127 samples used for statistical validation of ΔSOM were removed from the PFB dataset to avoid their use during model calibration and 10-fold cross-validation (see above).

3.3.2 Covariates

In line with the digital soil mapping methodology (McBratney et al., 2003), we used covariates as model independent variables (i.e. explanatory variables or features) that were representative of the soil-forming factors: climate, organisms, relief (topography), parent material (geology) and time (Dokuchaev, 1899; Jenny, 1941). In order to map SOM in 3D space and time, we extended upon established methods by using covariates that were static (2D), variable in time (2D+T) (Heuvelink et al., 2020; OpenGeoHub et al., 2022) and variable over depth and time (3D+T). All covariates were prepared at 25 m resolution, for the standard depth layers specified by GlobalSoilMap (Arrouays et al.,

2015) (GSM; 0 - 5 cm, 5 - 15 cm, 15 - 30 cm, 30 - 60cm, 60 - 100 cm and 100 - 200 cm) in the case of the 3D+T covariate, and for every year from 1953 to 2022 for 2D+T and 3D+T covariates.

Static covariates

Covariates were designated as static based on limited temporal variation or unavailable data over the 70-year period. Specifically, climate, relief, and parent material exhibited little temporal variability, while satellite-derived land cover indices were incomplete for the entire 70-year period. Although climate change may have impacted SOM, its effects were considered smaller than that of the dynamic covariates we selected (see below). Table 2.2 and Supplement S2 of Chapter 2 provide an overview of the static covariates. Additionally, we obtained monthly mosaics of Sentinel 2 RGB and NIR bands from 2015 onwards, removing monthly mosaics with more than 1 % clouds. We computed eight indices from these mosaics: Brightness Index (BI), Saturation Index (SI), Hue Index (HI), Coloration Index (CI), Redness Index (RI), Carbonate Index (CaI), Grain Size Index (GSI), and Normalized Difference Vegetation Index (NDVI), following Loiseau et al. (2019). To improve the signal-to-noise ratio and reduce data volume, we further processed the spectral indices into the first three principal components over all months and years (long-term yearly aggregates), as well as the long-term monthly mean and standard deviation. In total, we considered 318 static covariates for further model selection.

Dynamic 2D+T and 3D+T covariates

In recent decades and perhaps even centuries, anthropogenic activity has arguably altered soil characteristics more than any natural soil-forming factor (Amundson et al., 2015). To account for SOM changes between 1953 and 2022 in the Netherlands, we chose land use and the occurrence of peat as dynamic covariates due to their important link to SOM and high temporal variability.

Acquiring harmonized and spatially exhaustive information on land use changes from the 1950s to the present is difficult, as high-resolution remote sensing products are not available prior to the 1980s. However, in the Netherlands, the main land use categories were carefully mapped using topographic maps since the 1900s. Largely based on these maps, digital historical land use maps were created for around 1900, 1960, 1970, 1980, and 1990 (Kramer et al., 2010) (Table 3.6). Since the 1980s, national land use maps have been developed by combining information from the Dutch key registries for topography, land cover, agricultural parcels, urban areas, and nature, as well as remote sensing data, to provide greater detail (Hazeu, 2014). These maps have been regularly updated using the latest data and improved methods and since 2018 are updated annually. In total, we used five historical land use maps (1900, 1960, 1970, 1980, and 1990) and eleven recent land use maps from 1984 to 2021 to derive dynamic land use covariates.

Table 3.6: Table of the dynamic covariates variable in time (2D+T) related to land use (LU). Names denote the names used in the model code. Note that categories in more recent land use maps (1984 - 2021) (Hazeu, 2014) were aggregated to the nine classes of the historical land use maps (Kramer et al., 2010) shown here.

Name	Categories	Derived from
LU_xyt	Grassland Cropland	Historical land use maps
LU_xyt_delta5	Fruit orchards & tree nurseries Forest	(Kramer et al., 2010): 1900, 1960, 1970, 1980, 1990
LU_xyt_delta10	Heathland Sand & dunes	Land use maps (Hazeu, 2014):
LU_xyt_delta20	Swamps, marshes & fens Bogs	1984-2021 (since 2018 annual)
LU_xyt_delta40	Built-up	

All land use maps were reclassified into nine general classes (Fig. 3.1a & b; Table 3.6), which were limited by the classes contained in the historical land use maps. We obtained the land use for every location, with coordinates x and y for every year t between 1953 and 2022 (LU_{xyt}), by assigning the same class as in the temporally nearest year for which a map was available. If t was exactly in between two years for which a map was available, the older map was used. In the same manner, we further defined $LU_{xyt}\_\Delta 5$, $LU_{xyt}\_\Delta 10$, $LU_{xyt}\_\Delta 20$ and $LU_{xyt}\_\Delta 40$ by assigning the land use class that occurred most frequently in the 5, 10, 20 and 40 years prior to and including year t, respectively. These modal classes were assigned to account for the delayed response of SOM to land use change. An example of a 2D+T dynamic land use covariate (LU_{xyt}) is shown in Fig. 3.1a & b for the years 1953 and 2022, respectively. Previous studies have employed comparable approaches to derive dynamic covariates that serve as proxies for land cover and land cover changes, however, they utilized continuous covariates sourced from Landsat, MODIS or AVHRR products from more recent years (Heuvelink et al., 2020; OpenGeoHub et al., 2022).

Furthermore, we derived dynamic covariates for peat occurrence, considering that the majority of SOM in the Netherlands is found in soils with a peat layer. Peat soils have undergone substantial changes over time due to historical excavation, drainage, compaction, and agricultural management (de Vries et al., 2009; van Kekem et al., 2005; Kempen et al., 2012a; van den Akker et al., 2008; Hoogland et al., 2012; van Asselen et al., 2018; Knotters et al., 2022). In order to account for temporal changes in peat soil horizons, we used the original and updated version of the national soil map of the Netherlands (1:50 000) (de Vries et al., 2003) (Fig. 3.5), which is available at BROloket and bodemdata.nl. We included 2D+T dynamic covariates of different peat classes and a 3D+T covariate of peat occurrence because the combination of both helped explain SOM variability.

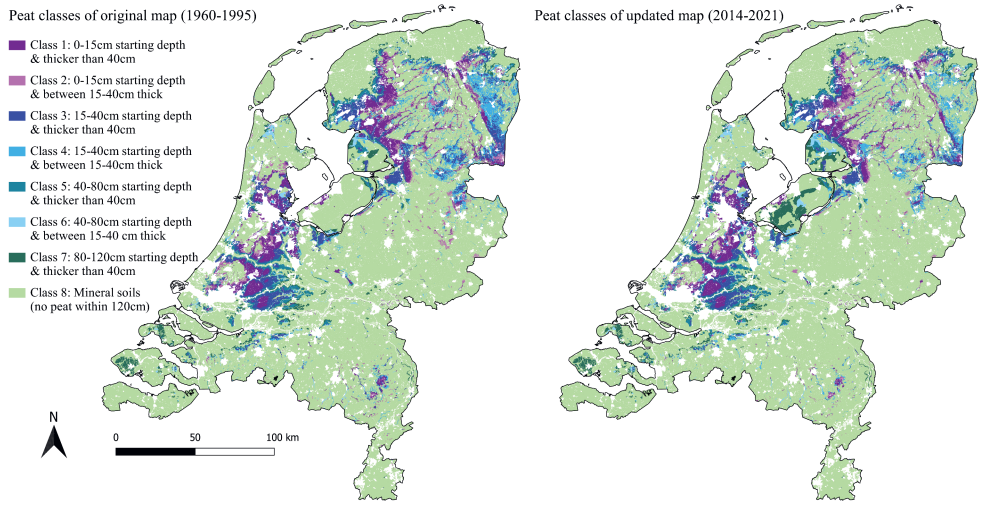


Figure 3.5: Maps of sub-classes of soils containing peat in the original (left) and updated (right) national soil map of the Netherlands (1:50 000) (de Vries et al., 2003). Following the Dutch soil classification (ten Cate et al., 1995; de Bakker & Schelling, 1966, 1989), soils containing peat can be classified according to the starting depth and thickness of the peat horizon. These maps were used to derive eight fuzzy subsets of peat classes (one for each class) as 2D+T dynamic covariates and one fuzzy subset of peat occurrence as a 3D+T dynamic covariate. Note that the national soil maps exclude urban areas (white). The national soil map of the Netherlands can be downloaded and viewed at `broloket.nl` and `bodemdata.nl`.

For deriving 2D+T dynamic covariates of the occurrence of different peat classes, we used the concept of fuzzy membership (Zadeh, 1965), which has also been widely used in soil classification and mapping (McBratney & Odeh, 1997). Let $p(x, y, t)$ denote the peat class at a location with coordinates x and y at time t, where t is any year between 1953 and 2022. In the Dutch soil classification system, used in the national soil map, soils containing peat can be designated into eight sub-classes, based on the starting depth and thickness of peat horizons (Fig. 3.5). Thus, each $p(x, y, t) \in \{1, 2, 3, 4, 5, 6, 7, 8\}$, where the first seven classes refer to different peat soils and where class 8 contains no peat within 0 - 120 cm (i.e., mineral soil). The information about which peat class occurs at a particular location was limited to a maximum of two points in time, t_o for the original mapping year, and t_u for the year the map was updated. t_o and t_u varied in space and t_u was only available for part of the country, since the soils of the Netherlands were systematically mapped, region by region, between the 1960s and 1990s, and some regions were updated once between 2014 - 2021. Many of the updated regions were areas with less than 2 m of peat within the top 2 m and reclaimed land, where soil characteristics or the depths at which these characteristics occur have substantially changed since the original map was made due to peat oxidation, compaction and land subsidence. For regions that were not updated, the

old and updated maps were the same. For $t < t_o$, we assumed $p(x, y, t) = p(x, y, t_o)$, i.e. for years between 1953 and the first mapping campaign we designated the same peat class as in the original map. For $t > t_u$, we assumed $p(x, y, t) = p(x, y, t_u)$, i.e. for years between which a region was updated and 2022, we designated the same peat class as in the updated map. However, for $t_o \leq t \leq t_u$, peat class was designated a fuzzy membership value that was a combination of the classes at t_o and at t_u. We let the membership of the class at t_o linearly decrease from one to zero in the period from t_o to t_u, and similarly we let it linearly increase from zero to one for the observed class at t_u, so that the sum of the memberships was always one. We used the 2D+T membership values for each of the eight peat classes as covariates in the random forest model.

For deriving the 3D+T dynamic covariate of peat occurrence, we used the peat starting depth and thickness information contained in the peat classes to derive another fuzzy variable of the occurrence of peat depending on location, depth and time. The fuzzy membership that was derived was as before a number between 0 and 1, where 0 means that peat does not occur, and 1 means that peat occurs. Any membership value in between 0 and 1 means that the soil at that location, depth and time was in a transition period from peat to non-peat or vice versa, in accordance with the soil peat class membership value at that location and time.

Note that some assumptions were needed to derive 3D+T peat occurrence depending on the peat class, depth and expert knowledge. For example, if $p(x, y, t) = 4$, which is the peat class for peat starting between 15 and 40 cm and having a thickness between 15 - 40 cm (Fig. 3.5), we assumed that peat occurs between $15 \leq d \leq 80$ cm because the exact depth range of peat between this minimum and maximum depth were not known more precisely. Likewise, for $p(x, y, t) \in \{1, 3, 5, 7\}$, where peat thickness was greater than 40 cm, we assumed that peat occurs for the entire depth range between the minimum starting depth for each class, that is 0, 15, 40 and 80 cm, respectively, and the maximum depth modelled (200 cm; Fig. 3.5). We made this assumption based on explanations from soil surveyors, who informed us that areas mapped with peat layers thicker than 40 cm tend to have substantially greater thicknesses. In order to predict at the standard depth layer specified by GSM (Arrouays et al., 2015), the upper and lower depth boundaries of 0 - 5 cm, 5 - 15 cm, 15 - 30 cm, 30 - 60cm, 60 - 100 cm and 100 - 200 cm layers were used for each year to derive peat occurrence according to t and d. During this step, assumptions were made regarding the overlap between GSM depth layers and defined depths of peat classes. When peat occurrence was greater than 0, the GSM depth layer needed to overlap by ≥ 1 cm with the peat layer, whereas when peat occurrence was equal to 0, we permitted no overlap with the peat layer. These assumptions were made because even if there is only a small overlap of peat, it will lead to substantially higher SOM values than on purely mineral soils.

Depending on the peat class in the original and updated national soil map, d, t and the

location, this ultimately resulted in one of three possible outcomes: no changes in peat occurrence, peat "appearing" or peat "disappearing". An example of the 3D+T dynamic peat occurrence covariate is shown in Fig. 3.1d & e for the year 1987 for depths 5 - 15 cm (d) and 60 - 100 cm (e).

3.3.3 Overlay and regression matrix

We created a regression matrix containing SOM and covariate values by performing a spatial overlay for static covariates, a space-time overlay for 2D+T covariates and a space-depth-time overlay for 3D+T covariates. t was equal to the year at which a SOM observation was made. For deriving peat occurrence according to t and d, the upper and lower sampled horizon boundaries were used, whereby the same assumptions were made regarding the overlap of depth layers as when deriving peat occurrence for the GSM depth layers (see above).

Sampling depth information, more specifically the upper and lower boundary and mid-point of each sampled horizon, were included as covariates in the regression matrix so that predictions could easily be made at any chosen depth and depth interval. Including these as covariates also supported accounting for changes in SOM over depth, in addition to the other 3D covariate. See Ma et al. (2021) for an overview of models using depth as a covariate in comparison to non-3D digital soil mapping methods.

3.3.4 Model selection, tuning and calibration

For model selection, defined here as selecting the best model based on their performance (Hastie et al., 2009), we first reduced the number of static covariates (n = 318). We removed covariates in a two-step procedure using de-correlation followed by recursive feature elimination as in Poggio et al. (2021). From any pair of covariates for which the Pearson correlation coefficient was greater than 0.85 or less than -0.85, the covariate that was more correlated with all remaining covariates was removed. Recursive feature elimination (Guyon et al., 2002) was implemented using the `caret` package (Kuhn, 2019). This resulted in a set of 16 static covariates. These, in addition to the three depth covariates, the five 2D+T covariates of land use, eight 2D+T covariates of peat classes and the 3D+T covariate of peat occurrence, were selected for model tuning and calibration and can be found in the variable importance plot (SI2).

For model tuning, we grew random forest models (Breiman, 2001) with the goal of optimizing hyper-parameters for mean predictions. Model tuning was performed using a location-grouped 10-fold cross-validation of the calibration data, wherein all observations from the same location were forced to be in the same fold. Each hold-in fold contained a random selection of 90 % of the laboratory measurements and all field observations. Each hold-out fold contained the remaining 10 % of laboratory measurements. In this way, models were calibrated using both laboratory measurements and field estimates, while

performance to optimize hyperparameters was assessed only using laboratory measurements. We evaluated all combinations of the same hyper-parameters as in Sect. 2.2.4 and selected the combination with the best performance.

In order to account for the lower accuracy of field estimates compared to laboratory measurements, we assigned larger weights to the laboratory measurements. Weight values of two, five, ten and fifteen times the weight of field estimates were tested. The final set of weights and other hyper-parameters was chosen based on the lowest root mean squared error (RMSE; Eq. 3.3) across the cross-validation. When the increase in RMSE was below 0.1 %, the model with fewer trees was chosen to reduce computation time. Note that an alternative way to account for differences in observation quality would be to derive the weights from the measurement error and residual variance, as in error-filtered machine learning (van der Westhuizen et al., 2022).

The final quantile regression forest used for model prediction was fitted using all soil observations in the calibration set (n = 15 312 laboratory measurements and 840 638 field estimates), 33 covariates and the final set of hyper-parameters, as optimized using random forest. We used the `ranger` package (Wright & Ziegler, 2017) with the option "quantreg" to grow a quantile regression forest and without it to grow random forest models.

3.3.5 Variable importance

During model fitting, we used impurity as a measure of variable importance (SI2). Impurity assesses the total reduction in heterogeneity that a covariate generates on the response variable. It is calculated by summing up all the reductions in the heterogeneity index in the tree nodes where a covariate was selected for splitting (Breiman, 2002). It is important to note that impurity has a bias towards covariates with more distinct values, making it negatively biased towards categorical covariates, as they have a finite number of binary splits due to their limited number of classes (Sandri & Zuccolotto, 2008, 2010). While impurity was used in this study, the more appropriate permutation measure to assess variable importance is dependent on the out-of-bag error (Breiman, 2002). As we assigned larger weights to lab measurements, there were not enough unselected soil samples available to calculate the out-of-bag error, making it impossible to use permutation to measure variable importance.

3.3.6 Prediction maps

The calibrated quantile regression forest was used to derive the median (0.50 quantile; $q_{0.50}$), 0.05 quantile ($q_{0.05}$) and 0.95 quantile ($q_{0.95}$) at every 25 m pixel and each standard depth layer specified by GSM (Arrouays et al., 2015) for every year t from 1953 to 2022 over the Netherlands. In addition, spatially explicit 90 % prediction intervals (PI90) were obtained at every 25 m pixel as a measure of prediction uncertainty as follows:

$$PI90 = q_{0.95} - q_{0.05} \tag{3.1}$$

Absolute mass percentage changes in SOM (ΔSOM) over the 70 year period were mapped by subtracting the 2022 and 1953 SOM [%] median prediction maps. For prediction, the depth covariates were equal to the upper and lower boundary and midpoint of each standard depth layer specified by GSM (Arrouays et al., 2015). However, note that the model can predict at any depth, so in order to analyze changes in SOM over smaller depth increments, we also predicted SOM [%] at 5 cm depth intervals between 0 and 2 m along a transect (Fig. 3.3e & f). The location of this transect was chosen such that it contained fen and bog peat soils of varying thickness, reclaimed land and mineral soil under different land use types.

3.3.7 SOM changes based on soil type and land use

In order to gain insight into average changes in peat vs. mineral soils and the dominant land uses in the Netherlands (grassland, cropland and forest), we computed average ΔSOM values between 1953 and 2022 for each combination of these soil types and land uses for the top 30 cm (Table 3.2). This depth interval was chosen because the topsoil is usually most relevant for agricultural and ecological purposes. Weighted averages were computed for the upper layers using the ΔSOM maps from 0 - 5 cm, 5 - 15 cm, 15 - 30 cm.

In order to compare our findings with other studies (Fernández-Ugalde et al., 2020; Reijneveld et al., 2009; Chardon et al., 2009; Conijn & Lesschen, 2015; Knotters et al., 2022), conversions from soil organic carbon to SOM were necessary. We used the same conversion factor that was used for soils in the Netherlands in Knotters et al. (2022), so soil organic carbon values reported in other studies were multiplied by 2.000. Note that this conversion ratio depends on the soil type and we did not account for its uncertainty, as shown for soils in the Netherlands in Fig. 53 of van Tol-Leenders et al. (2019). If other studies reported values in g/kg, values were divided by 10 to obtain absolute mass percentages.

3.3.8 Model accuracy assessment

In order to assess model accuracy, we used prediction uncertainty, cross-validation (1953 - 2011), design-based inference (1993 - 2000 & 2018) and non-design based inference to evaluate SOM temporal changes (ΔSOM; Table 3.4).

Prediction uncertainty

At the location, depth and year of a SOM measurement, all quantiles from 0 to 1 at steps of 0.01 were predicted to obtain the PI90 (Eq. 3.1) as well as the prediction interval coverage probability (PICP) of all prediction intervals. The PICP is the proportion of observations that fall into the corresponding prediction interval (Papadopoulos et al., 2001). It is an indication of how accurately quantile regression forest quantifies uncertainty. Prediction

uncertainty using PI90 is an example of model internal accuracy assessment since it is model (quantile regression forest) dependent.

Cross-validation (1953 - 2011)

In order to obtain an overall indication of 3D+T SOM mapping accuracy, we used a location-grouped 10-fold cross-validation with only laboratory measurements from the 1953 - 2011 dataset in the hold-out folds, similar as during model tuning (see above). Cross-validation is commonly used for digital soil mapping assessment (Piikki et al., 2021) and was also used in other space-time soil mapping studies (Heuvelink et al., 2020).

To obtain commonly used accuracy metrics, median predictions were used to calculate residuals. From these residuals we estimated the mean error (ME or bias), the RMSE and the model efficiency coefficient (MEC):

$$\widehat{ME} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{y}_i) \tag{3.2}$$

$$\widehat{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2} \tag{3.3}$$

$$\widehat{MEC} = 1 - \frac{\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \tag{3.4}$$

where n is the number of observations, y_i and $\widehat{y}_i$ are the i^{th} observation and prediction, respectively, at a certain location, depth and year, and $\overline{y}$ is the mean of all test set observations. We computed these accuracy metrics for all observations (Fig. 3.4a) and separated into observations in 0 - 30, 30 - 100 and 100 - 200 cm depth layers (Table 3.3), as the latter was necessary for design-based inference (see below). The midpoint of the depth layer was used to designate it into the corresponding depth layer. For example, an observation from 20 - 60 cm, having a midpoint of 40 cm, was designated to the depth layer 30 - 100cm.

Design-based inference (1993 - 2000 and 2018)

Since the 1993 - 2000 and 2018 datasets are probability samples over 2D space (Sect. 3.3.1), we used design-based inference to compute accuracy metrics for 0 - 30, 30 - 100 and 100 - 200 cm depth layers (Table 3.3), in the same manner as in Sect. 2.2.6. This included the lower and upper 97.5 % confidence limits of the accuracy metrics, which together give the 95 % confidence intervals (de Gruijter et al., 2006, Sect. 7.2.4). Design-based inference using a probability sample is recommended for map validation because it yields unbiased estimates of the accuracy metrics and allows computing confidence intervals (Brus et al., 2011; Wadoux et al., 2021a; Brus, 2022).

Evaluation of SOM temporal changes (ΔSOM)

Using the ΔSOM validation dataset (Sect. 3.3.1 and locations in Fig. 3.3c), we computed the difference between a SOM measurement in 2022 and a legacy SOM measurement called Δy as follows:

$$\Delta y \;=\; y(x, y, d, t_{2022}) \;-\; y(x, y, d, t_i) \tag{3.5}$$

where y is a SOM observation, x and y are the coordinates of a location, d is the sampled depth and t_i is the year between 1953 and 1999 when the legacy soil sample was collected before at that location and depth. We compared predictions to observations and also computed accuracy metrics using Eq. 3.2 - 3.4, except that y_i and $\widehat{y}_i$ were the i^{th} ΔSOM measurement and prediction, respectively (Fig. 3.4b).

3.4 Code and data availability

An adapted version of the code and data used in this scientific manuscript is available at the following repository: `https://git.wur.nl/helfe001/bis-4d_masterclass`, which also includes a hands-on tutorial (00_TUTORIAL_guide.html) and an extensive readme describing the code scripts and datasets. In the interest of facilitating reproducibility in the modelling workflow and decreasing the computational demand, some minor adjustments were made compared to the original code and data used for the research presented herein. These adjustments within the code and data repository encompass the following:

(a) The resolution of covariates and prediction maps were aggregated to 1 km, as opposed to the 25 m resolution utilized in the paper.

(b) SOM field estimates from the BPK dataset (Table 3.4) were not included in model calibration (Sect. 3.3.1).

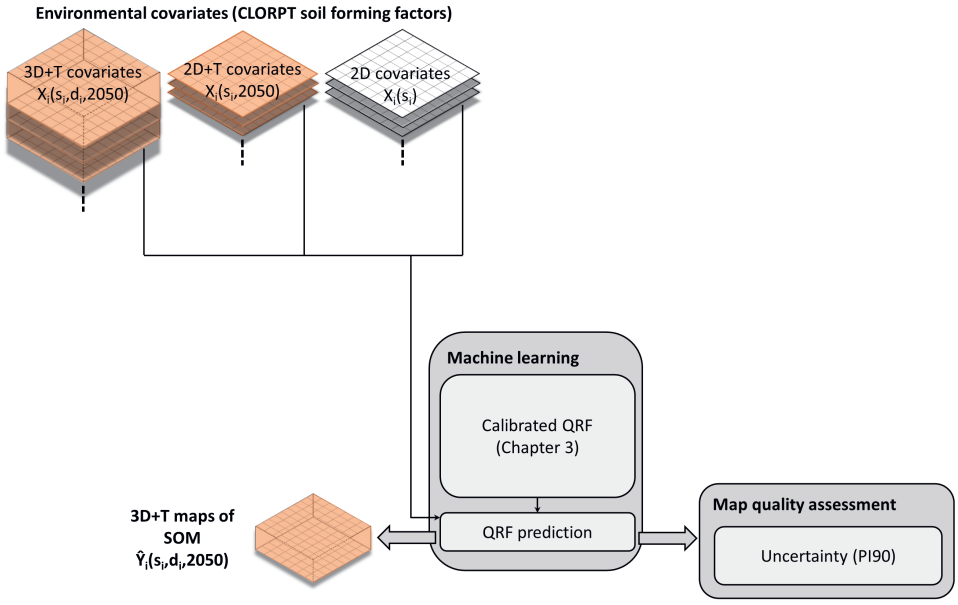
(c) The LSK and CCNL datasets (Table 3.4) are not included in the repository, as they are not publicly available.

Due to these changes, the results (e.g. prediction maps and accuracy metrics) are also slightly different than in the paper. However, besides these adjustments, the underlying code remains unaltered.

For an overview of soil data in the Netherlands from the Dutch National Key Registry of the Subsurface (BRO, in Dutch), please visit bodemdata.nl. Soil data in the Netherlands is maintained within the BIS database by Wageningen Environmental Research (Bodemkundig-Informatie-Systeem-BIS-Nederland.htm).

Chapter 4

A nature-inclusive future with healthy soils? Mapping soil organic matter in 2050 in the Netherlands



This chapter is based on:

Helfenstein, A., Mulder, V.L., Hack-ten Broeke, M.J.D., Breman, B., 2024. A nature-inclusive future with healthy soils? Mapping soil organic matter in 2050 in the Netherlands. *European Journal of Soil Science*. Under review.

Abstract

Nature-inclusive scenarios of the future can help address numerous societal challenges related to soil health. As nature-inclusive scenarios imply sustainable management of natural systems and resources, land use and soil health are assumed to be mutually beneficial in such scenarios. However, the interplay between nature-inclusive land use scenarios and soil health has never been modelled using digital soil mapping. We predicted soil organic matter (SOM), an important indicator of soil health, in 2050, based on a recently developed nature-inclusive scenario and machine learning in 3D space and time in the Netherlands. By deriving dynamic covariates related to land use and the occurrence of peat for 2050, we predicted SOM and its uncertainty in 2050 and assessed SOM changes between 2022 and 2050 from 0 - 2 m depth at 25 m resolution. We found little changes in the majority of mineral soils. However, SOM decreases of up to 5 % were predicted in grasslands used for animal-based production systems in 2022, which transitioned into croplands for plant-based production systems by 2050. Although increases up to 25 % SOM were predicted between 0 - 40 cm depth in rewetted peatlands, even larger decreases, on reclaimed land even surpassing 25 % SOM, were predicted on non-rewetted land in peat layers below 40 cm depth. There were several limitations to our approach, mostly due to predicting future trends based on historic data. Furthermore, nuanced nature-inclusive practices, such as the adoption of agroecological farming methods, were too complex to incorporate in the model and would likely affect SOM spatial variability. Nonetheless, 3D-mapping of SOM in 2050 created new insights and raised important questions related to soil health behind nature-inclusive scenarios. Using machine learning explicit in 3D space and time to predict the impact of future scenarios on soil health is a useful tool for facilitating societal discussion, aiding policy making and promoting transformative change.

4.1 Introduction

International organisations such as the Intergovernmental Panel on Climate Change (IPCC) and the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) call for urgent action and transformative change to address the challenges that negatively affect our planet, such as climate change and loss of biodiversity (IPBES, 2019; Pörtner et al., 2021). For transformative change, we need approaches that address the interdependent challenges in an integrated way to avoid negative trade-offs and feedbacks (Larrosa et al., 2016). One such approach is envisioning nature-inclusive scenarios for the future in order to help us resolve challenges we are facing today (Keesstra et al., 2018; Sowińska-Świerkosz & García, 2022).

In the Netherlands, a scenario of a nature-inclusive society for the National Nature Outlook 2050 was jointly developed by the Netherlands Environmental Assessment Agency and Wageningen University & Research (Breman et al., 2022). In this scenario, a narrative was developed in which more nature-inclusive types of land use could help to tackle several topical and urgent societal challenges, such as 1) nature conservation and biodiversity, 2) climate change, 3) quality of living, 4) farming transition, 5) energy transition and 6) water quality. Nature-inclusive transformations could have a big potential in the Netherlands. For example, a farming transition has the potential to increase the functioning of ecosystem services and improve the quality of life. In the Netherlands, historic land use changes were mainly conducted with the aim to intensify agriculture. 17 % of the present day land surface was reclaimed from water and 70 % of peatlands have disappeared in the last 2000 years (Erkens et al., 2016; Vos et al., 2020). Today, the Netherlands is the second largest exporter of agricultural products in the world (Jukema et al., 2023) and has the highest livestock density of all EU member states (Eurostat, 2022, p. 32). While this resulted in short-term economic growth, it had numerous negative effects for the environment and human well-being, such as nitrogen pollution and water eutrophication (Stokstad, 2019; de Vries et al., 2021). Consequently, parts of society are demanding a transformation to more sustainable practices (Erisman, 2021; Aarts & Leeuwis, 2023). The nature-inclusive scenario for 2050 addresses these and other challenges in an integrated way and would allow an increase in the provision of multiple ecosystem services and the quality of the human environment (Breman et al., 2022).

Soils play a pivotal role in the delivery of ecosystem services and the quality of the human environment. An increase in the provision of multiple ecosystem services largely depend on the soil's capacity to function within natural or managed ecosystem boundaries, to sustain plant and animal productivity, maintain or enhance water and air quality and promote plant and animal health (Lehmann et al., 2020; Creamer et al., 2022). Understanding the spatial variability, the current condition and the potential of the soil is essential for adopting nature-inclusive planning. In return, more nature-inclusive land use could also enhance soil health, defined as the continued capacity of soils to support ecosystem services

(European Commission, 2021), as such an approach implies sustainable management of and investing in natural systems and resources (Doorn et al., 2016). Thus, nature-inclusive scenarios may be beneficial for implementing pressing soil health initiatives like the Soil Deal for Europe and the recent Directive on Soil Monitoring and Resilience (European Commission, 2021, 2023b). In summary, soil health and nature-inclusive land use are deemed mutually beneficial.

To the best of our knowledge, the interplay between soil health and nature-inclusive land use scenarios has not been studied using digital soil mapping (DSM). DSM is the computer-assisted production of soil type and soil property maps, using statistical models to infer the relationship between a soil property and spatially exhaustive environmental explanatory variables (McBratney et al., 2003; Scull et al., 2003). While some studies have mapped temporal changes in soil properties (Chapter 3; Gasch et al., 2015; Stockmann et al., 2015; Hengl et al., 2017a; Sanderman et al., 2017; Stumpf et al., 2018; Huang et al., 2019; Szatmári et al., 2019; OpenGeoHub et al., 2021, 2022), few have used DSM for modelling future scenarios. Gray & Bishop (2016, 2019) used DSM to map soil properties in south-eastern Australia until 2070 based on projected climate change scenarios. Yigini & Panagos (2016) mapped soil organic carbon stocks in Europe in 2050 based on climate and land use scenarios. These studies were based on likely climate, and for the latter, land use projections, as opposed to scenario modelling based on future visions assuming the immediate adoption of sustainable practices.

In this study, we used the nature-inclusive land use scenario for 2050 (Breman et al., 2022) and a DSM model in 3D space and time (3D+T; Chapter 3) to predict soil organic matter (SOM) and its uncertainty at 25 m resolution between 0 - 2 m depth for 2050 in the Netherlands. SOM is linked to six of the eight mission objectives of the Soil Deal for Europe (European Commission, 2021), increasing SOM is one of the main challenges related to soil health (Vanino et al., 2023). Moreover, in this study we demonstrate how it also links to various societal priorities addressed in the National Nature Outlook 2050. SOM and absolute changes in SOM between 2022 and 2050 (ΔSOM) were expressed as mass percentages. Our aim was to explore whether a nature-inclusive scenario for 2050 is conducive to enhancing SOM-related soil health.

4.2 Methods

4.2.1 Nature-inclusive scenario for 2050

The nature-inclusive outlook was one of three scenarios that were developed to explore the future of nature and related ecosystem services in the Netherlands (Hinsberg et al., 2020; Breman et al., 2022). In contrast to the other scenarios that focused mainly on biodiversity goals by protecting natural habitats and species, in the nature inclusive scenario, nature and its related ecosystem services were to be enhanced as much as possible throughout the

entire country, not only in protected nature areas. The starting point was the upscaling of existing and promising nature-inclusive practices, such as:

- Greening of cities and ecological design and management of urban green spaces;
- Rewetting peatlands to mitigate further land subsidence and CO_2 emissions (Fig. 4.1c);
- Stream valley restoration for increasing water storage, reducing flood risk, improving water quality and enhancing biodiversity;
- Transition to more agroecological and plant-based production systems where possible, in order to improve the efficiency of food production, enhance biodiversity (at soil, crop, parcel and landscape level) in agricultural systems and reduce emissions from animal-based production systems;
- Adding trees, hedges and ponds to the landscape to sequester carbon, store water and create corridors and stepping stones for biodiversity;
- Increasing plant biodiversity along river dikes, roadsides and train tracks to enhance drought resistance, strengthen natural corridors and biodiversity as a whole;

In the nature-inclusive scenario, these existing nature-based solutions were upscaled and implemented to a national level in 2050, based on detailed knowledge of landscapes and soils and the overarching principle that "function follows form". For example, peatlands were mainly rewetted where the starting depth of a peat layer was within the uppermost 40 cm depth (Fig. 4.1d), based on the soil landscape map (van Delft & Maas, 2022, 2023). Plant-based agricultural production was concentrated in areas with fertile soils suitable for crop growth, whereas animal-based production was concentrated in less productive areas where it can often be combined with other functions (Breman et al., 2022).

4.2.2 3D+T SOM model

In this study, we used an existing high-resolution soil modelling and mapping platform for the Netherlands. Over the last few years, we have developed 3D maps for a wide range of soil properties, such as soil pH (Chapter 2). More recently, we extended the model to predict changes in SOM between 1953 - 2022 in 3D+T (Chapter 3). The 3D+T SOM model is based on well-established DSM practices, while also developing innovative and improved methods, such as assessing map accuracy using design-based statistical inference (Chapter 2) and developing novel covariates, or spatial-explicit environmental variables, to map SOM in 3D+T (Chapter 3). In the 3D+T SOM model, some of the covariates are static, such as soil-forming factors representing climate, topography and parent material. However, other covariates in the model are dynamic in 2D space and time (2D+T) or 3D+T. For example, land use change and peat occurrence were covariates which have a greater propensity for change over several decades than climate, topography and parent

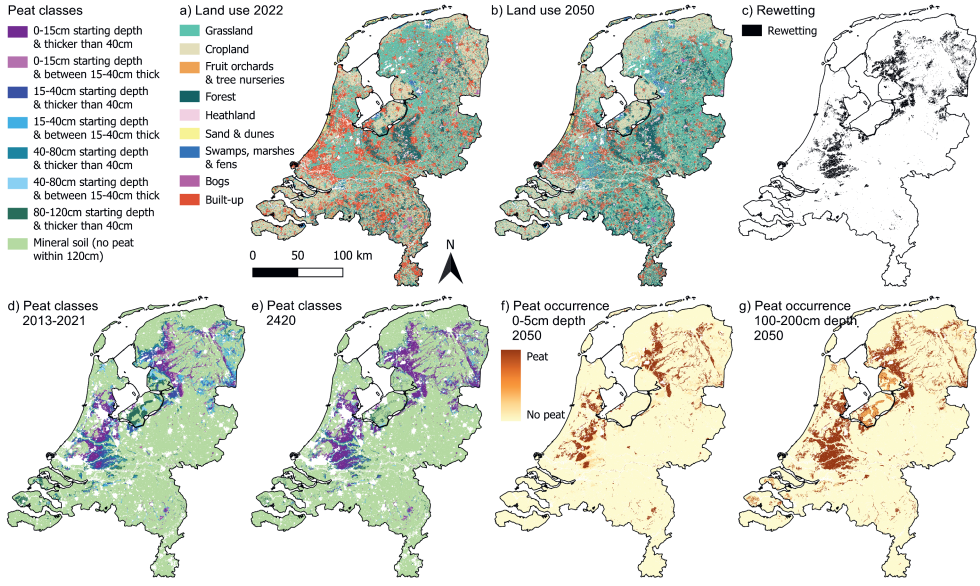


Figure 4.1: Land use in 2022 derived from Hazeu et al. (2023) (a) and in 2050 derived from Breman et al. (2022) (b), rewetted peatland areas (c), peat classes based on the 2021 version of the national soil map of the Netherlands (1:50 000; de Vries et al., 2003) (d), peat classes in 400 years (e), and peat occurrence in 2050 for 0 - 5 cm depth (f) and 100 - 200 cm depth (g). Land use in 2050 and rewetted peatlands are based on the nature-inclusive vision for the Netherlands in 2050 (Breman et al., 2022). Land use and rewetting in 2050 were in turn used to modify the map of peat classes for 400 years from now (e) and derive 3D+T dynamic peat occurrence covariates for 2050 (f & g).

material and were important for quantifying temporal SOM dynamics. In the model, land use (Fig. 4.1a & b) and peat classes (Fig. 4.1d & e) were variable in 2D+T, while peat occurrence was variable in 3D+T (Fig. 4.1f & g).

The 3D+T SOM model was calibrated using 869 094 SOM observations from 339 231 point locations in the Netherlands (approximately 10 locations per km^2) and 33 covariates, the latter of which were selected based on rigorous model tuning of hundreds of covariates relating to the soil-forming factors (Chapter 3). We used quantile regression forest (QRF; Meinshausen, 2006) to infer the relationship between SOM observations and the covariates. Ensemble decision tree models such as QRF have repeatedly outperformed other DSM models (e.g. Nussbaum et al., 2018) and QRF has the unique advantage that it delivers a probability distribution of the modelled response. Thus, the 90$^{\text{th}}$ prediction interval (PI90), i.e. the 95$^{\text{th}}$ quantile minus the 5$^{\text{th}}$ quantile, can be used as a measure of the prediction uncertainty. All information about the soil point data, covariates, model selection, tuning and calibration and model accuracy assessment using design-based sta-

tistical inference and spatially explicit prediction uncertainty is described in Chapter 3. In this study, we take the 3D+T SOM model a step further and explore to what extent it has the potential to simulate a future scenario.

4.2.3 2050 scenario modelling

Using the 3D+T SOM model, we predicted SOM in 2050 by deriving simulated, dynamic land use and peat covariates based on the nature-inclusive land use scenario for 2050. The nature-inclusive land use map for 2050 (Breman et al., 2022) needed to be reclassified to the same general land use classes that were used when calibrating the 3D+T SOM model (Fig. 4.1a, Chapter 3, Table 3.6), which resulted in the map shown in Fig. 4.1b. The 3D+T SOM model uses dynamic covariates of land use variable in 2D+T during year t, as well as the land use class that occurred most frequently in the 5, 10, 20 and 40 years prior to and including t. These modal classes were assigned to account for the delayed response of SOM to land use change (Chapter 3). However, since the land use between 2022 and 2050 was unknown in this simulated future scenario, we simply used the re-categorized nature-inclusive land use map for 2050 for all dynamic 2D+T land use covariates. This assumes that the envisioned land use changes were implemented already several years prior to 2050. More importantly, reclassifying land use led to the oversimplification of nuanced, nature-inclusive practices envisioned for 2050, particularly with regards to crop diversity and management practices. For example, it was not possible to distinguish land use and management practices such as strip cropping, biodiversity strips and alternative crops within the general "cropland" and "grassland" classes used in the 3D+T SOM model. In general, we were not able to incorporate numerous aspects of the nature-inclusive practices (Sect. 4.2.1) if they were not directly linked to land use, peat classes or peat occurrence as these were the only dynamic covariates used in the 3D+T SOM model.

Deriving simulated peat classes in 2D +T and future peat occurrence in 3D+T for 2050 proved more challenging than deriving land use and required making several general assumptions. In the 3D+T SOM model, covariates of 2D+T peat classes and 3D+T peat occurrence were derived from the peat class categories found in the national soil map of the Netherlands (1:50 000; de Vries et al., 2003). In the national soil map, soil type was mapped region by region between the 1960s and 1990s. Some regions, especially areas with peat soils, were updated between 2014 and 2021. For the 2050 scenario, we used the 2021 updated map of peat classes (Fig. 4.1d) as a starting point and assumed a peat growth rate of 1 mm/yr only in areas subject to peatland rewetting strategies in the nature-inclusive scenario for 2050 (Fig. 4.1c). A detailed comparison between the 2021 map of peat classes and the rewetting areas chosen based on the soil landscape map (Sect. 4.2.1; van Delft & Maas, 2022, 2023) revealed some discrepancies. For example, while all areas with peat starting at a depth between 0 - 15 cm and 15 - 40 cm were rewetted, only part of areas with peat starting between 40 - 80 cm depth and thicker than 40 cm were rewetted. Furthermore, none of the areas with peat starting between 40 - 80 cm depth and between

15 - 40 cm thick and peat starting between 80 - 120 cm depth were rewetted. In summary, peat growth was assumed only in areas where both of the following conditions were true: there already was a peat layer (Fig. 4.1d) and where rewetting occurred (Fig. 4.1c).

Based on literature, peat accumulation rates vary between 0.5 - 10 mm/yr (Witte & Van Geel, 1985; Charman, 2002; Joosten & Clarke, 2002; Höper et al., 2008; Stivrins et al., 2017; Craft, 2022), but 1 mm/yr is most commonly used as a general estimate. Hence, in order to change from peat starting between 15 - 40 cm depth to 0 - 15 cm depth (Fig. 4.1d), up to 25 cm of peat would need to grow under rewetted circumstances, which would take approximately 250 years. Similarly, to change from peat class starting between 40 - 80 cm depth to 15 - 40 cm depth, up to 40 cm of peat would need to form over approximately 400 years. As the latter was the longest time needed of any change between classes, a map of peat classes in 2420 was made (Fig. 4.1e). The 1 mm/yr peat accumulation rate is itself highly uncertain, partly because the estimated rate is based on natural peatland growth. Also, the land use oversimplification contributes to uncertainty, as some areas in the rewetted peatlands in the nature-inclusive scenario could be used for the production of crops suitable to these conditions, such as cattail, cranberries, reed and rice (Breman et al., 2022). Although crop growth under water saturated conditions would decrease the rate of or hinder peat mineralization, it is generally thought unlikely to lead to additional peat growth (Tanneberger et al., 2022).

Using the 2021 updated and the 2420 simulated peat classes maps (Fig. 4.1d & e), we derived 2D+T peat class covariates and 3D+T peat occurrence covariates of any year up to 2420 using fuzzy memberships, in the same manner as during model calibration between 1953 - 2022 and explained in Sect. 3.3.2. For our purpose, we thus derived 2D+T peat class covariates and 3D+T peat occurrence for 2050, the latter of which are shown for 0 - 5 cm depth and 100 - 200 cm depth in Fig. 4.1f & g.

The calibrated model from 1953 - 2022, the static covariates and the dynamic 2D+T and 3D+T land use and peat covariates for 2050 were used to predict SOM and its uncertainty for 2050 across six standard depth layers (0 - 5 cm, 5 - 15 cm, 15 - 30 cm, 30 - 60cm, 60 - 100 cm and 100 - 200 cm).

4.3 Results

4.3.1 SOM trends at the national scale

We predicted decreases of more than 1 % SOM on 22 % of the land surface (7390 km^2) and corresponding increases of more than 1 % SOM on 14 % of the land surface (4740 km^2) between 2022 and 2050 based on the nature-inclusive scenario. Additionally, we predicted decreases of more than 10 % SOM on 4 % of the land surface (1300 km^2) and concurrent increases of more than 10 % SOM on 2 % of the land surface (670 km^2) over these 28

years. Thus, under the nature-inclusive land use scenario for 2050, the prevalence of areas predicted to experience SOM decreases exceeds those showing an increase. ΔSOM maps shown for 0 - 5 cm (Fig. 4.2c), 15 - 30 cm (Fig. 4.3a) and 100 - 200 cm (Fig. 4.3b) support these findings.

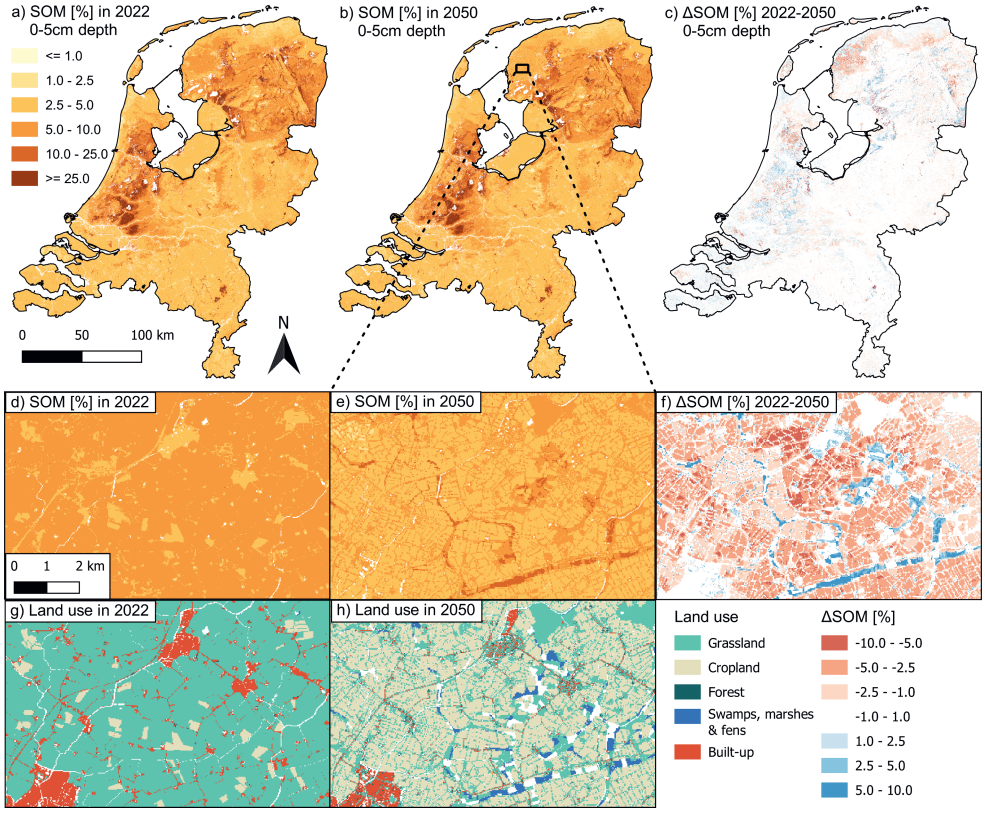


Figure 4.2: Predicted SOM [%] from 0 - 5 cm depth in 2022 (a), 2050 (b) and the difference in SOM between 2050 and 2022 (ΔSOM; c); zoom-in maps from the same depth layer and years (d-f) alongside land use (g & h) for an area in the province of Friesland.

4.3.2 SOM trends in mineral soils

For the majority of the regions with mineral soils in the Netherlands (Fig. 4.1d & e), there was little to no change in SOM between 2022 and 2050 (Figs. 4.2c, 4.3a & b). However, in the uppermost centimeters of some mineral soils, SOM decreased by up to 5 %, for example in the northern province of Friesland (Fig. 4.2). These were usually areas where grassland was turned into cropland based on the reclassified land use categories of the 3D+T SOM model. In areas where cropland in 2022 was also cropland in 2050, SOM

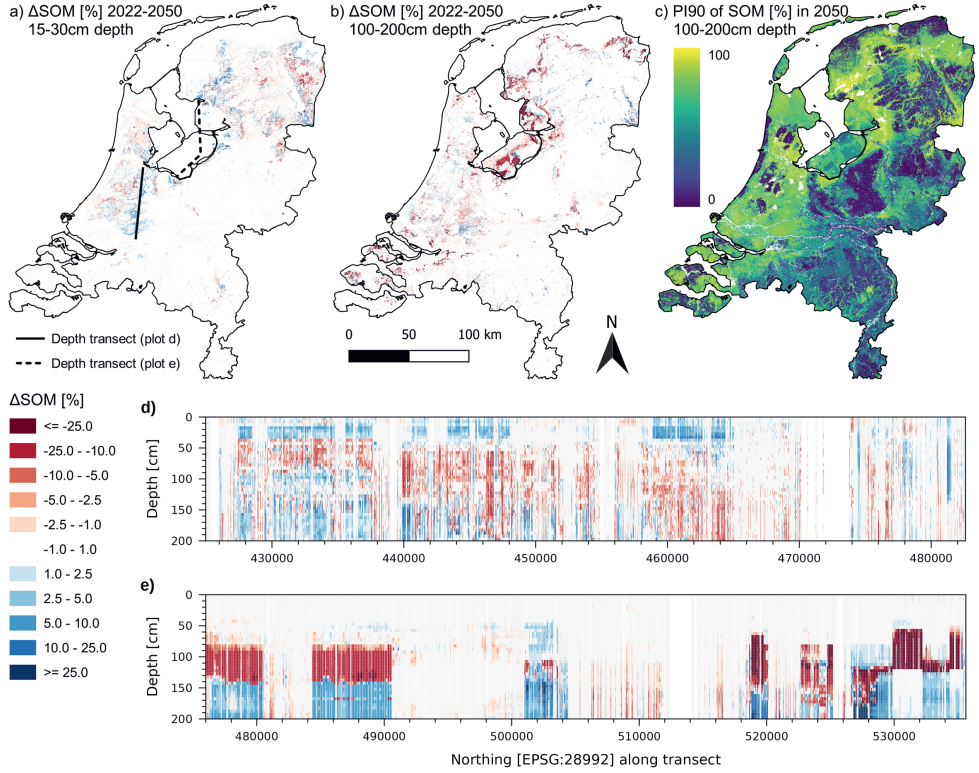


Figure 4.3: Predicted difference in SOM between 2050 and 2022 (ΔSOM) at 15 - 30 cm depth (a) and 100 - 200 cm depth (b); the 90th prediction interval (PI90) for SOM predictions [%] at 100 - 200 cm depth as a measure of uncertainty (c); ΔSOM depicted over depth [cm] vs. Northing [EPSG: 28992] in a region in the low-lying fen peatlands (d) and on reclaimed land in the province of Flevoland (e). The location of the depth transects (d & e) are shown in map a.

remained constant or decreased by less than 2.5 %. However, along narrow strips bordering crop parcels designated as grassland, SOM mostly remained constant or increased slightly if it was cropland in 2022. In the nature-inclusive scenario, these narrow borders were mostly envisioned as buffer and biodiversity strips along the edges of agricultural parcels (Breman et al., 2022). In addition, SOM increased up to 10 % between 2022 and 2050 in areas turned into nature reserves such as swamps and marshes in stream valleys and along waterways in the nature-inclusive scenario. Furthermore, SOM remained constant or slightly increased in built-up areas such as towns, cities and infrastructure, which can be explained by land use change from built-up to either grassland or forest as a result of greening in cities and alongside roads (Breman et al., 2022).

4.3.3 SOM trends in peatlands

In peatlands (Fig. 4.1d & e), we predicted more changes in SOM between 2022 and 2050 than in mineral soils (Figs. 4.2c, 4.3a & b). In peatlands that were rewetted in the nature-inclusive scenario (Fig. 4.1c), SOM increased by up to 25 % in the upper 40 cm (Fig. 4.3a & d). However, below 40 cm in rewetted peatlands, there was no clear pattern of SOM changes; SOM sometimes decreased and sometimes increased (Fig. 4.3b & d).

The 3D+T SOM model predicted the largest changes in SOM in areas with peat layers that were not rewetted (Figs. 4.1c, 4.3b & e). More specifically, these were in some of the areas where peat started between 40 - 80 cm depth and all areas where peat started below 80 cm depth (Fig. 4.1d). For example on reclaimed land in the province of Flevoland (Chapter 3, Fig. 3.1c), land subsidence caused peat layers below 80 cm to shift upwards (Brouwer et al., 2018), leading to SOM increase below 150 cm (Fig. 4.3e). However, above 150 cm depth, this shifting up of peat layers resulted in large decreases of SOM above 10 % and even above 25 % farther North along the transect in Fig. 4.3e.

4.3.4 Model uncertainty

Prediction uncertainty, provided by the PI90 of the predicted probability distribution of QRF, was very high for 2050 (Fig. 4.3c). Uncertainty was especially high where SOM predictions were high, e.g. in peatlands, and generally increased with increasing depth. One of the main limitations of the 3D+T SOM model is that it cannot provide uncertainty of ΔSOM because it does not account for cross- and spatial correlation in prediction errors (Chapter 3). These correlations can be accounted for by choosing a multivariate or geostatistical approach (Szatmári et al., 2021; van der Westhuizen et al., 2022; Wadoux & Heuvelink, 2023), but this was beyond the scope of this research. Nonetheless, while not demonstrated using our approach, we expect the uncertainty of ΔSOM to be high where the PI90 of SOM predictions for 2050 were also high (Fig. 4.3c).

4.4 Discussion

This study found that a majority of the nature-inclusive practices led to increases in SOM and were therefore beneficial for soil health (Table 4.1). SOM increased in areas designated as grasslands, forests, swamps, marshes, fens or bogs in 2050 as a result of the greening of cities, stream valley restoration and adding trees, hedges and ponds to the landscape. However, rewetting peatlands and transitioning to agroecological and plant-based farming systems were only partially beneficial for SOM-related soil health, as SOM increased and decreased as a result of these practices (Table 4.1). For example, SOM increased by as much as 25 % in the top 40 cm where peatlands were rewetted, but this effect varied at lower depths. Similarly, the conversion of cropland to grassland in mineral topsoils showed an increase in SOM, whereas the reverse land use change during the farming transition

resulted in a notable decreases up to 5 % SOM. It is crucial to acknowledge that the partially beneficial outcomes observed in rewetting peatlands and farming transitions may be influenced by limitations inherent in our scenario modelling method, as discussed further below. Additionally, the 3D+T SOM maps unveiled substantial SOM losses, sometimes exceeding 25 %, in expansive regions with peat layers below 40 or 80 cm depth. Largely located on reclaimed land, these areas were not rewetted and the subsoil was not influenced by other nature-inclusive practices, resulting in SOM loss due to factors such as land subsidence and peat oxidation. In summary, model predictions underscore the importance of implementing nature-inclusive practices for sustainable soil management. Moreover, SOM maps in 3D+T emphasize the potential consequences of neglecting nature-inclusive practices and their limitations for positively contributing to soil health at lower depths in the soil profile.

Table 4.1: Overview of the impact of nature-inclusive practices on SOM changes in 2050, based on the nature-inclusive land use scenario (Breman et al., 2022) and the 3D+T SOM model (Chapter 3). Based on these findings, we assessed whether nature-inclusive land use was beneficial for SOM-related soil health.

Nature-inclusive practice	**ΔSOM**	**Beneficial**
Greening of cities	Increase	Yes
Rewetting peatlands	Increase/decrease	Partially
Stream valley restoration	Increase	Yes
Transition to agroecological & plant-based production systems	Increase/decrease	Partially
Adding trees, hedges and ponds to the landscape	Increase	Yes
Increasing plant biodiversity along river dikes & transportation infrastructure	Unknown	Unknown

The nature-inclusive practices related to rewetting peatlands and transitioning to more agroecological and plant-based farming systems require a more nuanced evaluation. In the rewetted, low-lying, fen peatlands in the West of the Netherlands (Fig. 4.3d), increases up to 25 % SOM above 40 cm were due to dynamic peat class and peat occurrence covariates, which indicated that peat was accumulating (Fig. 4.1d-g). However, since peat started below 15 or 40 cm in these areas already in 2022, dynamic peat occurrence remained constant and indicated the presence of peat below these depth thresholds in 2050, while dynamic peat class covariates in 2050 were changing because of increasing peat thickness as a result of slow peat growth. Consequently, SOM decreases below 40 cm depth are likely attributed to the dynamic peat class covariates and are deemed implausible within this scenario. While the validity of this assumption could be examined by excluding 2D+T peat class covariates and relying solely on 3D+T peat occurrence, such an analysis was not conducted in this study, as the inclusion of both peat class and peat occurrence improved model performance during the calibration period (Chapter 3). When considering

these constraints, the modelling results support the notion that rewetting peatlands tends to increase SOM, a conclusion supported by numerous empirical field experiments (e.g. Ballantine & Schneider, 2009; Negassa et al., 2019).

Another major limitation was the 3D+T SOM model's inability to incorporate agroecological farming methods in plant-based production systems envisioned in the nature-inclusive scenario for 2050 (Breman et al., 2022). Methods such as conservation tillage, mulching, cover crops and especially growing crops where less soil disturbance is needed, such as perennial crops, have shown to achieve improvements in maintaining SOM in croplands compared to conventional methods (Crews & Rumsey, 2017). However, crop type and management practices such as tillage were not included as covariates. Moreover, the model was calibrated without accounting for sustainable management practices, as such practices were not the standard during the model calibration period (1953 - 2022; Chapter 3, Table 3.2). Extrapolating this conventional farming scenario into the future likely caused an overestimation of SOM losses in croplands and with the conversion of forest or grassland into cropland. In essence, the model represented a simplified version of the past reality, limiting its ability to predict a highly complex vision of a potential future reality. Yet, it also shows that if we do not transition to nature-inclusive farming systems, on the long term our soils will be less capable to provide multiple ecosystem services needed to sustain plant and animal productivity, maintain or enhance water and air quality and promote plant and animal health.

Limitations related to land use and management practices also applied to peatlands. Some rewetted peatlands areas in the nature-inclusive scenario would be used for paludiculture or crops suitable for growth under water saturated conditions (Breman et al., 2022). While this may prevent peat mineralization and lead to constant SOM levels, it is unlikely that new peat grows in these areas (Tanneberger et al., 2022). In summary, in its limited ability to account for nature-inclusive land use practices, our approach may have overestimated SOM decrease in mineral croplands and SOM increase in rewetted peatlands.

Another methodological limitation in our study was that climate was a static covariate. We included long-term minimum, maximum and average temperature and precipitation data between 1981 and 2010 as static covariates in our modelling approach (Chapter 2, Table 2.2). Hence, the temporal dynamics inherent in climate covariates, such as precipitation and temperature, were not accounted for in either the model calibration period (1953 - 2022) or the projection for the 2050 scenario, even though climate change affects carbon dynamics in the soil (Beillouin et al., 2022, 2023). While other DSM studies modelling future scenarios accounted for climate change (Gray & Bishop, 2016, 2019; Yigini & Panagos, 2016), we posited that, within our specified timeframe and under the prevailing conditions, the impacts of temperature and precipitation on SOM dynamics were of lesser consequence compared to changes in land use, peat class, and peat occurrence. The current time-frame is less than 30 years in the future, with an expected increase of

1.6°C and decrease of 17 mm (-2 %) in rainfall under a high emissions and dry scenario projected for the Netherlands (KNMI, 2023; van Dorland et al., 2023). Nonetheless, we recommend studies, especially ones over longer scenario timeframes, to include dynamic changes in covariates related to the climate. For our model, it would have required deriving dynamic covariates based on temperature and precipitation maps between 1953 - 2050 and recalibration of the model, which was outside of the scope of this study.

Despite the limitations in the model, mapping SOM in 3D space in 2050 and assessing SOM changes compared to 2022 in the context of a nature-inclusive scenario yielded valuable insights. Although regional soil conditions were considered for developing the nature-inclusive outlook for 2050 (Breman et al., 2022), this study creates new insights and raises important questions related to soil health about some of the notions and assumptions behind the scenario. For example, switching from animal-based to plant-based production systems is expected to bring many advantages for the environment and human well-being (Breman et al., 2022). Yet even with the adoption of agroecological practices, achieving SOM levels akin to those in permanent grasslands (e.g. pastures) within croplands presents a formidable challenge (Crews & Rumsey, 2017). Conversely, the conversion of less suitable croplands into grazing lands for extensive animal production systems is expected to offset this challenge.

Another valuable insight is that rewetting peatlands as a nature-inclusive practice prevented the continuation of substantial SOM decreases in expansive areas in the majority of fen peatlands in the West and bogs and brook valleys in the East that were predicted between 1953 - 2022 (Chapter 3). The major potential in rewetting peatlands lies in its potential to impede rapid and sustained SOM decrease, as we found that SOM mainly decreased where soils were not rewetted (Fig. 4.3b & e). Although preventing further SOM loss can be immediate or within a few years, peat growth in rewetted peatlands takes decades to centuries, operating on time-scales over multiple generations. Moreover, it is also dependent on the land use, for example natural peatland vs. peatland under paludiculture. The modelled scenario is only around 25 years in the future, but for some aspects of soil health to substantially change, so that humans and other organisms in return benefit from the ecosystem services that soils provide, a longer time window might be necessary.

Consistent with mapping SOM in 3D+T between 1953 - 2022 (Chapter 3), we also predicted substantial decreases in SOM on reclaimed land in the nature-inclusive scenario for 2050. It is improbable that nature-inclusive farming methods alone will suffice to prevent SOM decrease in these regions, given ongoing land subsidence, compaction and upward shifting of peat layers, and lower groundwater levels for part of the year, all of which contribute to SOM mineralization. Soils in areas with deep peat layers, not designated as peat soils in the soil landscape map (Sect. 4.2.1; van Delft & Maas, 2022, 2023), may currently by suitable for crop production. However, our 3D+T approach showed

that nature-inclusive scenarios only based on the dominating soil conditions in the topsoil may have severe consequences and lead to soil health deterioration if adopted by policy makers. In line with Chapter 3, this highlights the strengths of the 3D+T approach and inadequacy of evaluating soil health at point scale or static mapping at a single depth for policymaking.

Although negative trends in SOM-related soil health found over the last 70 years (Chapter 3) continued up to 2050 on reclaimed land, nature-inclusive practices benefited SOM in many areas, suggesting that, as opposed to a "business as usual" scenario, a nature-inclusive transition can improve soil health, thereby also benefiting society. The quantitative modelling of prospective scenarios, facilitated by our innovative 3D+T method, yields insights that may be valuable for guiding strategic spatial planning decisions. This is particularly relevant in the context of aligning with targets delineated in national policies such as the Climate Agreement of the Netherlands (Government of the Netherlands, 2019), as well as adhering to international frameworks like the European Soil Deal (European Commission, 2021), Green Deal (European Commission, 2023a), and Sustainable Development Goals (United Nations, 2015). Our findings underscore the potential of envisioning nature-inclusive transitions as a proactive and impactful approach to address soil health concerns and contribute to broader sustainability goals.

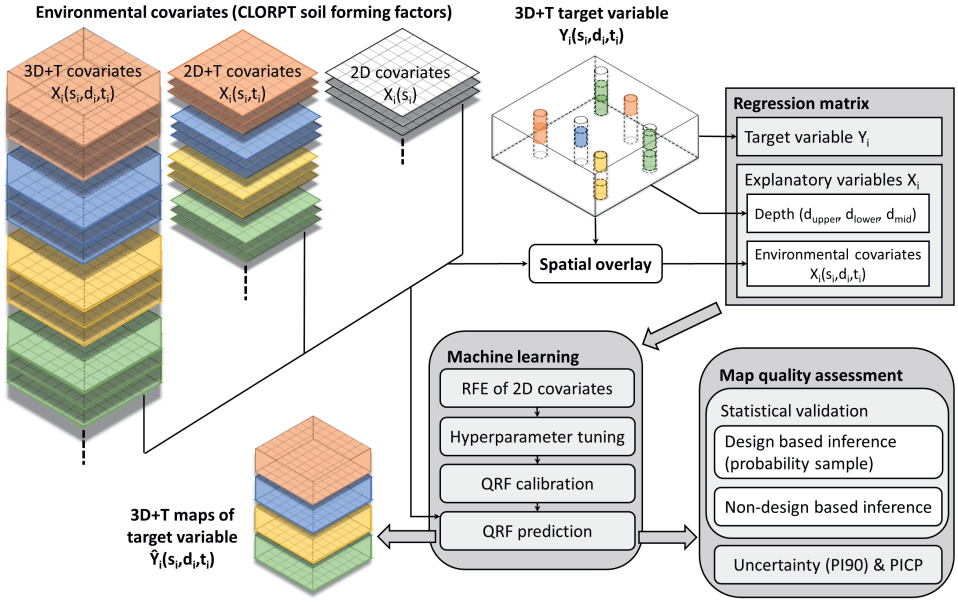
4.5 Conclusion

In this study, we demonstrated that 3D+T mapping of SOM for a future scenario is a pivotal tool to move from soil health-related proceedings to actions on a national scale. Beyond functioning as visual aids to underscore the societal importance of soils, our approach generated novel insights and prompted pertinent questions within the context of nature-inclusive scenarios. These insights require thoughtful consideration for the enhancement of soil health and the facilitation of broader societal transformations. By linking the nature-inclusive outlook to soils and thereby capturing the potential benefits and overlooked opportunities within spatial planning for soil-based ecosystem services, we have introduced an innovative and indispensable tool for policymakers. Space-time scenario modelling of soils not only aids in developing future plans but also provides a framework for gauging the temporal efficacy of implemented practices. However, it is equally imperative to emphasize the necessity of field monitoring and measurement to ensure the effectiveness of these practices over time.

In alignment with Breman et al. (2022), adopting nature-inclusive forms of spatial planning across the entirety of the Netherlands represents a major challenge. Key factors in realising this agenda will include a clear spatial policy strategy, sustainable business models, and a structured behavioural change. Despite the recognized challenges, we contend that ambitious visions stimulate a broader dialogue on the significance of soil health in the context of sustainable development and are catalysts for societal transformation.

Chapter 5

BIS-4D: Mapping soil properties and their uncertainties at 25 m resolution in the Netherlands



This chapter is based on:

Helfenstein, A., Mulder, V.L., Hack-ten Broeke, M.J.D., van Doorn, M., Teuling, K., Walvoort, D.J.J., and Heuvelink, G.B.M. BIS-4D: Mapping soil properties and their uncertainties at 25 m resolution in the Netherlands. *Earth System Science Data Discussions* 1–43 [preprint], `https://doi.org/10.5194/essd-2024-26`, accepted, 2024.

Abstract

In response to the growing societal awareness of the critical role of healthy soils, there is an increasing demand for accurate and high-resolution soil information to inform national policies and support sustainable land management decisions. Despite advancements in digital soil mapping and initiatives like GlobalSoilMap, quantifying soil variability and its uncertainty across space, depth, and time remains a challenge. Therefore, maps of key soil properties are often still missing on a national scale, which is also the case in the Netherlands. To meet this challenge and fill this data gap, we introduce BIS-4D, a high resolution soil modelling and mapping platform for the Netherlands. BIS-4D delivers maps of soil texture (clay, silt and sand content), bulk density, pH, total nitrogen, oxalate-extractable phosphorus, cation exchange capacity and their uncertainties at 25 m resolution between 0 - 2 m depth in 3D space. Additionally, it provides maps of soil organic matter and its uncertainty in 3D space and time between 1953 - 2023 at the same resolution and depth range. The statistical model uses machine learning informed by soil observations numbering between 3815 - 855 950, depending on the soil property, and 366 environmental covariates. We assess the accuracy of mean and median predictions using design-based statistical inference of a probability sample and location-grouped 10-fold cross-validation, and prediction uncertainty using the prediction interval coverage probability.

We found that the accuracy of clay, sand and pH maps was highest, with the model efficiency coefficient (MEC) ranging between 0.6 - 0.92 depending on depth. Silt, bulk density, soil organic matter, total nitrogen and cation exchange capacity (MEC = 0.27 - 0.78), and especially oxalate-extractable phosphorus (MEC = -0.11 - 0.38), were more difficult to predict. One of the main limitations of BIS-4D is that prediction maps cannot be used to quantify the uncertainty of spatial aggregates. A step-by-step manual helps users decide whether BIS-4D is suitable for their intended purpose, an overview of all maps and their uncertainties can be found in the supplementary information (SI[1]), openly available code and input data enhance reproducibility and future updates, and BIS-4D prediction maps can be easily downloaded at `https://doi.org/10.4121/0c934ac6-2e95-4422-8360-d3a802766c71` (Helfenstein et al., 2024a). BIS-4D fills the previous data gap of a national scale GlobalSoilMap product in the Netherlands and will hopefully facilitate the inclusion of soil spatial variability as a routine and integral part of decision support systems.

[1]Supplementary information of Chapter 5 is available at `https://doi.org/10.5194/essd-2024-26` under "Assets".

5.1 Introduction

Life on Earth, including that of humans, relies fundamentally on the availability and quality of air, water, and soil. These essential resources exhibit spatial variations in accordance with Tobler's first law of Geography, asserting that "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). However, the spatial heterogeneity of soil properties stands out prominently over short distances compared to air and water. This disparity arises from the multifaceted nature of soil, comprising solid, liquid, and gaseous phases, rendering it less mobile and unable to create homogeneous mixtures akin to air or water. Moreover, soil formation is a gradual process unfolding over hundreds to millions of years, shaped by intricate interactions between the climate, organisms (including humans), topography, and parent material (Dokuchaev, 1899; Jenny, 1941). Some of these soil-forming factors themselves exhibit high heterogeneity over short distances. Consequently, achieving a comprehensive understanding of soil spatial variability demands a high sampling density, a task hindered by the inherent difficulty, time consumption, and expense associated with collecting soil samples. These challenges underscore the complexity of quantifying soil variation, highlighting the formidable task of mapping soils in 3D space and time (3D+T).

With the rising awareness of soil health among diverse stakeholders such as governmental bodies and value chains (Lehmann et al., 2020), soil scientists are increasingly dedicated to deliver high-resolution, accurate soil maps. Internationally prominent examples of policies for which spatio-temporal soil information is essential include several of the Sustainable Development Goals, such as "Zero hunger" and "Life on land" (United Nations, 2015) and, in Europe, the Green Deal, Common Agricultural Policy and Zero Pollution (Panagos et al., 2022b). The importance of soil information for these policies has led to the EU Soil Strategy for 2030, the Soil Deal (European Commission, 2021) and most recently, the Proposal for a Directive on Soil Monitoring and Resilience (European Commission, 2023b). For such policies to have an impact, it is essential that soil scientists deliver information required to facilitate land use decisions and management practices at multiple scales.

In the Netherlands (land area = 33 481 km^2), the demand for soil information is also large. Located in the midst of Europe's largest delta, soils in the Netherlands are naturally very fertile (Edelmann, 1950; Römkens & Oenema, 2004). As one of Europe's most densely populated countries, multi-functional land use decisions made at national or regional level, need to be implemented at the field level, involving a broad range of diverse stakeholders. This spectrum of stakeholders collaborates on initiatives like the "Smart Land Use" project, which aims to sequester an additional 0.5 Mton CO_2-eq per year to Dutch mineral agricultural soils (Slier et al., 2023). Spatial information of soil properties can be used to evaluate soil health on Dutch agricultural fields using tools such as the Open Soil Index (OSI; Ros et al., 2022; Ros, 2023) and Soil Indicators for Agriculture

(BLN 2.0; Ros et al., 2023) and for assessing soil functions at different scales (Schulte et al., 2015). Information on soil texture and soil organic matter (SOM) are necessary for greenhouse gas reporting of the Land Use, Land Use Change and Forestry (LULUCF) sector for the United Nations Framework Convention on Climate Change and the Dutch LULUCF submission under the Kyoto Protocol (KP-LULUCF; Arets et al., 2020). Data of basic soil properties serve as inputs for modelling agricultural suitability (Mulder et al., 2022), crop precision agriculture (Been et al., 2023) and Soil-Water-Atmosphere-Plant interactions (SWAP; van Dam et al., 1997; Kroes et al., 2017). Furthermore, soil property maps contribute to initiatives such as the Watervision Agriculture and Nature (Hack-ten Broeke et al., 2019), Hydrological Instrumentations of the Netherlands (NHI, 2023) and Delta Program 2024 (Delta Programme, 2023).

Soil maps can also be used to identify and prioritize threats to soil health, as reviewed for the Netherlands by Römkens & Oenema (2004) and Hack-ten Broeke et al. (2009). Specific threats to soil health in the Netherlands include soil compaction (van den Akker & Hoogland, 2011; van den Akker et al., 2012), subsidence of peat due to oxidation and compaction (Brouwer et al., 2018; van Asselen et al., 2018), subsidence of young clay soils due to ripening on reclaimed land (Brouwer et al., 2018), and soil erosion (Hessel et al., 2011). Recently, Chapter 3 mapped SOM in 3D+T, which identified decreases in SOM at high resolution in 3D space. Spatial soil information is also crucial for agricultural businesses, both for optimizing fertilizer and manure applications for crop growth, but also for environmental accounting. The demand for such information is especially high in the Netherlands (Stokstad, 2019; Erisman, 2021; Aarts & Leeuwis, 2023), as it has the highest livestock density in the EU (Eurostat, 2022, p. 32) and ranks as the world's second-largest agricultural exporter (Jukema et al., 2023). An estimated 1 300 000 ha are phosphate saturated soils, where phosphate loss due to leaching exceeds ecologically tolerable limits (Römkens & Oenema, 2004). Hence, providing spatially explicit soil information is crucial to adhere to Targets 4.2 and 4.3 of the Soil Deal for Europe, which aim to reduce fertilizer use by at least 20 % and reduce nutrient losses by at least 50 % by 2030 (European Commission, 2021). In summary, the pressure of using soils sustainably in the Netherlands is immense.

Between the 1950s and 2000, conventional soil maps were completed in many countries. Today, the well-established discipline of digital soil mapping (DSM) has been widely adopted to meet the demands for accurate and high-resolution soil information for a wide range of purposes. Since DSM was first conceptualized (McBratney et al., 2003; Scull et al., 2003), maps of soil properties and soil types have been produced from local to global scales. These advances were propelled by initiatives like GlobalSoilMap (GSM) under the support of the International Union of Soil Sciences (Arrouays et al., 2014a; Hempel et al., 2014a; Arrouays et al., 2015) and the availability of openly accessible tutorials elucidating standard DSM workflows (Malone et al., 2017; Hengl & MacMillan, 2019; Brus et al., 2017; Brus, 2019, 2022).

Historically, the Netherlands was at the forefront of soil mapping. Scientific soil investigations in the Netherlands were started by Winand C.H. Staring in the mid-1800s followed by Jan van Baren and David J. Hissink in the early 1900s (Bouma & Hartemink, 2003). The first publication of the spatial distribution of soil properties in the Netherlands dates back to the 19th century (Felix, 1995). Systematic soil mapping became institutionalized with the establishment of the Dutch Soil Survey institute, or "Stichting voor Bodemkartering" (StiBoKa) in 1945 (Hartemink & Sonneveld, 2013). From 1950 to 1995, StiBoKa conducted conventional soil surveys (Buringh et al., 1962; de Bakker & Schelling, 1989; ten Cate et al., 1995) and produced regional maps (1:10 000 and 1:25 000 scale) and a national map (1:50 000 scale) of soil types (de Vries et al., 2003). After the development of DSM as a research field, various studies used (geo-)statistical methods to map qualitative and quantitative soil properties using the data collected by StiBoKa (Chapters 2 - 3; Brus & Heuvelink, 2007; Brus et al., 2009; Kempen et al., 2014; van den Berg et al., 2017). Several regions of the national soil map have since been updated (Kempen et al., 2009, 2011, 2012a; de Vries et al., 2014, 2017, 2018; Brouwer et al., 2018; Brouwer & Walvoort, 2019, 2020; Brouwer et al., 2021, 2023) and a variety of thematic maps were derived, such as a map of re-worked soils (Brouwer & van der Werff, 2012), a peat thickness map (Brouwer et al., 2018), a map of soil landscapes (van Delft & Maas, 2022, 2023) and the soil physical units map of the Netherlands (BOFEK; Heinen et al., 2022).

DSM has established itself and is routinely implemented across the world, but various challenges remain (Chen et al., 2022; Wadoux et al., 2021b). Maps of basic chemical, physical and especially biological soil properties are often missing (Chen et al., 2022; Wadoux et al., 2021b, challenge 8). Approximately 78 % of articles reviewed by Chen et al. (2022) mapped SOM, carbon content and carbon stocks. If a DSM product is available, predictions are often only made for one depth layer. Half of the studies reviewed by Chen et al. (2022) focused on soil properties at less than 30 cm depth only. However, users also require soil information at deeper depths and could benefit from models being able to predict at any desired depth in 3D, and for dynamic soil properties, in 3D+T (Chen et al., 2022; Wadoux et al., 2021b, challenge 5). In addition, there are numerous challenges relating to the accuracy of soil maps (Wadoux et al., 2021b, challenges 5 and 9). With regards to the accuracy, a major challenge is that the uncertainty of soil maps are often not quantified. A recent review showed that only 35 % of studies mapping continuous soil properties estimated prediction uncertainty (Piikki et al., 2021). Without providing the uncertainty of a map, users cannot determine its fitness for use. Moreover, assessing map accuracy is not straightforward and involves many demanding pre-requites, for example the sampling design of the locations used for statistical validation. According to Piikki et al. (2021), only 13 % of studies used probability sampling for map validation, which is the best approach for assessing map accuracy (Brus et al., 2011; Wadoux et al., 2021a; de Bruin et al., 2022). When using a soil map in a model or analysis, the uncertainty may be so large that it compromises the quality of the outputs of the model or analysis,

posing risks of erroneous conclusions and decisions for end users (Knotters & Vroon, 2015; Knotters et al., 2015a,b; Heuvelink, 2018). The efficacy of uncertainty propagation analysis relies on quantifying input uncertainty realistically, emphasizing the consistent need to quantify uncertainty in soil maps. The above challenges also apply to the Netherlands, where there is not yet a product that meets all these requirements.

To meet these challenges and demands, we introduce a high resolution soil modelling and mapping platform for the Netherlands called BIS-4D (Fig. 5.1). It delivers maps of key soil properties according to GSM specifications and assesses their accuracy using prediction uncertainty and statistical validation. The platform provides maps of soil texture (clay, silt and sand content), bulk density (BD), pH, total nitrogen (N_{tot}), oxalate-extractable phosphorus (P_{ox}) and cation exchange capacity (CEC) at 25 m resolution between 0 and 2 m depth in 3D space (Table 5.1). Furthermore, we provide maps of SOM in 3D+T between 1953 - 2023 at the same resolution and depth range, since SOM has changed substantially over time. Note that for soil pH and SOM, specific updates were made compared to previous versions (Sect. 5.2.7; Chapters 2 - 3). These nine soil properties were chosen based on those prioritized by GSM (Arrouays et al., 2014a; Hempel et al., 2014a; Arrouays et al., 2015), end-user needs in the Netherlands and data availability. In collaboration with soil surveyors, database maintainers and experts on Dutch soils from Wageningen University and Research, we assess the strengths and limitations of the BIS-4D maps and recommend potential map applications. Finally, model inputs, outputs (BIS-4D maps) and code, using free and open source software, are made available, easily accessible and well documented so that BIS-4D can be updated for future applications.

5.2 Materials and Methods

We predicted soil properties $\hat{Y}$ in 3D space, and SOM in 3D+T, using well established DSM methods (Fig. 5.1). BIS-4D uses machine learning to model the relationship between a soil property measured at point locations as the model response Y (Tables 5.1 - 5.3) and environmental covariates as the explanatory variables X (Table 5.5).

5.2.1 Soil point data

BIS-4D uses laboratory measurements and field estimates of soil properties from point locations collected in the Dutch soil database, or "Bodemkundig informatie systeem" (BIS). Definitions and laboratory measurement and field estimation methods for the soil properties mapped using BIS-4D are described in Table 5.1. We only included observations between 0 and 2 m depth excluding the O horizon (humus layer).

Note that clay, silt and sand content are particle size fractions (PSF) which together constitute soil texture. Thus, soil texture is a compositional variable: each PSF must be non-negative and together they must add up to 100 % (Pawlowsky-Glahn & Buccianti,

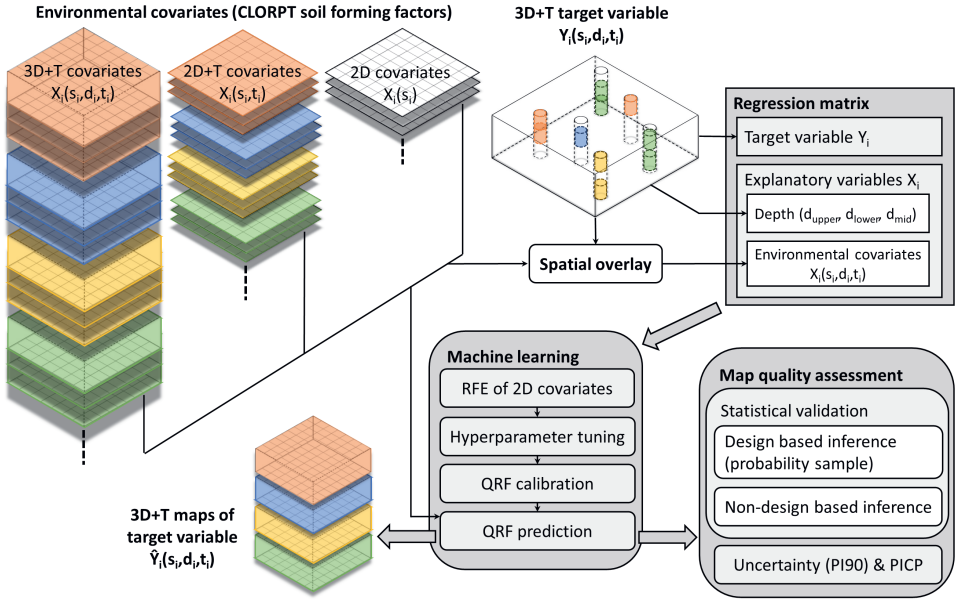


Figure 5.1: Graphical abstract of the BIS-4D soil modelling and mapping platform, where Y is a target soil property and X are covariates that vary in 2D space (s), depth (d) and, for SOM, in time (t). RFE = recursive feature elimination; QRF = quantile regression forest; PI90 = 90^{th} prediction interval width; PICP = prediction interval coverage probability (Sect. 5.2.3 - 5.2.6).

2011; Pawlowsky-Glahn et al., 2015). In order to achieve this, soil texture can be spatially interpolated as a compositional variable using geostatistical models (Odeh et al., 2003; Lark & Bishop, 2007; Wang & Shi, 2017), e.g. compositional kriging (de Gruijter et al., 1997; Walvoort & de Gruijter, 2001), machine learning (Akpa et al., 2014; Amirian-Chakan et al., 2019; Poggio & Gimona, 2017a; Poggio et al., 2021; Malone et al., 2021; Varón-Ramírez et al., 2022), and other techniques (Buchanan et al., 2012; Román Dobarco et al., 2017). Most commonly, these studies used the additive log-ratio transformation with the Gauss-Hermite quadrature (Aitchison, 1986). When not modelled as a compositional variable, other approaches include estimating two of the three PSFs and calculating the third by subtracting the sum of the two estimates from 100 % (Adhikari et al., 2013) or modelling all three PSF separately (Viscarra Rossel et al., 2015; Chagas et al., 2016; Mulder et al., 2016a; Taghizadeh-mehrjardi et al., 2016; Pahlavan-Rad & Akbarimoghaddam, 2018) and post-processing the predictions to ensure that they are all non-negative and sum to 100 %. For BIS-4D, we decided to model PSFs separately followed by post-processing (Sect. 5.2.5) for three reasons. Firstly, we wanted to use the additional locations where only one or two PSFs were observed (Table 5.2). Secondly,

Table 5.1: Acronyms, units and description of methods used for laboratory measurements and field estimates of target soil properties. *Mineral soil is defined as the dried soil fraction (105°C) put through a 2 mm sieve after removal of SOM and $CaCO_3$.

Soil property	Acronym	Unit	Description
Clay	-	%	Soil particles < 2 µm as a mass percentage of the mineral soil* fraction. Measured in the laboratory using the pipette method (NEN 5753, 2020) and estimated in the field following ten Cate et al. (1995); de Bakker & Schelling (1966, 1989).
Silt	-	%	Soil particles 2 - 50 µm as a mass percentage of the mineral soil* fraction measured in the laboratory using the pipette method (NEN 5753, 2020).
Sand	-	%	Soil particles 50 - 2000 µm as a mass percentage of the mineral soil* fraction measured in the laboratory using the pipette method (NEN 5753, 2020).
Bulk density	BD	g/cm^3	Dry bulk density of the oven-dry fine earth fraction.
Soil organic matter	SOM	%	Measured in the laboratory using loss on ignition at 550°C as a mass percentage of the mineral soil* fraction; or estimated in the field following ten Cate et al. (1995); de Bakker & Schelling (1966, 1989).
pH [KCl]	pH	-	Measured in the laboratory using pH in 1M KCl soil suspension.
Total N	N_{tot}	mg/kg	Measured in the laboratory mainly using Jodlbauer method (Maring et al., 2009, Appendix E, p. 79).
Oxalate-extractable P	P_{ox}	mmol/kg	Measured in the laboratory mainly using extraction with NH_4-oxalate at pH 3 (Maring et al., 2009, Appendix E, p. 81).
Cation exchange capacity	CEC	mmol(c)/kg	Measured in the laboratory mainly using extraction with silverthioureum or Ca-acetate at pH 6.5 (Maring et al., 2009, Appendix E, p. 81).

modelling soil texture as a compositional variable does not necessarily improve model performance (Amirian-Chakan et al., 2019). Thirdly, modelling separately followed by post-processing is easy to implement.

Soil point data for model calibration

We used laboratory measurements and field estimates from the "Boring Bodemkundig pakket" (BPK) and "Profielbeschrijving" (PFB) datasets in BIS for model selection, tuning and calibration (Tables 5.2 & 5.3, Fig. 5.3). Observations in BPK and PFB were made by soil horizon. Laboratory measurements and field estimates were available for all depths between 0 and 2 m (Table 5.3). All laboratory measurements were made at PFB locations. These locations are arranged in a purposive sampling design selected in the past to create the national 1:50 000 scale soil type map (de Vries et al., 2003). For the majority of the target soil properties, these locations covered soil variability in the

Netherlands well (Fig. 5.2). The majority of field estimates are part of the BPK dataset and are spatially clustered in specific areas for regional soil mapping purposes or specific projects (Chapter 3, Fig. 3.2). Most soil properties follow a skewed distribution, especially SOM, N_{tot}, P_{ox} and CEC (Fig. 5.3). However, pH, sand and to a lesser extent, silt, followed bimodal distributions. The distributions of the target soil properties likely affected model predictions (Sect. 5.3.1).

Table 5.2: Descriptive statistics of soil point data used for model calibration (field estimates and laboratory measurements) across all depths. Obs. = observations; Min. = minimum; Max. = maximum; Year = years during which observations were made. Minimum, median, mean and maximum values are in units of measurement of each soil property (Table 5.1). Soil point data used for model calibration is publicly available (Sect. 5.4).

Soil property	Dataset	Method	Locations	Obs.	Min.	Median	Mean	Max.	Year
Clay	PFB	Lab	3489	13 140	0	7	14.82	90.3	1953-2012
	PFB, BPK	Field	200 427	618 586	0	18	20.47	95	1955-2022
Silt	PFB	Lab	3376	12 912	0	17.8	24.29	97.5	1953-2002
Sand	PFB	Lab	3386	12 918	0	73.95	60.68	100	1953-2007
BD	PFB	Lab	951	3362	0.1	1.43	1.33	1.96	1957-1988
	PFB, BPK	Field	2586	12 509	0.1	1.5	1.49	2	1955-2002
pH	PFB	Lab	4216	15 248	0.9	4.8	5.2	9	1953-2010
SOM	PFB	Lab	4298	15 312	0	2.1	7	99.9	1953-2011
	PFB, BPK	Field	334 668	840 638	0	4	15.33	99	1954-2022
N_{tot}	PFB	Lab	2511	5739	0	1300	3287.38	36 700	1953-2003
P_{ox}	PFB	Lab	1655	6084	0	3.44	8.34	95.2	1955-2011
CEC	PFB	Lab	1332	3815	0	103	165.73	1541	1955-2010

Table 5.3: Number of laboratory measurements (lab) and field estimates (field) used for model calibration per standard GSM depth layer for each soil property.

Observation type	Depth [cm]	Clay	Silt	Sand	BD	pH	SOM	N_{tot}	P_{ox}	CEC
Lab	0-5	400	299	299	65	919	1049	765	311	556
	5-15	3844	3838	3840	3080	4524	5538	3258	2967	933
	15-30	1803	1794	1802	632	2519	2500	1200	961	502
	30-60	3731	3723	3725	2568	5392	5329	1192	3308	824
	60-100	4397	4294	4291	3630	5228	6170	1667	3972	728
	100-200	1262	1261	1258	1328	2329	2249	149	1566	272
Field	0-5	12 184	-	-	1547	-	18 873	-	-	-
	5-15	124 749	-	-	1372	-	230 710	-	-	-
	15-30	57 050	-	-	1360	-	117 800	-	-	-
	30-60	134 156	-	-	2242	-	209 918	-	-	-
	60-100	129 640	-	-	2395	-	138 122	-	-	-
	100-200	171 859	-	-	3593	-	130 836	-	-	-

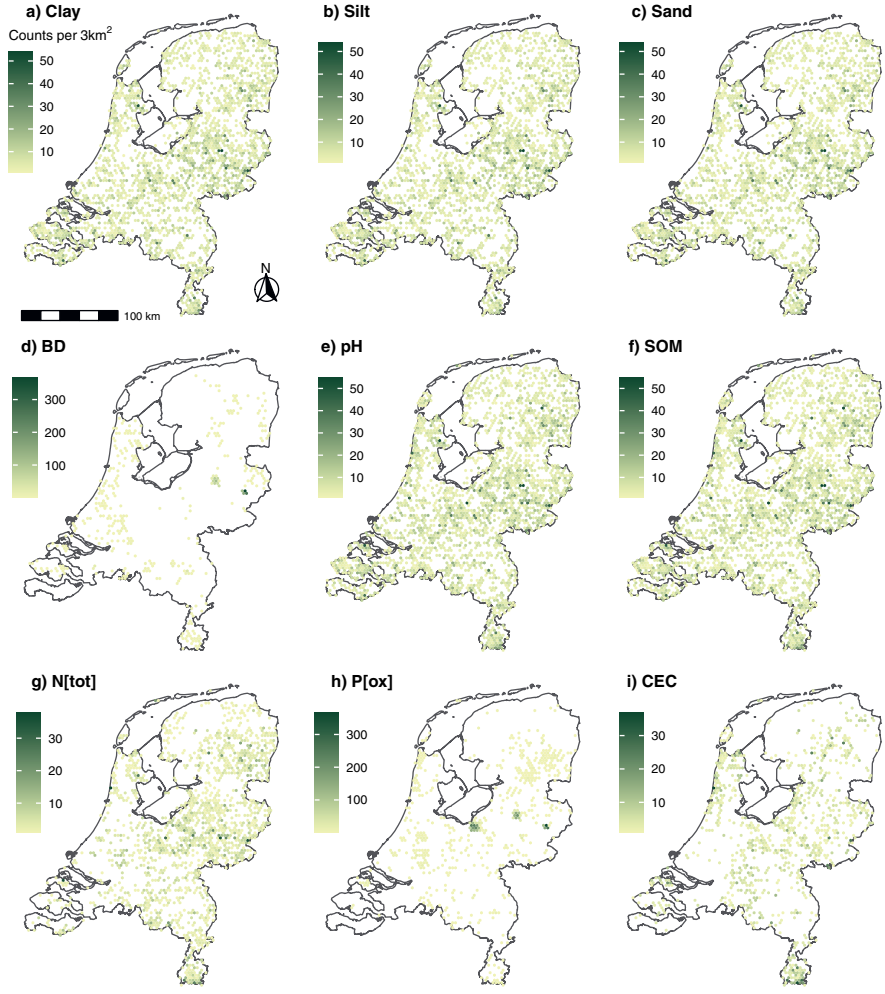


Figure 5.2: Observation density of locations with laboratory measurements used for model calibration of all BIS-4D target soil properties. All of these locations are part of the PFB dataset.

The laboratory measurements were deemed more important than field estimates because they are more accurate and locations with laboratory measurements were less spatially clustered. Nevertheless, field estimates from BPK and PFB also provide valuable information, expanding spatial coverage and, for SOM, also temporal coverage from 1953 - 2022 (Table 5.2). In addition, since around 2000, most observations that were added to the BIS are field estimates, a trend which is likely to continue into the future due to limited budgets for laboratory measurements. Other national mapping studies have also used

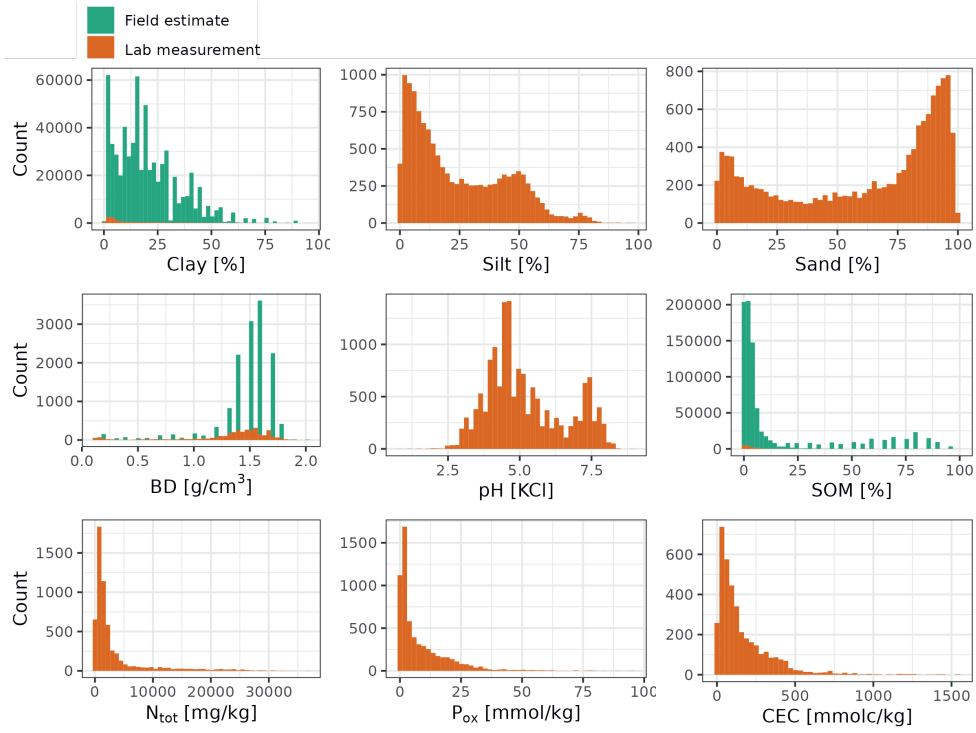


Figure 5.3: Histograms of soil property observations used for model calibration, colored by observation type.

field estimates in the past (van den Berg et al., 2017). We accounted for differences in data quality between laboratory measurements and field estimates using rigorous model tuning based on optimizing model performance (Sect. 5.2.3). Field estimates were removed if there was a laboratory measurement available from the same location and soil horizon (and year, in case of SOM). Methods for estimating clay content, BD and SOM in the field are described in ten Cate et al. (1995); de Bakker & Schelling (1966, 1989).

Soil point data for statistical validation

For clay, silt, sand and CEC, no separate dataset with laboratory measurements was available for statistical validation. Therefore, statistical validation of these four soil properties was conducted using PFB laboratory measurements and a cross-validation approach (Sect. 5.2.6).

For BD, pH, SOM, N_{tot} and P_{ox}, laboratory measurements from either the "Landelijke Steekproef Kaarteenheden" (LSK) or "Carbon Content NL" (CCNL) dataset were available for model validation (Table 5.4). LSK is a separate and independent dataset gathered

between 1993 and 2000, where locations were determined using probability sampling. The stratified simple random sample contains 94 strata defined based on soil type and groundwater class (Finke et al., 2001; Visschers et al., 2007), with the original purpose to validate the national soil type map (de Vries et al., 2003). Observations were made for each soil horizon. Statistical validation of BD, pH, SOM, N_{tot} and P_{ox} maps was conducted using LSK because map accuracy should preferably be estimated with design-based statistical inference using a probability sample (Brus et al., 2011). LSK data were also used to validate earlier versions of soil pH (Chapter 2) and SOM maps (Chapter 3).

Table 5.4: Descriptive statistics of separate soil point datasets used for statistical validation across all depths. Note that for statistical validation only laboratory measurements were used. Separate datasets were not available for clay, silt, sand and CEC. Obs. = observations; Min. = minimum; Max. = maximum; Year = periods during which observations were made.

Soil property	Dataset	Locations	Obs.	Min.	Median	Mean	Max.	Year
BD	LSK	1363	5644	0.17	1.43	1.29	1.69	1993-2000
pH	LSK	1363	5663	1.9	5.2	5.54	8.2	1993-2000
SOM	CCNL	1144	2284	0.5	3.4	7.51	78.7	2018
	LSK	1185	4952	0.1	2.5	6.52	93.6	1993-2000
	ΔSOM	63	276	0	1.9	9.97	96.9	1953-1995
N_{tot}	CCNL	1145	2286	0	1360	2784.85	24690	2018
P_{ox}	LSK	1480	6220	0	3.98	7.05	96.55	1989-2000

For SOM and N_{tot}, the CCNL dataset was used for statistical validation (Table 5.4). The CCNL dataset consists of all LSK locations that were still accessible in 2018. In contrast to LSK, CCNL locations were re-sampled at two fixed depth layers (0 - 30 cm and 30 - 100 cm) instead of by soil horizon. LSK and CCNL datasets were also used and their methodological sampling differences were explained in van Tol-Leenders et al. (2019); van den Elsen et al. (2020); Knotters et al. (2022). Since LSK was sampled by soil horizon, at more locations and also below 1 m depth, it is preferential to use it rather than CCNL.

For 3D+T maps of SOM, four different datasets were used for statistical validation with the specific purpose to assess SOM maps for specific years (Chapter 3): location-grouped 10-fold cross-validation of PFB data (1953 - 2011; lab measurements shown in Table 5.2 and Fig. 5.3), design-based inference using LSK (1993 - 2000), design-based inference using CCNL (2018) and a separate set of PFB locations that were re-sampled in 2022, used to assess changes in SOM over time (Table 5.4). Design-based inference and cross-validation procedures are explained in Sect. 5.2.6.

5.2.2 Covariates

In line with the DSM methodology (McBratney et al., 2003; Scull et al., 2003), we used 366 covariates as explanatory variables that were representative of the soil-forming factors:

climate, organisms, relief (topography), parent material (geology) and time (Dokuchaev, 1899; Jenny, 1941). Accounting for Tobler's first law of Geography (Tobler, 1970) and spatial auto-correlation, Easting (x-coordinate) and Northing (y-coordinate) were also included as covariates. Numerous studies have used spatial position and geographical distances as covariates (Li et al., 2011; Behrens et al., 2018b; Hengl et al., 2018; Møller et al., 2020; Sekulić et al., 2020). Sampling depth information, more specifically the upper and lower boundary and midpoint of each sampled horizon, were included as covariates so that predictions could be made at any chosen depth and depth interval. See Ma et al. (2021) for an overview of models using depth as a covariate in comparison to non-3D DSM methods. The majority of static covariates used in BIS-4D were previously used to map soil pH (Chapter 2). Others, mainly derivations of monthly mosaics from Sentinel 2 RGB and NIR bands, were added to map SOM (Chapter 3). In order to map SOM in 3D+T, we extended upon established methods by also deriving covariates variable in time (2D+T) and variable over depth and time (3D+T), as described in detail in Chapter 3. All covariates were resampled at 25 m resolution.

We created a regression matrix containing the BIS-4D target soil property observations and static covariate values by performing a spatial overlay. For SOM, this was extended to a space-time overlay for 2D+T covariates and a space-depth-time overlay for 3D+T covariates (Chapter 3).

5.2.3 Model selection, tuning and calibration

For model selection as defined by Hastie et al. (2009), we removed covariates in a two-step procedure using de-correlation followed by recursive feature elimination (RFE) as in Poggio et al. (2021). From any pair of covariates for which the Pearson correlation coefficient was > 0.85 or < -0.85, the covariate that was more correlated with all remaining covariates was removed. RFE (Guyon et al., 2002) was implemented using the `caret` package (Kuhn, 2019) and the number of covariates was chosen with the lowest root mean squared error (RMSE; Eq. 5.3). From 366 covariates, this resulted in a set of 20 - 50 covariates depending on the target soil property (Table 5.5), further used in model tuning, calibration and prediction.

For model tuning, we grew random forest (RF) models (Breiman, 2001) and optimized hyper-parameters for mean predictions. We tuned the model using a location-grouped 10-fold cross-validation, meaning that all measurements from the same soil profile location were forced to be in the same fold. Field estimates were excluded from the hold-out fold. We assessed all combinations of the same hyper-parameters as in Sect. 2.2.4 (Chapter 2) and chose the combination with the lowest RMSE (Eq. 5.3, Table 5.6).

For soil properties where both laboratory and field estimates were available (clay, silt, sand, BD and SOM), we also tuned whether designating a larger case weight for laboratory measurements improved model performance, in order to account for the lower accuracy of

Table 5.5: Covariates used during model calibration and prediction for different responses (soil properties), i.e. after covariate removal based on de-correlation and recursive feature elimination (RFE; Sect. 5.2.3). "All" implies that a covariate was used in tuning, calibration and prediction of all soil properties. Further information can be found in the metadata files and description of the provided covariates (Sect. 5.4).

Soil forming factor	Description	Source	Soil property
Soil	Peat classes starting depth and thickness	National soil map (de Vries et al., 2003)	Clay, BD, pH, N_{tot}, CEC
	Groundwater classes in agricultural areas; subsurface material in groundwater zones	de Gruijter et al. (2004); Hoogland et al. (2014); Knotters et al. (2018)	All
Climate	Long-term mean, min. & max. temperature	KNMI (2020)	BD, N_{tot}, P_{ox}, CEC
	Long-term mean precipitation	KNMI (2020)	Clay, silt, BD, SOM, N_{tot}, P_{ox}
Organism	Land use 1900, 1960, 1970, 1980 & 1986–2022	HGN (Alterra, 2004); LGN (WENR, 2020; Hazeu et al., 2020)	Clay, silt, sand, pH, SOM, N_{tot}
	Sentinel 2 RGB & NIR bands & spectral indices (2015-2022) as in Loiseau et al. (2019)	Roerink & Mücher (2023)	All
	Manure application, ammonia & total N emissions, management type	Besluit Gebruik Dierlijke Meststoffen (BGDM; RIVM, 2020); BIJ12 (2019)	Clay, silt, sand, BD, pH, P_{ox}, CEC
	Land cover & vegetation types	Bakker et al. (1989)	Clay, pH, CEC
	Forest vegetation types, tree species & age	de Vries & Al (1992); Clement (2001)	Clay, silt, sand, pH, P_{ox}, CEC
	Water drainage classes, areas behind dikes or not, riparian zone land cover	Maas et al. (2019)	Clay, silt, sand, pH
Relief	Digital elevation model (DEM) & derivatives	AHN (2023)	All
	Low- vs. high-elevation regions (binary)	Knotters et al. (2018)	Clay, silt, sand, pH
Parent material	Geological units/classes & chronostratigraphic formation period	Kombrink et al. (2012); van der Meulen et al. (2013)	Clay, silt, sand, pH, CEC
	Geomorphology based on geomorphological classes, genesis, form, formation time & relief	Koomen & Maas (2004); Maas et al. (2019)	Clay, silt, sand, BD, pH, SOM, N_{tot}, CEC
	Physical geographic regions & landscape types	EZK (2013)	Clay, silt, sand, BD, pH, SOM, N_{tot}, CEC
	(Paleo-) geographical maps (9000–250 B.C., 100–1850 A.D.)	Vos (2015); Vos et al. (2020)	Clay, silt, sand, BD, pH, SOM, N_{tot}
Spatial position	Easting & Northing	-	Clay, silt, sand, BD, pH, SOM, P_{ox}, CEC
	Upper, midpoint & lower boundary of soil layer	-	All
Time	2D+T dynamic covariates of land use (Chapter 3)	HGN (Alterra, 2004); LGN (WENR, 2020; Hazeu et al., 2020)	SOM
	2D+T & 3D+T dynamic covariates of peat classes & peat occurrence (Chapter 3)	Original (1960-1995) & updated (2014-2021) national soil map (de Vries et al., 2003)	SOM

Table 5.6: Final covariate count (post de-correlation and RFE) and optimized hyper-parameters for each modelled soil property. In instances without case weights, optimal performance was achieved excluding field estimates (silt and sand) or when the property was not estimated in the field (pH, N_{tot}, P_{ox}, and CEC).

Soil property	Number of covariates	Number of trees	Mtry	Min. node size	Sample fraction	Split rule	Case weight
Clay	50	500	12	1	0.8	Variance	5
Silt	50	500	10	1	0.8	Variance	-
Sand	50	500	10	1	0.8	Variance	-
BD	30	250	8	1	0.8	Variance	5
pH	50	500	12	1	0.8	Variance	-
SOM	33	500	7	1	0.8	Variance	10
N_{tot}	20	500	4	1	0.8	Variance	-
P_{ox}	30	500	6	1	0.8	Variance	-
CEC	50	500	10	1	0.63	Variance	-

field estimates compared to laboratory measurements. Values of two, five, ten and fifteen times the weight of field estimates were tested for laboratory measurements (Table 5.6). In addition, we also tested excluding field estimates entirely. The final set of hyper-parameters was chosen based on the lowest RMSE (Eq. 5.3) across the cross-validation. When the increase in RMSE was below 0.1 %, the model with fewer trees was chosen to reduce computation time during prediction. For silt and sand, model performance was highest when using only laboratory measurements, so field estimates were excluded in model calibration (Table 5.6).

For model calibration and prediction, we used RF to predict the mean and quantile regression forest (QRF) due to its ability to predict the entire conditional distribution (Meinshausen, 2006). The final QRF used for model prediction was fitted using all soil observations in the calibration set (Table 5.2), the selected covariates (Table 5.5) and the final set of hyper-parameters (Table 5.6).

5.2.4 Variable importance

During model calibration, we assessed variable importance using the permutation method for pH, N_{tot}, P_{ox}, CEC, silt and sand, and the impurity method for clay, BD and SOM. Permutation gives a better estimate of the variable importance than impurity because impurity has a bias towards covariates with more distinct values, making it negatively biased towards categorical covariates as they have a finite number of binary splits due to their limited number of classes (Sandri & Zuccolotto, 2008, 2010). However, the permutation measure is dependent on the out-of-bag error (Breiman, 2002). As we assigned larger weights to laboratory measurements for clay, BD and SOM models, there were not

enough unselected soil samples available to calculate the out-of-bag error.

5.2.5 Prediction maps

The calibrated RF and QRF and final set of covariates were used to estimate the mean, median (0.50 quantile; $q_{0.50}$), 0.05 quantile ($q_{0.05}$) and 0.95 quantile ($q_{0.95}$) at every 25 m pixel and each standard depth layer specified by GSM (0 - 5 cm, 5 - 15 cm, 15 - 30 cm, 30 - 60 cm, 60 - 100 cm and 100 - 200 cm) over the Netherlands. In addition, spatially explicit 90 % prediction interval widths (PI90) were obtained at every 25 m pixel as a measure of prediction uncertainty as follows:

$$PI90 = q_{0.95} - q_{0.05} \tag{5.1}$$

We post-processed the mean and median PSF prediction maps to ensure that the three PSF maps summed to 100 %. The predictions of clay, silt and sand were divided by the sum of the three at that location and multiplied by 100 for every 25 m pixel.

5.2.6 Accuracy assessment

We evaluated map quality using internal (model-based) and external (model-free) accuracy assessment. At the location and depth, and year in the case of SOM, of a soil property measurement, all quantiles from 0 to 1 at steps of 0.01 were predicted to obtain the PI90 (Eq. 5.1) as well as the prediction interval coverage probability (PICP) of prediction intervals between 0.02 and 1. The PICP is the proportion of independent observations that fall into the corresponding prediction interval (Papadopoulos et al., 2001). We refer to the PICP of the PI90 as the PICP90. The PICP is an indication of how accurately QRF quantifies uncertainty. Prediction uncertainty using PI90 is an example of a model internal accuracy assessment since it is QRF-dependent, whereas PICP is an external accuracy metric.

Besides PICP, we used two different statistical validation methods for an external accuracy assessment: 1) design-based inference (Brus et al., 2011; Brus, 2022), using either LSK or CCNL laboratory measurements, and 2) non-design-based inference using PFB laboratory measurements (Sect. 5.2.1, Table 5.4). We used the same approach as described in detail in Chapter 2 to adapt design-based inference for statistical validation of prediction maps at different depth layers. However, design-based inference was not used to assess clay, silt, sand and CEC predictions, as it was not measured in LSK or CCNL. For non-design-based inference, we used location-grouped 10-fold cross-validation of the PFB laboratory measurements, similar as during model tuning.

To obtain commonly used accuracy metrics, both mean and median predictions were used to calculate residuals. From these residuals we estimated the mean error (ME or bias),

the RMSE and the model efficiency coefficient (MEC):

$$\widehat{ME} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i) \tag{5.2}$$

$$\widehat{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2} \tag{5.3}$$

$$\widehat{MEC} = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{5.4}$$

where n is the number of validation observations, y_i and $\widehat{y}_i$ are the i^{th} observation and prediction, respectively, at a certain location, depth and year (for SOM), and $\overline{y}$ is the mean of all validation observations. Eq. 5.2 - 5.4 apply for non-design-based inference. The adapted equations for design-based inference are Eq. 2.5, 2.8 and 2.11. We computed these accuracy metrics for all observations and separated into observations pertaining to each depth layers, as the latter was necessary for design-based inference (Chapter 2).

In addition to rigorous quantitative accuracy assessment, we also evaluated the spatial patterns of BIS-4D prediction maps qualitatively by comparing them to existing soil maps in the Netherlands (de Vries et al., 2003; Brus et al., 2009; Schoumans & Chardon, 2015; van den Berg et al., 2017; Heinen et al., 2022; Knotters et al., 2022) and based on expert judgement. We acknowledge that qualitative evaluation was not definitive and indicative only.

5.2.7 BIS-4D updates: pH and SOM

Previous map versions of soil pH in 3D and SOM in 3D+T have recently been published using BIS-4D (Chapters 2 - 3). For soil pH, this version contains several important updates. Firstly, covariates of peat classes (de Vries et al., 2003), groundwater classes in agricultural areas (Knotters et al., 2018) and Sentinel 2 RGB and NIR bands and spectral indices (Roerink & Mücher, 2023) were added, all of which were selected and thus used for model calibration and prediction of the updated version (Table 5.5). We also included de-correlation and RFE to increase the signal to noise ratio and make models more parsimonious (Sect. 5.2.3). For 3D+T maps of SOM, we included the latest national land use map (year 2022) to derive the dynamic 2D+T land use covariates and predict SOM for the year 2023.

5.2.8 Software and computational framework

The computational framework of BIS-4D is entirely based on open source software and was operationalized on a Ubuntu 22.04 operating system with 48 cores and 128 GB working memory (RAM). Model input data (soil point data and covariates), scripts and model

outputs (BIS-4D soil property prediction maps and their associated uncertainty maps) are openly accessible (Sect. 5.4).

BIS-4D is mostly based on R (version 4.3.1; R Core Team, 2023), although GDAL (version 3.7.2; GDAL/OGR contributors, 2023) and SAGA-GIS (version 7.8.4; Conrad et al., 2015) were used during covariate preparation and processing because this massively decreased computation time compared to using similar functions in R. Further details about resampling, masking and processing of covariates and reclassification of categorical covariates can be found in Sect. 2.2.7 (Chapter 2). The indices necessary for the location-grouped 10-fold CV were made using the `CAST` R package (Meyer, 2023). The `caret` package (Kuhn, 2008, 2019, 2022) was used for tuning and selection of hyper-parameters. We used the `ranger` package (Wright & Ziegler, 2017) with the option "quantreg" to grow a QRF during calibration and without it to grow a RF during RFE and tuning. For predictions, the option "quantiles" was used to predict quantiles while the option "response" was used to predict the mean. A combination of the `ranger` and `terra` packages was used for predicting at all locations and depths. We used QGIS (version 3.32.3; QGIS Development Team, 2023) and the `rasterVis` (Lamigueiro & Hijmans, 2023) and `mapview` (Appelhans et al., 2023) R packages for exploratory and qualitative analysis and visualization of covariates and prediction maps. The computational workflow for all BIS-4D maps took approximately 5700 CPU-hours.

5.3 Results and Discussion

BIS-4D prediction maps for every GSM depth layer at 25 m resolution can be downloaded at `https://doi.org/10.4121/0c934ac6-2e95-4422-8360-d3a802766c71` (Helfenstein et al., 2024a). These include predictions of the mean, 0.05, 0.50 (median) and 0.95 quantiles and the PI90 of clay, silt, sand, BD, pH, N_{tot}, P_{ox} and CEC. For SOM, these prediction maps are available for the years 1953, 1960, 1970, 1980, 1990, 2000, 2010, 2020 and 2023 (Sect. 5.4). An overview of all prediction maps together with the associated accuracy metrics (ME, RMSE, MEC, PICP) and variable importances can be found in the supplementary information (SI), which is organized by target soil property.

5.3.1 Accuracy assessment

Quantitative accuracy assessment

The accuracy of the produced maps varied considerably depending on the soil property (Table 5.7, 5.8 & SI). Based on 10-fold cross-validation (Table 5.7), the accuracy of mean predictions over all depths for clay, sand, BD, pH and N_{tot} maps was highest (MEC > 0.70), followed by SOM and silt (MEC > 0.60). Mean predictions for P_{ox} and CEC were least accurate (MEC $= 0.54$ and 0.49, respectively). Design-based inference separated by depth layer confirms the high accuracy of pH prediction maps (Table 5.8). MEC values computed

for mean and median predictions using design-based inference were lower for BD (0.34 - 0.78) and N_{tot} (0.27 - 0.52) than when using 10-fold cross-validation. Mean and median P_{ox} maps were very inaccurate (MEC = -0.11 to 0.38) based on design-based inference. The large differences in accuracy between 10-fold cross-validation using PFB laboratory measurements and design-based inference using LSK laboratory measurements for BD and P_{ox} may be due to the clustered and limited spatial distribution of calibration data for those soil properties (Fig. 5.2d & h). Therefore, for BD and P_{ox}, metrics using 10-fold cross-validation are likely overly optimistic.

Table 5.7: Accuracy metrics of BIS-4D soil property maps using mean and median predictions, computed using 10-fold cross-validation (Sect. 5.2.6). Units of ME and RMSE are in units of the measured soil property (Table 5.1).

Soil property	ME (mean)	ME (median)	RMSE (mean)	RMSE (median)	MEC (mean)	MEC (median)	PICP90
Clay	-0.23	0.42	8.1	7.7	0.77	0.78	0.84
Silt	-0.28	0.59	12	13	0.62	0.57	0.91
Sand	0.35	-1.2	17	17	0.74	0.74	0.92
BD	-0.011	-0.032	0.21	0.22	0.71	0.68	0.86
pH	-0.010	-0.023	0.71	0.72	0.73	0.72	0.93
SOM	-1.0	0.97	9.5	9.7	0.64	0.64	0.88
N_{tot}	-37	390	2800	2900	0.72	0.69	0.91
P_{ox}	-0.33	1.5	7.5	7.7	0.54	0.52	0.92
CEC	-3.6	26	130	140	0.49	0.46	0.92

The RMSE and ME were low for most soil properties (Table 5.7 and SI). The RMSE of sand was higher than for clay and silt, even though the MEC of sand indicates higher model performance for sand than for silt. This can be explained by the high proportion of regions in the Netherlands with very high sand content ($> 75\,\%$), i.e. the Pleistocene sandy areas shown in pink in Fig. 5.4d & h. In comparison, laboratory measurements of clay and silt content were rarely $> 75\,\%$ (Fig. 5.3).

The differences in accuracy between mean and median prediction maps varied between soil properties. Based on 10-fold cross-validation, mean predictions were less biased than median predictions for all soil properties except SOM (Table 5.7). For soil properties where calibration data were positively skewed (Fig. 5.3), i.e. all soil properties except sand, BD and pH, the bias of mean predictions was negative, whereas the bias of median predictions was positive (Table 5.7). However, in contrast to the findings based on 10-fold cross-validation, design-based inference of N_{tot} revealed that median predictions were less biased (between -609 and 120 mg/kg; SI) than mean predictions (between -511 and -1408 mg/kg; SI). Higher accuracy of median predicted N_{tot} was also reflected in lower RMSE (Table S7) and higher MEC values (Table 5.8).

Table 5.8: MEC for mean and median predictions of BIS-4D soil property maps, separated by depth layer and computed using either 10-fold cross-validation (CV) of PFB laboratory measurements, or design-based inference (DBI) using LSK or CCNL data (Table 5.4). DBI for N_{tot} at 100 - 200 cm depth was not possible because soil samples were not collected below 100 cm in CCNL (Sect. 5.2.1 & 5.2.6). However, for this depth layer, CV metrics are included in the supplementary information (Table S7).

Statistical validation method		CV	CV	CV	DBI	DBI	DBI	DBI	DBI	CV
Prediction	Depth [cm]	Clay	Silt	Sand	BD	pH	SOM	N_{tot}	P_{ox}	CEC
Mean	0-15	0.84	0.70	0.80	0.39	0.71	0.52	0.44	0.25	0.59
	15-30	0.84	0.68	0.81	0.78	0.91	0.53	0.44	0.17	0.49
	30-60	0.77	0.62	0.75	0.54	0.73	0.34	0.27	-0.11	0.47
	60-100	0.69	0.54	0.67	0.49	0.74	0.46	0.27	0.04	0.38
	100-200	0.60	0.51	0.61	0.47	0.77	0.44	-	0.04	0.16
Median	0-15	0.84	0.67	0.79	0.34	0.71	0.48	0.52	0.20	0.56
	15-30	0.85	0.65	0.82	0.78	0.92	0.68	0.52	0.38	0.43
	30-60	0.79	0.58	0.75	0.54	0.72	0.27	0.41	0.05	0.42
	60-100	0.72	0.48	0.67	0.44	0.74	0.53	0.41	0.00	0.36
	100-200	0.63	0.44	0.61	0.41	0.76	0.54	-	0.11	0.26

Mean predictions are more sensitive to extreme values and outliers than median predictions. For instance, in mineral soils, the predicted conditional distribution of SOM, N_{tot}, P_{ox} and CEC was positively skewed and median predictions were usually smaller than mean predictions (e.g. von Hippel, 2005, Fig. 1). In peat soils, the opposite was the case. Here, the predicted conditional distribution of SOM, N_{tot}, P_{ox} and CEC was negatively skewed and median predictions were larger than mean predictions. For these soil properties, mean predictions were thus systematically higher than median predictions in mineral soils, whereas mean predictions were systematically lower than median predictions in peat soils.

The maps of prediction uncertainty (PI90) for every GSM depth layer revealed that uncertainty was high when mean and median predictions fell within a range with limited calibration data (SI). This meant that for most soil properties, uncertainty was high in areas where predictions were high due to the positively skewed distribution of observation data (Fig. 5.3). For example, the positive correlation between increasing uncertainty with increasing predictions can be clearly observed for clay and silt in Fig. 5.4e, f, i and j. The same positive correlation between predictions and uncertainty was observed for N_{tot} over depth (Fig. 5.5e & f). We found a similar pattern of high uncertainty in peatlands due to high predictions in these areas for SOM, P_{ox} and CEC. However, given its bimodal distribution, the uncertainty for sand was highest in areas where predictions ranged between 25 – 75 % (for example in the river areas) and uncertainty was comparatively low in marine clay areas (< 25 % sand) and Pleistocene areas (> 75 % sand) (Fig. 5.4c, g & k). Prediction uncertainty for most soil properties increased with increasing depth (e.g.

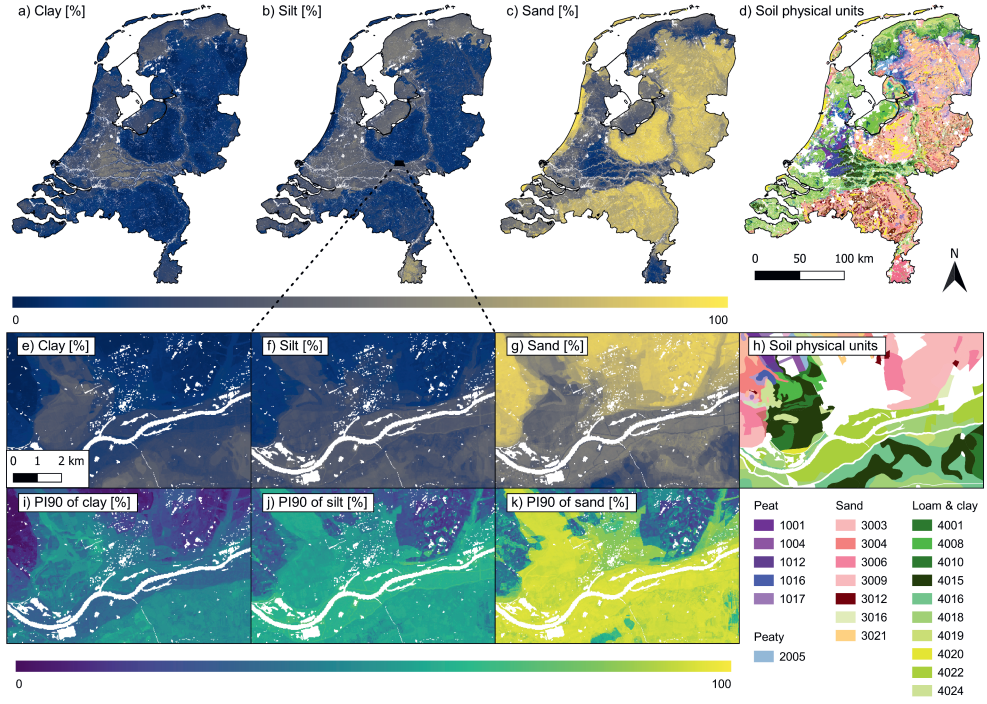


Figure 5.4: Mean predicted clay [%] (a & e), silt [%] (b & f) and sand [%] content (c & g) at 60 - 100 cm depth and associated prediction uncertainty (PI90 = 90th prediction interval) and the soil physical units map of the Netherlands (BOFEK, Heinen et al., 2022, d & h) in comparison. The soil physical unit codes can be found in Heinen et al. (2022); here grouped into the main categories (1xxx = peat, 2xxx = peaty, 3xxx = sand, 4xxx = loam/clay and 5xxx = loess). The zoom-in area around Wageningen was chosen since this area contains all main soil physical categories except loess.

Fig. 5.5f), except if mean and median predictions decreased substantially over depth, as was the case for P_{ox} (SI, Figs. S78 - S85). Higher uncertainty at lower depths is in line with worse accuracy metrics at lower depths (Table 5.8; SI) and this tendency was found in the majority of recently reviewed DSM studies (Chen et al., 2022). Finally, prediction uncertainty of most soil properties was also higher in urban areas, which can be attributed to limited soil samples and heavily disturbed soils in urban areas. With increasing population growth in an already densely populated country, this highlights the need to map urban soils (Römkens & Oenema, 2004; Vasenev et al., 2014, 2021; Kortleve et al., 2023).

The PICP90 (Table 5.7) and the PICP (SI) indicated that prediction uncertainty was estimated relatively accurate using QRF, but small differences were found among the pre-

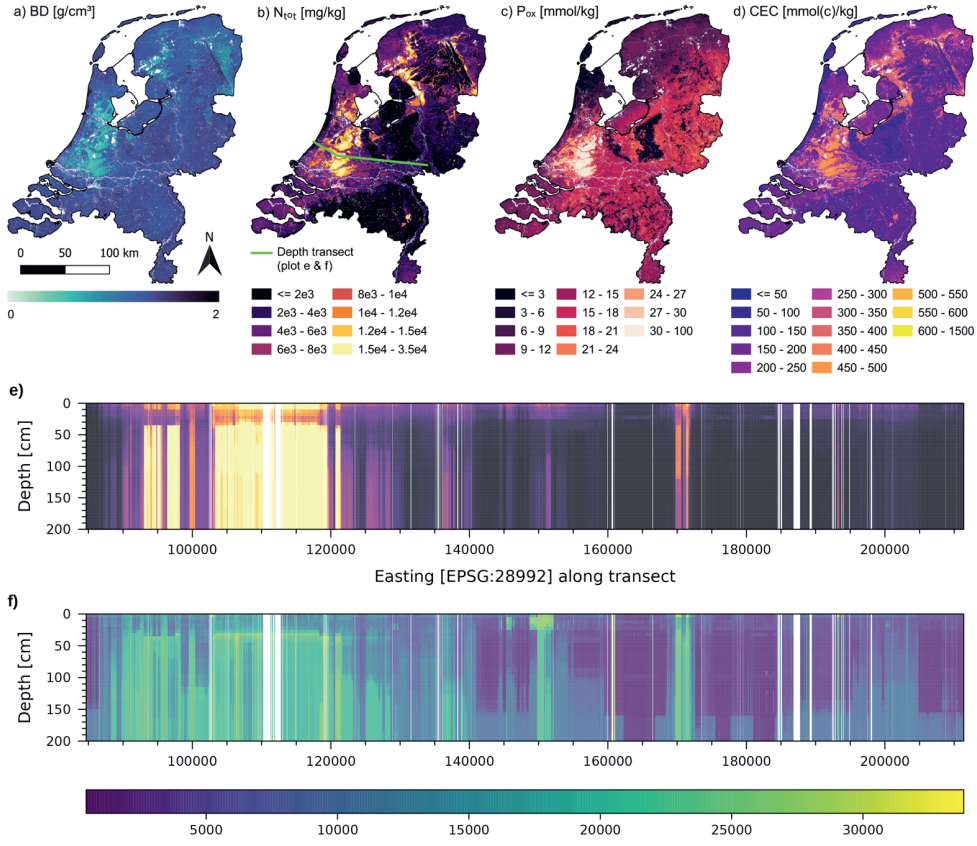


Figure 5.5: Median predicted BD [g/cm^3] (a), N_{tot} [mg/kg] (b), P_{ox} [mmol/kg] (c), and CEC [mmol(c)/kg] (d) at 0 - 5 cm depth; and median predicted N_{tot} (e) and PI90 (90th prediction interval) as a measure of the associated prediction uncertainty (f) along the depth transect shown in b.

dicted soil properties. For clay content, the PICP90 was between 0.82 - 0.86 (Table S1) and hence less than 0.90, indicating that the uncertainty of clay predictions was underestimated. The uncertainty of BD based on PFB laboratory measurements was slightly underestimated (0.86; Table 5.7), but was slightly overestimated based on LSK laboratory measurements (0.88 - 0.95; Table S4). For SOM, the PICP90 varied strongly with depth (0.75 - 0.96; Table S5), but the PICP overall was very accurate for all depths combined (Fig. S53). In our study, the soil properties for which field estimates were included during calibration were the only ones for which the PI90 was sometimes underestimated. Similarly, Chen et al. (2023) found that increasing the proportion of spectral estimates combined with conventional laboratory measurements decreased the PI90. Hence, if cali-

bration data are a smoothed version of the truth, which may be the case with predictions of spectral models and field estimates, this tends to lead to underestimation of the "true" uncertainty. The aim of sharp, i.e. narrow, conditional probability distributions by including various types of observational data is desirable only if ensuring that the uncertainty is still reliable, e.g. by computing the PICP (Schmidinger & Heuvelink, 2023). This is important to avoid presenting overoptimistic results to end users. Besides clay, BD and SOM, prediction uncertainty for the remaining target soil properties was accurate but marginally overestimated (0.89 - 0.97) based on the independent datasets used for statistical validation (LSK and CCNL; SI). Hence, the PICP indicates that silt, sand, pH, N_{tot}, P_{ox} and CEC maps are somewhat more accurate than suggested by the prediction uncertainty (PI90).

Qualitative accuracy assessment

The BIS-4D maps of the nine predicted soil properties align with the national soil map of the Netherlands (de Vries et al., 2003). This can be seen when comparing our maps (Figs. 5.4 & 5.5, SI) with the soil physical units map (BOFEK; Fig. 5.4d & h; Heinen et al., 2022), derived from the national soil map of the Netherlands. Further information on the comparison of BIS-4D maps to previous soil pH maps (Brus et al., 2009) can be found in Sect. 2.4.2 (Chapter 2), and to previous SOM maps (Brus et al., 2009; van den Berg et al., 2017; Knotters et al., 2022) in Chapter 3. Nonetheless, visual evaluation of the maps also revealed several limitations.

The maps of soil texture or particle size fractions (clay, silt and sand) of the mineral soil component should be used with caution in peatlands, since natural peat only consists of organic matter without a mineral component. However, the low-lying fen peatlands, located mostly in the West and Northwest of the Netherlands, typically also contain some clay, silt or sometimes even sand due to past flooding events (Edelmann, 1950; de Bakker & Schelling, 1966, 1989; Brouwer et al., 2023). Drained organic soils, particularly when under agricultural use, can also contain mineral components introduced or mixed in from mineral soil horizons from below or above the organic soil horizon. Nonetheless, 30 % clay content in a soil composed mostly of peat in absolute terms contains less clay than a mineral soil with 30 % clay content.

Visual examination of the BIS-4D maps reveals artifacts from the covariates. Although water and buildings were cropped out, some mapping artifacts remained, such as small buildings, roads and railways. For instance, the road on top of the dike, parallel to and South of the Rhine River is clearly visible in Fig. 5.4e-k. This highlights the difficulty of spatial modelling approaches such as DSM that rely strongly on remote sensing products. Other artifacts were due to the combination of several Sentinel 2 images from different days in one month to obtain one monthly mosaic (Sect. 5.2.2). Image mosaicing created artificial lines from images with more clouds, although overall, images contained very

few to no clouds. Finally, there were also orthogonal artifacts most likely due to using Northing and Easting coordinates as covariates, which can be largely removed by also including oblique axes in many additional directions (Møller et al., 2020).

5.3.2 Strengths

BIS-4D maps fill the missing data gap of spatial soil property information on a national scale in the Netherlands and bring substantial improvements to previously mapped soil properties. The main strengths of BIS-4D are: 1) the ability to provide information of soil properties as opposed to soil types; 2) the high spatial resolution (25 m); 3) accuracy and uncertainty assessment based on best practices; 4) the benefits of machine learning combined with large amounts of data; 5) the flexibility to predict in 3D and 3D+T; and 6) model code and data are openly available, making BIS-4D fully reproducible and easy to update.

The BIS-4D maps have several advantages compared to previous soil maps of the Netherlands. While categorical maps of soil type (de Vries et al., 2003) and derived thematic maps (Brouwer & van der Werff, 2012; Brouwer et al., 2018; van Delft & Maas, 2022, 2023; Heinen et al., 2022) are important and useful, many users require information on specific, numerical soil properties (Sect. 5.1). We acknowledge that clay content (Brus et al., 2009), SOM (Brus et al., 2009; van den Berg et al., 2017; Knotters et al., 2022), pH (Brus et al., 2009) and soil properties related to soil texture (Heinen et al., 2022) and P_{ox} (Schoumans & Chardon, 2015) have previously been mapped on a national scale in the Netherlands. However, these maps were at much coarser resolution, accuracy was either not assessed, or not assessed using design-based statistical inference, quantification and evaluation of uncertainty were missing, mapping approaches did not include machine learning, used only a few covariates, and predictions for one or several depth layers were modelled separately and only Knotters et al. (2022) assessed changes over time. The only standard GSM soil properties that we did not map are SOC, plant exploitable (effective) depth, depth to rock and coarse fragments (Arrouays et al., 2014a; Hempel et al., 2014a; Arrouays et al., 2015). We mapped SOM instead of SOC because, in the Netherlands, SOC was not included in routine soil analyses until recent years. However, SOC can be derived from SOM, as investigated in other studies in the Netherlands (van Tol-Leenders et al., 2019; van den Elsen et al., 2020; Teuling et al., 2021; Knotters et al., 2022). Plant exploitable (effective) depth is mostly limited by high groundwater levels in most regions of the country. Since groundwater levels have been extensively mapped in the Netherlands (de Gruijter et al., 2004; Hoogland et al., 2014; Knotters et al., 2018), mapping plant exploitable (effective) depth was not deemed necessary. Depth to rock and coarse fragments are not relevant on a national scale in the Netherlands, as the substrate materials of Dutch soils are almost exclusively either Pleistocene sand, fine-grained Quaternary sediments or peat.

Another strength of BIS-4D is that maps are at a high spatial resolution of 25 m. As covariates such as remote sensing products and national maps of land use (Hazeu et al., 2023) and digital elevation models (AHN, 2023) are nowadays available at 5 - 25 m resolution, useful information for modelling complex relationships between soil-forming factors such as land cover and topography and soil properties is provided at these scales. The increasing availability of high resolution information in soil-related domains has also increased the demand for high resolution soil maps. While high resolution products such as BIS-4D bring many advantages, it is crucial to emphasize that resolution is not an indicator of accuracy and should not be used solely to determine a map's fitness for use (de Bruin et al., 2001; Malone et al., 2013; Knotters & Walvoort, 2020; Szatmári et al., 2021).

One of the main advantages of BIS-4D is the rigorous map quality evaluation using design-based statistical inference and prediction uncertainty. Based on sampling theory (Cochran, 1977; de Gruijter et al., 2006; Gregoire & Valentine, 2007), map accuracy should be assessed with design-based statistical inference using a probability sample whenever possible, as this provides a better estimate of the "true" map accuracy compared to non-design-based approaches (Brus et al., 2011). Moreover, it also produces confidence intervals (Tables S4 - S8), so that we know how close the estimate of the map accuracy is to the true map accuracy. We were able to use design-based inference for BD, pH, SOM, N_{tot} and P_{ox} maps due to the availability of the LSK and CCNL datasets. We are not aware of any other GSM products that used design-based inference to evaluate map accuracy on a national scale. For soil properties for which design-based inference was not possible, i.e. for clay, silt, sand and CEC, we used location-grouped 10-fold cross-validation, as recommended in the case of non-clustered data (Wadoux et al., 2021a; de Bruin et al., 2022). In addition, BIS-4D maps provide spatially explicit estimations of prediction uncertainty (PI90), including GSM accuracy thresholds for soil pH (Chapter 2), and we evaluated the accuracy of the uncertainty using PICP.

Another strength of BIS-4D, for example when compared to previous soil property maps in the Netherlands (e.g. Brus & Heuvelink, 2007; Brus et al., 2009; van den Berg et al., 2017), is that machine learning leads to more accurate predictions than other geostatistical and regression techniques. Ensemble decision tree models such as RF and QRF have repeatedly outperformed other spatial interpolation methods (e.g. Hengl et al., 2015; Nussbaum et al., 2017; Keskin et al., 2019; Khaledian & Miller, 2020). Ensemble decision tree models are able to capture complex, non-linear relationships between the covariates and soil properties and are widely used in recent DSM studies (Vaysse & Lagacherie, 2017; Heuvelink et al., 2020; Poggio et al., 2021; Baltensweiler et al., 2021; Nussbaum et al., 2023).

BIS-4D maps for clay, silt, sand, BD, pH, N_{tot}, P_{ox} and CEC are in 3D (between 0 - 2 m depth) and for SOM, also dynamic (3D+T; SI and Chapter 3). This fills a largely missing

gap of soil information in deeper layers (Chen et al., 2022). In addition, BIS-4D can predict at any depth, as opposed to recalibrating models when mapping individual depth layers separately (Ma et al., 2021). This improves model flexibility and efficiency and a larger amount of data can be leveraged during model tuning and calibration. For example, routine agronomic soil sampling depths in the Netherlands are 0 - 10 cm for grasslands and 0 - 25 cm for croplands, thereby deviating from the GSM standard depths (Arrouays et al., 2014a; Hempel et al., 2014a; Arrouays et al., 2015). Predictions and associated uncertainty for those depths can be provided using BIS-4D without recalibrating models. This is particularly useful for uncertainty, which, unlike mean and median predictions, cannot be aggregated using e.g. weighted averaging over depth layers (Sect. 5.3.1). Finally, we developed innovative covariates explicit in 3D+T, presenting a novel opportunity to extend the predictive power of machine learning to 3D+T (Chapter 3). This provided a new opportunity for monitoring SOM-related soil health using a method that is explicit in 3D space.

Lastly, compared to the time-consuming effort of updating conventional soil maps, DSM products such as BIS-4D can easily be extended to other soil properties in BIS and can be updated and delivered on demand (Heuvelink et al., 2010; Kempen et al., 2009, 2012b, 2015). In comparison to an earlier version for soil pH (Chapter 2), the number of covariates has been substantially decreased during model selection (Sect. 5.2.3), which benefits reproducibility and possibilities to update maps. The model code, workflow, inputs and outputs are well documented and openly available, making procedures reproducible and easy to update (Sect. 5.4).

5.3.3 Limitations and improvements

Uncertainty in DSM products such as BIS-4D can be linked to three overarching sources: 1) the quantity and quality of soil point data, 2) the quantity and quality of covariates, and 3) the model structure (Heuvelink, 2014, 2018). Consequently, we discuss the limitations of BIS-4D maps with regard to uncertainties linked to soil point data, covariates, and model structure and suggest improvements to minimize these three sources of uncertainty.

Soil point data

Measurement errors and differences in measurement methods of the soil point data may have contributed to the uncertainty of BIS-4D maps. For example, Fe, Al and P extracted by oxalate extraction are considered to consist of amorphous Fe- and Al-(hydr)oxides and P bound to those oxides. However, a fraction of oxalate-extractable P in peat soils likely consist of P bound to organically complexed Fe and Al, since those are also partially extracted during the oxalate extraction (McKeague, 1967; McKeague et al., 1971; van der Zee et al., 1990; Schoumans, 2013; Schoumans & Chardon, 2015). Recent research has de-

vised methods to quantify the uncertainty of soil laboratory measurements (van Leeuwen et al., 2021) and to incorporate these errors into machine learning algorithms (van der Westhuizen et al., 2022). Furthermore, several slightly different methods, standards and laboratory facilities were used to measure N_{tot}, P_{ox} and CEC (Maring et al., 2009, Appendix E). This introduced uncertainty that can be minimized by standardizing laboratory measurements and procedures.

There were several limitations related to the spatial and spatio-temporal distribution of the soil point data used in BIS-4D. The calibration data of BD, P_{ox} and, to a lesser extent, CEC, were spatially clustered (Fig. 5.2), which most likely affected mapping accuracy of those soil properties (Sect. 5.3.1). In addition, no wet-chemical laboratory measurements were available as part of a probability sample (LSK and CCNL) for design-based statistical inference of clay, silt, sand and CEC prediction maps (Sect. 5.2.1). As most of the soil point data were collected between 1950 and 2000, soil measurement age and time should be addressed also for other soil properties besides SOM (Arrouays et al., 2017). N_{tot} and CEC are strongly linked to SOM and thus temporal changes may be similar to mapped SOM changes (Chapter 3). BD, pH, N_{tot}, P_{ox} and CEC likely changed due to land use and management. However, yearly variation in P_{ox} is relatively small since P binds strongly to soil particles and the plant available fractions of P with short turnover times are less than 15 % of the total reversibly bound P pool (Withers et al., 2014, Fig. 3), which is what is measured with P_{ox} (Lookman et al., 1995; Neyroud & Lischer, 2003). Large quantities of topsoil data are collected for agronomic surveys every four years in the Netherlands (BZK, 2022; Eurofins Agro, 2024a,b), but only a small part of these are not privacy-protected, making it challenging to incorporate in DSM approaches. Although the point data suggest good spatial coverage of most of the basic soil properties in the Netherlands, there is a major lack in repeated laboratory measurements collected using identical sampling strategies over time, as discussed in Knotters et al. (2022) and Chapter 3. A consistent national soil monitoring scheme would be beneficial for modelling dynamic soil properties in 3D+T, updating static BIS-4D maps and for accuracy assessment with more recent data.

Covariates

Although BIS-4D was able to make use of a large range of high-quality, country-specific covariates (Table 5.5), the main variable missing in our modelling approach is more detailed land management data, which is a common challenge in DSM (Finke, 2012; Arrouays et al., 2021). Land cover and land use covariates only indirectly provided information on land management. From 2005 onwards, annual data on the specific crop type for every agricultural parcel in the Netherlands was available ("BRP Gewaspercelen"; EZK, 2019), but these were never selected among the final covariates used for model calibration (and therefore not shown in Table 5.5). This implies that they did not provide additional useful information for the spatial distribution of the target soil properties at the national

scale, although different drivers may be relevant at regional and local scales (Sect. 5.3.4). However, regardless of the scale, national crop parcel data do not capture information on management decisions such as fertilizer inputs, liming and ploughing frequency on agricultural lands and maintaining forests and nature areas. These management decisions are highly relevant for many of the mapped soil properties, with the exception of particle size fractions. For example, BD is strongly dependent on the size and driving frequency of tractors on agricultural fields (Stettler et al., 2014).

As another example, P_{ox} exhibits considerable small-scale spatial variability, as discussed and made evident by the high nugget in the semivariogram in Fig. 6 of Lookman et al. (1995). As P in the form of phosphate is bound in the soil much stronger than N or other plant nutrients affected by the base cation saturation and CEC, there are large legacy effects due to historic management not captured in the covariates currently used in BIS-4D. In our study, the three most important covariates for modelling P_{ox} were the covariates related to soil horizon sampling depth (Fig. S88). The relationship of P_{ox} to soil depth is supported by empirical findings of the maximum P sorption capacity decreasing with soil depth, especially in sandy soils. Moreover, given that P_{ox} map quality was poor (Table 5.8 and SI), the relative importance of depth suggests that the other covariates did not explain the spatial variation of P_{ox} well, likely due to missing (historic) management data. Although not solving the problem of missing management data, one easy step to improve the accuracy of BD and P_{ox} and other management-dependent soil properties is to only map them for agricultural areas, as was done in the Netherlands for amorphous Iron- and Aluminium-(hydr)oxides (van Doorn et al., 2024). We expect that including dynamic covariates of land management and climate, as discussed in Chapter 3, would likely also improve modelling dynamic soil properties in 3D+T.

Model structure

Despite the many advantages of using QRF for DSM (Sect. 5.3.2), predictions may be further improved using methods such as convolutional or recursive neural networks (deep learning; Behrens et al., 2005, 2018a; Padarian et al., 2019b; Wadoux, 2019; Wadoux et al., 2019) or transfer learning (Liu et al., 2018; Padarian et al., 2019a; Seidel et al., 2019; Helfenstein et al., 2021; Baumann et al., 2021), defined as the process of sharing intra-domain information and rules learned by general models to a local domain (Pan & Yang, 2010). We recommend future research to investigate the use of deep learning and transfer learning in the Netherlands for SOM, due to the large amount of SOM data and more opportunities in accounting for differences in observational quality (field estimates and laboratory measurements) using more complex models. However, to the best of our knowledge, deep learning has only outperformed ensemble decision tree models when using a small number of covariates covering only some of the soil-forming factors, from which hyper-covariates are then derived (Wadoux, 2019). Hence, deep learning may not improve predictions in the Netherlands, where large amounts of high-quality covariates are readily

available for all soil-forming factors. In addition, quantifying model-based uncertainty using deep learning remains a challenge. Although model-free approaches of estimating uncertainty using deep learning have been used, e.g. involving bootstrapping (Padarian et al., 2019b; Wadoux, 2019), we are not aware of studies that have compared the accuracy of these uncertainty estimations to QRF-based uncertainty (PI90).

One of the main limitations of the BIS-4D modelling approach is that QRF predictions cannot be used to compute the uncertainty of spatial aggregates, for example when aggregating prediction maps of different depth layers or computing average values of a soil property for a specific land use or province. This requires quantifying cross- and spatial correlation in prediction errors, which can be accounted for by taking a multivariate or geostatistical approach (Szatmári et al., 2021; van der Westhuizen et al., 2022; Wadoux & Heuvelink, 2023).

5.3.4 Assessment scale

We recommend using BIS-4D maps on a national scale, as long as the map quality based on the provided accuracy metrics (Tables 5.7 & 5.8; SI) and prediction uncertainty (Figs. 5.4i-k & 5.5f; SI) is sufficient for the intended use. The model was developed for the national scale for multiple land uses. Foremost, BIS-4D maps contribute to the GSM project by delivering high-resolution, 3D (and 3D+T for SOM) maps of key soil properties with quantified uncertainty according to GSM specifications for the Netherlands. The BIS-4D maps may be especially useful for initiatives that require spatially explicit soil information across all land uses and soil types of the Netherlands. This may include national contributions to United Nations and pan-European directives and policies (Panagos et al., 2022b), such as the Green Deal, the Common Agricultural Policy, Zero Pollution, the EU Soil Strategy for 2030, the Soil Deal (European Commission, 2021) and the Proposal for a Directive on Soil Monitoring and Resilience (European Commission, 2023b). For example, clay, silt, sand and SOM maps can be used to improve estimates of soil-derived greenhouse gas emissions from the LULUCF sector for the Netherlands (Arets et al., 2020).

Many potential users of BIS-4D soil property prediction maps on a national scale may require information specifically for one land use and soil type. Perhaps most commonly, users may need information for agricultural soils. For example, maps of clay, silt, sand and SOM can provide information used to estimate the carbon sequestration potential for the "Smart Land Use" project (Slier et al., 2023), which is focused specifically on mineral soils under agricultural use on a national scale. As policy makers are mostly more interested more complex soil information, such as soil health, soil functions or soil-based ecosystem services, BIS-4D maps of several soil properties separately or combined can serve as inputs for a variety of tools that assess soil health or ecosystem services. For agricultural soils, these tools include OSI (Ros et al., 2022; Ros, 2023) and BLN 2.0 (Ros et al., 2023). Although pH, N_{tot}, P_{ox} and CEC can be used as approximations on

a national scale, pH, plant-available N and P and CEC are part of routine agronomic soil analyses. Therefore, maps of these soil properties may be more useful for forests and nature areas, where the base cation occupation of the CEC and pH should generally remain above a certain threshold to prevent Al-toxicity. For soil pH, BIS-4D predictions should be compared to predictions in Wamelink et al. (2019), who mapped soil pH in Dutch nature areas. As soil variability is linked to soil type (e.g. mineral vs. organic) and land use, we expect that BIS-4D model predictions would improve when modelling only one land use or soil type separately. However, this was not the scale of assessment aimed for with BIS-4D.

BIS-4D maps may also be used on a regional scale, as long as the accuracy allows and no better product is available. By regional scale we mean the level of provinces, regional water authorities, which are typically composed of one or more polders or watersheds, or large municipalities. These recommendations hold true especially for clay, sand and pH, which were predicted with higher accuracy than the other soil properties. However, regional management decisions come with social and economical risks. The costs of poor management decisions due to the use of inaccurate or not detailed enough soil information are often several magnitudes larger than investments for conducting a more detailed regional soil survey (Knotters & Vroon, 2015; Keller et al., 2018). In agreement with Chen et al. (2022), more research is necessary in relating DSM performance indicators such as uncertainty to cost-benefit and risk assessment analysis for improving decision support. We do not recommend the use of BIS-4D maps on a farm or field scale, as the uncertainty of predictions is most likely too high for the precision required by the farmer. Drivers of soil variation vary locally and were presumably not captured at this scale by the soil point data, covariates and model structure. As shown in Chapter 2, even for a soil property like soil pH, which was relatively easy to predict, less than 10 % of the map pixels were designated with one of the highest two GSM accuracy thresholds (AA and AAA). On such a local scale, we expect that the time and costs invested in a new soil survey outweigh the risks of using inaccurate soil data (Lemercier et al., 2022).

5.3.5 BIS-4D user manual

Based on the accuracy assessment of BIS-4D maps (Sect. 5.3.1), clay, sand and pH maps were most accurate. This is in agreement with Chen et al. (2022), who found that pH was the best predicted standard GSM soil property, followed by BD, PSFs (i.e. clay, silt and sand) and SOM, based on a review of 244 articles. BIS-4D map quality of silt, BD, SOM, N_{tot} and CEC were lower. For P_{ox}, we only recommend it as a baseline for an overview of the distribution of this soil property in the Netherlands.

Beyond these general recommendations, we have summarized the following simple chronological steps for users to help decide whether BIS-4D maps may be suitable for their intended purpose:

1. Choose one or multiple soil properties of interest.
2. Choose the depth layer(s) (0 - 5 cm, 5 - 15 cm, 15 - 30 cm, 30 - 60 cm, 60 - 100 cm and 100 - 200 cm) and, in the case of SOM, the year (1953, 1960, 1970, 1980, 1990, 2000, 2010, 2020, 2023), for which soil information is needed.
3. Consult the accuracy metrics of mean and median predictions of the soil property and depth layer of interest (Tables 5.7 & 5.8; SI), keeping in mind that the accuracy in the intended area of use may differ from the overall accuracy of the map. If the overall map quality based on these accuracy metrics is within an acceptable range for its purpose, continue to step 4. Use accuracy metrics based on design-based statistical inference for the soil properties for which it is available, i.e. BD, pH, SOM, N_{tot} and P_{ox} (Table 5.8; SI).
4. Choose whether to use mean or median prediction maps by comparing accuracy metrics of mean and median predictions (Tables 5.7 & 5.8; SI). Consult Sect. 5.3.1 for differences in mean and median predictions found using BIS-4D.
5. Download the mean and/or median prediction maps for the chosen soil property and depth layer as well as maps of the associated uncertainty (0.05 quantile, 0.95 quantile and/or PI90) and open them using GIS software. If soil information is required for a specific area, continue to step 6.
6. Prediction uncertainty is only useful for end-users if it is reliable (Schmidinger & Heuvelink, 2023). Therefore, check whether the prediction uncertainty is reliable by consulting the PICP. If the PICP90 is close to 0.90 (Table 5.7) and the PICP plot close to the 1:1 line (SI), then the provided prediction uncertainty map is reliable.
7. Ideally, prediction uncertainty should also be sharp, i.e. the PI90 should be as narrow as possible (Schmidinger & Heuvelink, 2023). Decide whether the PI90 is within an acceptable range for its purpose. If possible, fitness for use can be determined by analyzing how uncertainties in BIS-4D maps propagate through the intended usage, for example for an environmental model that uses BIS-4D maps as input. Commonly used uncertainty propagation methods include the Taylor Series or Monte Carlo methods (Heuvelink, 2018).

5.4 Data and code availability

The BIS-4D soil property prediction maps at 25 m resolution can be downloaded at `https://doi.org/10.4121/0c934ac6-2e95-4422-8360-d3a802766c71` (Helfenstein et al., 2024a). Prediction maps of the mean, median, 0.05 and 0.95 quantiles and the PI90 are available for each standard depth layer specified by GSM (0 - 5 cm, 5 - 15 cm, 15 - 30 cm, 30 - 60 cm, 60 - 100 cm and 100 - 200 cm). For SOM, maps at the same resolution

and for the same depth layers are available for the years 1953, 1960, 1970, 1980, 1990, 2000, 2010, 2020 and 2023.

Regarding BIS-4D model inputs, the soil point data of laboratory measurements and field estimates used during model calibration (PFB and BPK data) are publicly available at `https://doi.org/10.4121/c90215b3-bdc6-4633-b721-4c4a0259d6dc` (Helfenstein et al., 2024c). The georeferenced soil point data of PFB and BPK can also be viewed at `https://bodemdata.nl/bodemprofielen`. LSK and CCNL data used for design-based inference are not open due to privacy agreements. The pre-processed covariates that were openly available can be downloaded at 25 m resolution at `https://doi.org/10.4121/6af610ed-9006-4ac5-b399-4795c2ac01ec` (Helfenstein et al., 2024b). This includes the majority of the covariates used for BIS-4D, with the main exception being the covariates related to the national forestry inventory, since these data are closed. A public repository of the BIS-4D code is available here: `https://git.wageningenur.nl/helfe001/bis-4d`. The GitLab code repository is complete with the exception of BIS database credentials and the LSK and CCNL data. All data and code is available under the CC BY 4.0 license, except for the covariates (Helfenstein et al., 2024b), which are available under the CC BY-NC-SA 4.0 license.

Chapter 6

Synthesis

6.1 Introduction

In this thesis, I have developed BIS-4D, a high-resolution soil modelling and mapping platform for the Netherlands in 3D+T. In Chapter 2, I first directed attention towards a crucial part of DSM: how to best quantify the accuracy of a produced map, in this case for soil pH. Using established sampling theory, I developed an approach to statistically validate prediction maps at various depths using design-based inference based on a probability sample. In doing so, I provided unbiased accuracy metrics with associated confidence intervals, making the maps unique compared to other national-scale GSM products. Also in Chapter 2, I implemented GSM accuracy thresholds, which were devised in GSM specifications (Arrouays et al., 2015) but to my knowledge never used. I discussed the potentials of using GSM accuracy thresholds as ratings to communicate the quality of soil maps in a simple approach with tiers. Lastly, I developed a 3D DSM model that can predict at any user-defined depth.

In Chapter 3, I added the temporal dimension to BIS-4D, allowing for predictions of a dynamic soil property like SOM in 3D+T. Using machine learning in 3D+T, I predicted SOM for every 25 m pixel at any depth between 0 - 2 m for each year between 1953 - 2022. This allowed me to assess changes in SOM-related soil health in 3D+T, presenting a new paradigm compared to monitoring at the point scale or static soil maps at a single depth.

In Chapter 3, I stated that the 3D+T approach would also allow for predictions into the future, which is what I next investigated in Chapter 4. Using the same model, I predicted SOM in 2050 based on a nature-inclusive land use scenario. I found that nature-inclusive land use changes were mostly beneficial for SOM-related soil health, but that SOM decreased in areas overlooked in the nature-inclusive scenario. I discussed how using DSM in 3D+T for future scenarios might aid policymakers and spark discussions on sustainable transitions.

In Chapter 5, I demonstrated that BIS-4D can be used to fill the missing data gap of soil properties in the Netherlands by mapping nine different physical and chemical soil properties and their uncertainties: clay, silt and sand content, BD, SOM, pH, N_{tot}, P_{ox} and CEC at 25 m resolution between 0 - 2 m depth. By using standardised, highly automated and efficient workflows and making BIS-4D model inputs, code and outputs publicly available, I made BIS-4D more reproducible and easy to update.

In this last chapter, the synthesis, I will discuss whether the objectives of this thesis were achieved (Sect. 6.2), compare the findings and their implications with the literature and suggest future research directions and practical implementations (Sect. 6.3). Finally, I will relate this thesis within a broader, historical and contemporary context, add a personal reflection, and conclude (Sect. 6.4).

6.2 Overview and implications of findings

The aim of this thesis was to develop an operational, high-resolution, national-scale soil modelling and mapping platform that delivers predictions of soil properties and their uncertainties in 3D, and for dynamic soil properties, in 3D+T. By developing and applying this platform, BIS-4D, I expected to contribute to filling the gap of spatially explicit soil property information in the Netherlands, deliver a state-of-the-art product for GSM, while also addressing research challenges and answering pertinent questions (Sect. 1.6). In this section, I will discuss whether the objectives and research questions were achieved and adequately answered based on the findings from Chapters 2 - 5 and discuss their broader implications.

6.2.1 Assessing map accuracy

In Sect. 1.4.2, I outlined that there are several research gaps and challenges related to the best way to quantify map quality. These challenges relate to accuracy assessment methods that are either a) external, for example model-free, statistical validation, or b) internal, for example model-based, prediction uncertainty. For the soil pH maps in Chapter 2, I determined map accuracy by comparing various statistical validation strategies and by evaluating the prediction uncertainty. I compared the accuracy metrics derived from out-of-bag, location-grouped 10-fold cross-validation, an independent validation set and a stratified random sample of the independent validation set separated by depth layer. I obtained large differences depending on the four external accuracy assessment strategies and the depth layer. Design-based inference was most indicative of map accuracy based on sampling theory. Prediction uncertainty was slightly overestimated. Using the GSM accuracy thresholds, fewer than 10 % of pixels were designated with the highest ratings (AA and AAA) and therefore I recommended future studies to also test the feasibility of high quality thresholds for Tier 4 GSM maps.

Although design-based statistical inference is well-established in sampling theory (Cochran, 1977; de Gruijter et al., 2006; Gregoire & Valentine, 2007; Brus, 2022), Chapter 2 marked an additional step by applying the theory to soil layers at multiple depths, i.e. in 3D space (Research question 1a). To my knowledge, the few studies that used design-based inference with a probability sample elsewhere were at the regional scale (Malone et al., 2011; Lacoste et al., 2014). More importantly, these studies ignored the fact that the criteria of a probability sample were only met in 2D, but not 3D space. Hence, the approach of using design-based inference in 3D space developed in Chapter 2 bridged the gap between established sampling theory and much needed 3D DSM models. The soil pH maps for the Netherlands serve as a best-practice approach for other national-scale GSM products, in that they provide unbiased accuracy metrics including confidence intervals for maps in 3D space. In addition, although Brus et al. (2011) explained that using probability sampling for statistical validation of soil maps provides unbiased estimates

of map accuracy, comparing accuracy metrics of the same map based on design-based and various non-design based methods was rarely done (Research question 1a). Besides Chapter 2, one recent exemption is Wadoux et al. (2021a), which used a numerical experiment to show that differences in design-based and non-design-based, commonly used cross-validation approaches lead to substantially different conclusions about map quality. Finally, Chapter 2 identified accuracy thresholds delineated in the GSM specifications proposing tiered DSM products (Arrouays et al., 2015) as an additional powerful tool of communicating map quality (Research question 1b). The implemented spatially explicit ratings (none, A, AA and AAA) have the advantage that they can serve as a simple tool for communicating uncertainty with end-users and provide a standardized and easy way to directly compare DSM products across scales and study regions. In summary, I successfully fulfilled objective 1 to develop a high-resolution, national-scale DSM model of soil pH with improved ways of assessing map quality (Sect. 1.6).

The findings of Chapter 2 imply that a probability sample serves as the foundation for spatial soil information systems, in that they provide the basis for properly quantifying map accuracy. This has also been recognized in other studies. For example, van Doorn et al. (2024) recently assessed maps of amorphous iron- and aluminium-(hydr)oxides at various depths in agricultural soils in the Netherlands using the approach developed in Chapter 2. Findings of map quality obtained using design-based compared to other statistical inference methods need to be investigated in other study regions as well. However, based on established theory (Cochran, 1977; de Gruijter et al., 2006; Gregoire & Valentine, 2007; Brus et al., 2011; Brus, 2022), the results I found are likely study area independent. Thus, investing in a probability sample so as to assess map accuracy using design-based statistical inference is always advisable, also in other regions of the world. Considering the potential risks involved of subsequent map applications when accuracy is poorly estimated, investments in a probability sample for soil map validation are an added value and worthwhile, especially in the long-term. However, soil sampling is no doubt a costly and time-consuming affair and many areas do not have probability samples available. If resources are not available, it is advisable to use weighted cross-validation (de Bruin et al., 2022), not spatial cross-validation (Wadoux et al., 2021a), to assess mapping accuracy. If the sampling locations used for statistical validation are not spatially clustered, regular (random) cross-validation can also provide relatively unbiased metrics of map accuracy (de Bruin et al., 2022).

Also in Chapter 2, I addressed the question of how to quantify and visualize prediction uncertainty in a simple, standardized and appealing way by using GSM accuracy thresholds (Research question 1b). I found that GSM accuracy thresholds are a powerful and simple way to differentiate map quality, but that the GSM specifications possibly need to be revised. Even in the Netherlands, a study region with a large amount of high-quality data, and for soil pH, which had the highest prediction accuracy out of the soil properties mapped in Chapter 5, fewer than 10 % of the map pixels were designated with one of

the highest two GSM accuracy thresholds (AA and AAA). Lilburne et al. (2023) recently applied the GSM accurracy thresholds to clay, silt and sand predictions from SoilGrids (Poggio et al., 2021) in the Netherlands and New Zealand. Out of the three predicted soil properties at the six standard GSM depth layers in both countries, only predictions of sand from 5 - 15 cm depth met even the lowest accuracy threshold (A), and only for fewer than 20 % of the Netherlands. Recently, Nussbaum et al. (2023) found that approximately 50 % of predictions fall within one of the three accuracy thresholds (A, AA, AAA) when using a hierarchical modelling approach to predict soil pH in Swiss forests on a national scale. This was a large improvement compared to approximately only 10 % falling within the rating without the hierarchical modelling approach. The findings of Nussbaum et al. (2023) suggest that hierarchical predictions better reflected the conditional distribution of the response and prediction intervals were narrower. However, Nussbaum et al. (2023) did not assess whether the uncertainty was still reliable, for example using PICP or other uncertainty evaluation methods (Schmidinger & Heuvelink, 2023). Thus, prediction uncertainty using a hierarchical model may have decreased, but without an indication of the reliability, it unknown whether the "true" uncertainty of the map was also lower. The simplicity and ease of comparing rating achievements of A, AA and AAA from these three studies alone (Chapter 2; Lilburne et al., 2023; Nussbaum et al., 2023) highlights their advantage for end-users and comparing map quality across scales and study areas. Easy comparison of DSM products is much needed as an increasing number of maps become available at different scales (Mulder et al., 2016b; Lemercier et al., 2022).

Probabilistic models are paramount for providing spatially explicit uncertainty of soil maps, but assessment methods and metrics of varying degrees of complexity are necessary to directly compare, summarize and easily communicate prediction uncertainty of DSM products (Fig. 6.1). In this thesis, I used QRF due to its ability to predict conditional probabilities, used the PI90 as a measure of prediction uncertainty and the PICP to assess the reliability of the prediction uncertainty. However, the PICP used in this thesis and other DSM studies is ignorant of potential one-sided bias of its boundaries (Schmidinger & Heuvelink, 2023), which is for example apparent in soil texture maps (Lilburne et al., 2023). Therefore, improved ways to evaluate probabilistic predictions include the quantile coverage probability (QCP) and the probability integral transform (PIT) histogram (Schmidinger & Heuvelink, 2023). In addition, proper scoring rules for relative comparisons of competing probabilistic models include the Kolmogorov-Smirnov test statistic D and Anderson-Darling two sample test used by Nussbaum et al. (2023), or the interval score (IS) and continuous ranked probability score (CRPS), which can be decomposed into a reliability part (RELI)(Fig. 6.1; Schmidinger & Heuvelink, 2023). Although these evaluation methods offer improved ways of evaluating the magnitude and reliability of probabilistic predictions, I believe simple ratings, such as the GSM accuracy thresholds used in Chapter 2 are also needed to convey uncertainty in a simplified manner.

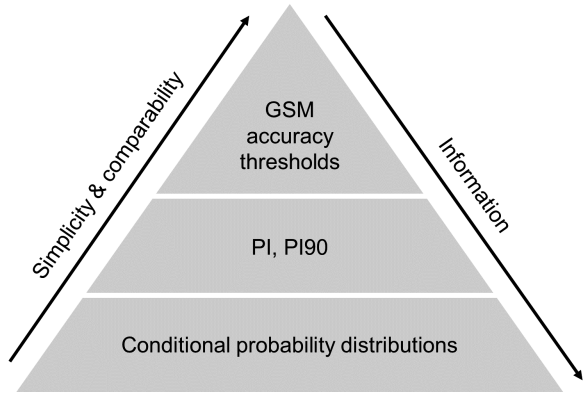


Figure 6.1: Probabilistic predictions expressed as conditional probability distributions are the basis of providing spatially explicit uncertainty. However, although metrics such as prediction intervals (PI), the 90th prediction interval width (PI90), and particularly the accuracy thresholds of GlobalSoilMap, reduce the amount of information related to prediction uncertainty, they have the advantage that they are simpler and easier to compare.

6.2.2 Mapping in 3D

Besides improved ways of assessing map quality, the first objective in Sect. 1.6, also addressed in Chapter 2, was to develop a DSM model for predicting soil pH in 3D space. This objective was achieved by using depth as a covariate. Compared to 2.D approaches, this had the advantage that sampling depth information was an important covariate and improved model performance. Furthermore, only one model needed to be calibrated to predict at any depth. Despite these practical and computational advantages, in the literature there is no general consensus on whether 2.5D models or 3D models are better and the choice between approaches is probably case-specific, depending on the soil property, its vertical variation and the sampling density (Nauman & Duniway, 2019; Roudier et al., 2020; Ma et al., 2021; Chen et al., 2022). However, according to the extensive review of Chen et al. (2022), the accuracy of 3D models for most soil properties was greater than that of the 2.5D models, except for bulk density. Moreover, compared to other 3D methods (Poggio & Gimona, 2014; Orton et al., 2016, 2020; Hengl et al., 2014, 2015; Meersmans et al., 2009a,b; Kempen et al., 2011; Liu et al., 2016; Ottoy et al., 2017; Rentschler et al., 2019), it is easier to implement, requires less steps, does not require tedious derivation of parameter depth functions and uses the predictive performance of machine learning. In summary, including depth as a covariate brings with it more practical advantages than 2.5D and other 3D methods, and perhaps leads to slightly more accurate maps.

Nonetheless, using depth as a covariate has become relatively common (e.g. Akpa et al., 2014; Filippi et al., 2019, 2020; Hengl et al., 2017b; Ramcharan et al., 2018; Zhang et al., 2020). In Chapter 2, one limitation was that the covariates used to predict soil pH

themselves were only variable in 2D space. The challenge of deriving covariates variable in more than 2D was next investigated in Chapter 3.

6.2.3 Mapping in 3D+T

In Sect. 1.4.4, I outlined that space-time DSM is a major challenge, especially in 3D+T, and that finding solutions to this challenge is important because many soil properties are highly variable in 3D+T. In Chapter 3, I developed a statistical modelling platform informed by 869 094 SOM observations from 339 231 point locations and the novel use of environmental covariates variable in 2D+T and 3D+T. The DSM model predicted SOM and its uncertainty annually at 25 m resolution between 0 - 2 m depth from 1953 - 2022 in the Netherlands. SOM decreased by more than 25 % in peatlands and on average by 0.1 - 0.3 % in cropland mineral soils. In addition, the model predicted increases between 10 - 25 % on reclaimed land due to land subsidence and vertical shifting of peat layers. The DSM model explicit in 3D+T quantified the substantial variations of SOM in space, depth, and time, highlighting the inadequacy of evaluating SOM dynamics at point scale or static mapping at a single depth for policymaking.

Chapter 3 successfully fulfilled Objective 2 of this thesis, to develop a high-resolution, national-scale DSM model of SOM explicit in 3D+T (Sect. 1.6). I was able to derive covariates variable in 2D+T, namely land use and peat classes, and in 3D+T, namely peat occurrence, as drivers of spatio-temporal SOM dynamics (Research question 2a). I believe Chapter 3 marks a pivotal next step in the field of predictive, spatial modelling by extending the advantages of machine learning to 3D+T. SOM enhances the availability of plant nutrients, improves moisture retention, stabilizes soil structure, increases permeability and chemical buffering and influences the biodegradability of pollutants. Hence, while a comprehensive evaluation of soil health requires consideration of various soil properties, amongst basic soil properties commonly mapped using DSM, SOM stands out as a fundamental indicator of soil health. Moreover, as SOM is highly variable in 3D space and time, the developed 3D+T model marks a new paradigm for not only mapping, but also monitoring soil health (Research question 2b). Even though successfully addressing Objective 2 of this thesis and introducing a new methodology to DSM, the 3D+T modelling approach also has several limitations.

By mapping SOM in 3D+T, it became apparent that an important limitation of the modelling approach used in BIS-4D is that it does not quantify the uncertainty of differences in SOM (ΔSOM), relevant for Chapters 3 and 4, and of spatial aggregates, relevant in all chapters throughout the thesis. This was discussed in detail in Sect. 3.2.4, but to summarize, these uncertainties can only be obtained if cross- and spatial correlation in prediction errors are quantified. These correlations can be accounted for by taking a multivariate or geostatistical approach (e.g. Gasch et al., 2015; Szatmári et al., 2021; van der Westhuizen et al., 2022; Wadoux & Heuvelink, 2023). Given the challenges and complexities such

analysis would involve (Sect. 3.2.4), this was no longer within the scope of this thesis, but should be investigated in future research. Investigating how to quantify uncertainty of SOM changes and spatial aggregates in 3D+T at scales relevant to management and policy is a critical task, since the uncertainty related to soil monitoring has prompted widespread doubts about the feasibility of measuring and verifying SOM and soil organic carbon changes (Moinet et al., 2023; Paul et al., 2023). The range of the estimated trend should be compared with the uncertainty to determine if trends are plausible or at the same order of magnitude as the cumulative errors linked to the spatial predictions at different dates (Chen et al., 2022). Thus, I advocate that future research should investigate the uncertainty quantification of SOM changes and SOM spatial aggregates.

Chapter 3 showed that predicting changes over time at point support is still a major challenge in DSM. This was also the case in other space-time DSM approaches using random forest (Heuvelink et al., 2020). Although using random forest with large amounts of point observations and covariates has repeatedly shown to outperform other mapping approaches (e.g. Hengl et al., 2015, 2017b; Nussbaum et al., 2018; Keskin et al., 2019; Khaledian & Miller, 2020), mechanistic models may capture temporal dynamics better. However, the practicality and performance of mechanistic models such as CENTURY (Parton et al., 1987), RothC (Coleman & Jenkinson, 1996) and Millenial (Abramoff et al., 2018, 2022) for spatial prediction tasks is poor. The respective limitations of both machine learning and process-based models highlight the challenge of accurately predicting in space and time, especially in 3D+T. However, a promising development is the combination of process-based with machine-learning models for space-time predictions (e.g. Zhang et al., 2024), or scientific machine learning (SciML; Iwema, 2023). While combining process-based and machine-learning models may improve predictions in cases for which it can be applied, one of the limitations is that it is considerably less flexible than the 3D+T approach used in Chapter 3. For example, for the calibration and statistical validation of process-based combined with ML models, repeated measurements at point locations are necessary. These were not available at point support for the Netherlands (Knotters et al., 2022, Sect. 2.1 and Appendix C) and are likely also not available in many other regions of the world. In addition, most process-based models are applicable for mineral soils under agricultural land use. This makes them difficult to apply on a national scale for all land uses, and in particular for the Netherlands, since 15% of the soils are organic soils.

Another finding of the 3D+T model in Chapter 3 is that there was very limited data for statistical validation of SOM temporal changes (ΔSOM), implying that investments are needed for soil monitoring. Furthermore, for many applications such as BIS-4D, monitoring is only useful if conducted using identical sampling techniques that minimize errors at point support. Even though the LSK probability sample locations were revisited in 2018 (CCNL), substantial methodological differences and positional uncertainty in the LSK and CCNL datasets prevented a temporal assessment at point scale (Sect. 3.3.1; Cavero Panez, 2021; Knotters et al., 2022, Sect. 2.1 and Appendix C). In light of the major advantages

of using a probability sample for design-based statistical inference (Chapter 2), the accuracy of 3D+T DSM models is ideally assessed using a a supplemented panel probability sample (Brus, 2022, Sect. 15.1). In other words, repeated sampling of the same 3D locations determined using a probability sample will allow design-based inference in four dimensions, i.e. space, depth and time. The benefits of monitoring these locations using identical procedures outweigh the costs in the long-term and is necessary for assessing soil health under the new proposed soil monitoring directive in the EU (European Commission, 2023b).

6.2.4 3D+T mapping for scenario modelling

In Sect. 1.4.4, I outlined that, to my knowledge, space-time DSM has never before been applied for future nature-inclusive land use scenarios and that combining these approaches may be a powerful way to link soil-related ecosystem services with sustainable transitions. In Chapter 4, I investigated the interplay between nature-inclusive land use scenarios and SOM using a DSM model in 3D+T. SOM was predicted in 2050 based on a recently developed nature-inclusive scenario (Breman et al., 2022) and machine learning in 3D+T on a national scale in the Netherlands. In contrast to Chapter 3, 2D+T and 3D+T covariates of land use, peat classes and peat occurrence were derived for 2050 based on nature-inclusive spatial planning decisions rather than historic land use and peat class maps. I predicted SOM and its uncertainty in 2050 and assessed SOM changes between 2022 and 2050 from 0 - 2 m depth at 25 m resolution.

In Chapter 4, I addressed Objective 3 of this thesis by applying the 3D+T SOM model for a future scenario. In doing so, Chapter 4 further demonstrated the flexibility of BIS-4D, since, given covariates are available, predictions can also be made in the future. Space-time modelling has rarely been applied for future scenarios (Meersmans et al., 2016; Yigini & Panagos, 2016; Gray & Bishop, 2016, 2019; Reyes Rojas et al., 2018; Adhikari et al., 2019). Furthermore, none of these studies used an approach explicit in 3D+T or scenarios based on nature-based solutions. Findings indicated that overall, nature-inclusive land use was conducive to enhancing SOM-related soil health (Research question 3a). For example, increases up to 25 % SOM were predicted between 0 - 40 cm depth in rewetted peatlands. However, the model also predicted SOM decreases in several areas throughout the Netherlands. For instance, SOM decreases up to 5 % were predicted in grasslands used for animal-based production systems in 2022, which transitioned into croplands for plant-based production systems by 2050. In addition, substantial decreases surpassing 25 % SOM were predicted on non-rewetted land in peat layers below 40 cm depth on reclaimed land. However, these decreases in SOM can mostly be attributed to the model's limitation to predict future trends based on historic data and to capture nuanced nature-inclusive practices, such as the adoption of agroecological farming methods. Nonetheless, there were several added values of using DSM for modelling SOM based on this future scenario (Research question 3b). For example, the findings from Chapter 4 pinpointed areas where

the nature-inclusive scenario overlooked potential threats to soil health in 3D space and at high resolution, such as deep peat layers. Furthermore, the findings also highlighted that, supported by the literature (Crews & Rumsey, 2017), achieving SOM levels akin to those in permanent grasslands within croplands presents a formidable challenge even with the adoption of agroecological practices. In summary, 3D-mapping of SOM in 2050 created new insights, raised important questions related to soil health behind nature-inclusive scenarios and may facilitate societal discussion, aid policy-making and promote transformative change.

Besides the difficulty of including nuanced nature-inclusive practices, another limitation, in both Chapter 3 and 4, was that climate was a not a dynamic covariate. Empirical evidence shows that climate change affects carbon dynamics in the soil (Beillouin et al., 2022, 2023). While long-term minimum, maximum and average temperature and precipitation between 1981 and 2010 were included as static covariates in the model (Chapter 2, Table 2.2), the temporal dynamics inherent in climate covariates were not incorporated during model calibration and prediction between 1953 - 2022, or during prediction for the 2050 scenario. While other DSM studies modelling future scenarios accounted for climate change (Gray & Bishop, 2016, 2019; Yigini & Panagos, 2016), I argued that, based on projected national climate scenarios for the Netherlands (KNMI, 2023; van Dorland et al., 2023), the impacts of temperature and precipitation on SOM dynamics were of lesser consequence compared to changes in land use, peat class, and peat occurrence. Nonetheless, I recommend space-time DSM studies to include dynamic changes in covariates related to the climate, for example through the combination with process-based climate models.

6.2.5 An operational platform for high-resolution soil property mapping

In Sect. 1.4.1 and 1.4.5, I described that there is a large gap and need for high-resolution maps of a broad range of soil properties, not only soil pH (Chapter 2) and SOM (Chapters 3 and 4), further highlighting the many advantages of creating an operational and efficient national-scale GSM product. Therefore, I next addressed the fourth and final objective of this thesis: to map a broad range of soil properties and their uncertainties using an operational, reproducible, standardised, largely automated and efficient workflow. In Chapter 5, I predicted clay, silt and sand content, BD, N_{tot}, P_{ox}, CEC, and updated versions of soil pH and SOM, and their uncertainties at 25 m resolution between 0 - 2 m depth. The accuracy of clay, sand and pH maps was highest, while silt, BD, SOM, N_{tot}, CEC and especially P_{ox} were more difficult to predict.

Based on the findings in Chapter 5, I further assessed the strengths and limitations of the BIS-4D prediction maps (Research question 4a). The main strengths of BIS-4D were: 1) the ability to provide information of soil properties as opposed to soil types; 2) the high spatial resolution (25 m); 3) accuracy assessment based on best practices; 4) the benefits

of using machine learning combined with large amounts of data; and 5) the flexibility to predict in 3D and, for SOM, in 3D+T. The main limitations of BIS-4D prediction maps were linked to uncertainty in the soil point data, covariates and model structure. With regards to the soil point data, the main limitations were that the calibration data of BD, P_{ox} and CEC were spatially clustered and that clay, silt, sand and CEC maps were not evaluated using design-based inference because they were not measured as part of a probability sample (LSK and CCNL). Despite having a wide range of high-quality covariates available in the Netherlands, the main limitations related to the covariates used in BIS-4D was limited information on detailed land management and that none of the covariates explained the spatial variation of P_{ox} well. Although it can be explored if slight improvements are achieved if using deep learning or transfer learning, the main limitation of the BIS-4D modelling structure was that probabilistic predictions cannot be used to compute the uncertainty of spatial aggregates and temporal differences (Sect. 3.2.4 and 6.2.3).

One main focus of Chapter 5 was the potential applications of BIS-4D, which was also a key part of answering Research questions 4a. I recommended the use BIS-4D maps on a national scale, as long as the quantified and provided map uncertainty was sufficient. I further outlined applications for a wide range of national contributions to international and EU policies such as the Green Deal and the Proposal for a Directive on Soil Monitoring and Resilience (European Commission, 2023b). In fact, the BIS-4D soil texture and SOM maps are currently already being used as inputs to improve the estimates of soil-derived greenhouse gas emissions for the Land Use, Land Use Change and Forestry (LULUCF) sector for the United Nations Framework Convention on Climate Change and the Dutch LULUCF submission under the Kyoto Protocol (Arets et al., 2020). However, besides modellers, many end-users are more interested in information derived from basic soil properties, such as soil health, soil functions or soil-based ecosystem services. I believe BIS-4D has made a substantial contribution for acquiring more complex information and knowledge about soil spatial variation in the Netherlands and similar regions of the world. Their application in tools such as the Open Soil Index (OSI; Ros et al., 2022; Ros, 2023) and Soil Indicators for Agriculture (BLN 2.0; Ros et al., 2023) to improve soil health is one example. Finally, I warned potential users against using BIS-4D when the uncertainty is too high for their intended use and discouraged from using the maps on a field scale.

The findings from Chapter 5 revealed that BIS-4D contributed to open science and provided a framework for making DSM models reproducible and easy to update (Research question 4b). One of the main advantages of DSM compared to conventional soil mapping is that mapping is more automated, less laborious, more reproducible and easier to update. Yet exploiting these advantages demands the scientific community to direct as much attention to making workflows findable, accessible, interoperable and reusable as to the results themselves. In the form of a data description manuscript (Chapter 5), I included a step-by-step user manual (Sect 5.3.5) and made model code and input data

openly available in publicly accessible and permanent repositories (Sect 5.4). Only a few parameters, such as the desired target soil property, prediction depth and year, need to be specified once in the beginning of a BIS-4D script, after which all computations run automatically. Moreover, BIS-4D was developed in close collaboration with soil surveyors and scientists, programmers and database maintainers from Wageningen Environmental Research, who can use BIS-4D in the future. BIS-4D prediction maps are already in the process of being uploaded to the national online soil portal (bodemdata.nl), where they will reach a different audience of potential users than the manuscript (Chapter 5) and published dataset (Helfenstein et al., 2024a). BIS-4D may serve as an example of open science, serving as an example for scientific projects to contribute to Targets 17.6 and 17.8 of the Sustainable Development Goals, which aim to improve global cooperation and access to science, technology, and innovation through knowledge sharing.

6.3 Future research and practical implementation

6.3.1 Improving predictions in 3D

A cornerstone of improving the accuracy of soil maps is improving predictions over depth, which requires developing better 3D models. Based on the implications from Chapter 3, one area of future research is deriving covariates that allow machine learning to capture soil variability over depth. One way of achieving this with covariates already available today is using maps of soil classes, as they are in essence 3D models of several soil properties (Kempen et al., 2011). However, rather than enforcing parametric depth functions (e.g. Meersmans et al., 2009a,b; Kempen et al., 2011; Liu et al., 2016; Ottoy et al., 2017; Rentschler et al., 2019), I suggest to simply derive the necessary covariate information from soil maps so that the machine learning algorithm can define these depth functions based on the error loss function during model calibration. Besides the example of peat occurrence (Chapter 3), another example that may be useful for mapping SOM and closely related soil properties is the presence of pedogenetic soil horizons that constitute sharp discontinuities in the soil profile (Fig. 1.4). For example the “O” or “Ah”, “E” and “Bh” horizons in podzols or increasing clay content with increasing depth in vertisols (IUSS Working Group WRB, 2014). A starting point to determine what kind of covariates can be derived explicitly in 3D space may be found in the studies that defined parametric depth functions, e.g. Kempen et al. (2011) for the case of the Netherlands. If choosing such an approach, one disadvantage will be the bias introduced by using soil point data for calibration that were also used to develop the soil class map, as discussed in Chapters 2 and 3.

3D predictions may also be improved by applying random forest spatial interpolation (RFSI; Sekulić et al., 2020) for 3D DSM. Using synthetic data and meteorological datasets, Sekulić et al. (2020) included nearby observations and their distances as covariates in a

random forest model. Machine learning models do not account for spatial autocorrelation, and although other studies have included geographic space as covariates (Li et al., 2011; Behrens et al., 2018b; Hengl et al., 2018; Møller et al., 2020), by including the observations and their distances, RFSI uses the core idea behind deterministic geostatistical methods like kriging. Although initial tests of applying RFSI for DSM in 3D space for soil pH in the Netherlands did not improve model predictions (Helfenstein et al., 2022), I recommend further testing for different soil properties and case studies. This approach may be promising for 3D modelling because presumably, no covariate can better detect vertical variation and depth discontinuities than observations from nearby profiles themselves.

6.3.2 A new paradigm: scalable soil estimation services

Over the course of history, our knowledge of soil and its spatial variation has advanced: from the rudimentary soil map crafted some 4000 years ago, to Dokuchaev's description of soil-forming factors around 1900, and the meticulous work of soil surveyors for conventional soil mapping throughout the 20th century. However, the development and widespread adoption of DSM in the past three decades has catalyzed a profound transformation in pedology and soil mapping, propelling the research field forward at an astonishing pace. While I believe this thesis has contributed to the pedometrics and DSM research fields, this thesis has also made evident that major challenges still remain. These challenges, together with the ever increasing importance of spatio-temporal soil information, technological advances and a society in transition, will prompt a further paradigm shift in delivering soil information in 3D+T.

To meet the challenges of monitoring soil health in 3D+T, I believe DSM needs a paradigm shift: from maps to scalable soil estimation services. To date, most DSM studies produce maps for one target scale with a fixed spatial extent. For example, BIS-4D was specifically designed to provide users with soil information on a national level, whereas, in consideration of the prediction uncertainty and accuracy metrics, I did not recommend its use on a field scale (Sect. 5.3.4). I also discussed that different covariates may be relevant at different scales (Sect. 5.3.3). For example, covariates of detailed crop data or plowing data may not be useful for improving predictions on a national level across various land uses. On the other hand, a covariate representing geology will likely not be useful on a local level, as the substrate material of soil formation on a farm will likely exhibit little to no spatial variability. Another issue is that if other countries also develop similar platforms to BIS-4D, for example neighboring countries to the Netherlands, there will likely be national border artifacts since different models were used. This challenge applies to all "bottom-up" GSM approaches by mapping across the globe region by region. Therefore, to address these challenges, I think the potential of DSM to meet future user demands is in developing scalable modelling and mapping platforms that can tailor predictions on-the-fly to any target region.

One potential of achieving scalable soil estimation services is by combining DSM with soil spectroscopy models. Using soil spectroscopy models, large soil spectral libraries can be efficiently mined based on a few local samples, so that relevant information from the spectral library is employed for improving local predictions (Lobsey et al., 2017; Liu et al., 2018; Padarian et al., 2019a; Seidel et al., 2019; Helfenstein et al., 2021; Baumann et al., 2021). Future research should investigate whether a similar approach can be used in DSM using several steps. Firstly, a minimum number of representative samples from a local target region are collected in the field. Secondly, using soil spectroscopy, the spectral signals of newly collected local samples are compared to data in large soil spectral libraries via data mining algorithms. Since soil spectral libraries also contain reference wet-chemistry measurements of soil samples, the reference measurements determined as representative for the local study area based on their spectral signal are subsequently used in a DSM model. The requirements for such an approach are soil spectral libraries and covariates at different scales. However, soil spectral libraries are already available at regional (e.g. Wetterlind & Stenberg, 2010), national (e.g. Baumann et al., 2021) and global scale (Viscarra Rossel et al., 2016; Shepherd et al., 2022), and many covariates, such as digital elevation models and remote sensing data, are useful from the field to the global scale. By comparing soils from the local target region with soils in a soil spectral library, the idea is comparable to the concept of using "Homosoils", or locations sharing similar soil-forming factors (Mallavan et al., 2010), for DSM (e.g. Nenkam et al., 2022). However, whereas "Homosoil" is based around similarity of soil-forming factors, i.e. covariates, soil-spectroscopy allows a more promising approach by rapidly comparing the soil itself. Soil spectral signatures are indicative of a variety of biological, chemical and physical soil properties (Viscarra Rossel et al., 2006, 2008) and are thus a holistic fingerprint of multiple soil properties. One project already underway to monitor soils using soil spectral libraries and hyperspectral satellites is the "WORLDSOILS" project of the European Space Agency (Yagüe et al., 2023; Dvorakova et al., 2021, 2023; Francos et al., 2023).

However, scalable soil estimation services will bring about new challenges and new research opportunities. For example, instead of calibrating one DSM model for one specific scale, predicting soil properties for any user-defined scale will require flexible platforms in which models are constantly re-tuned, re-calibrated and predicted. Solutions for these technical challenges may already be available from other fields, for example in computer science and agro-environmental applications, such as robotics and precision farming. However, the main challenge will not be in the tuning, calibration or prediction steps, but in accurracy assessment, posing a new challenge for spatial sampling theory. For example, how can predictions from scalable soil estimation services be evaluated, e.g. using design-based inference, without one fixed sample population? These and other challenges related to this paradigm shift will hopefully lead to closer collaboration between soil surveyors, modellers (pedometricians), trans-disciplinary scientists and a broad range of stakeholders interested

in increasingly accurate soil information.

6.3.3 Modern soil surveys to meet modern demands

Most DSM research uses legacy soil data collected by national soil survey institutes between World War II and the 1990s (Sect. 1.1). Besides a field campaign in 2022 (SI3 of Chapter 3[1]), BIS-4D is no exception to that, using mostly data collected by the StiBoKa, the former national soil survey institute of the Netherlands. Recall from Sect. 1.1.2 that due to their key role, soil surveying was a relatively common profession in the field of environmental and agricultural sciences. In 1998, there were on average 23 soil surveyors per 1000 km^2 agricultural land in the Netherlands, although numbers were lower, ranging between three and six, in France, Denmark and the United Kingdom (van Baren et al., 2000). National soil surveys were used to develop conventional soil maps, but after completion, or in many countries also before completion, funding for national soil survey institutes was ceased. Ever since, soil surveys of such scale and magnitude have to my knowledge not been undertaken. In the Netherlands, one exception was the CCNL field campaign, but recall from Chapter 3 the numerous limitations and lack of soil monitoring data at point scale. Other recent soil surveys are relatively small, regional scale projects with limited budget, which also means that new data are usually field estimates rather than expensive, but more accurate laboratory measurements (e.g. Fig. 3.2b). Finally, there are numerous general limitations revolving around the use of soil legacy data. For example, soil legacy data exhibit uncertainty related to the age of the soil sample and their geographic position, which was determined before the use of GPS.

We cannot expect to solve modern-day and future challenges alone with legacy soil data, prompting the necessity of large-scale soil surveys using state-of-the-art techniques. The transition from DSM to scalable soil estimation services will be key (Sect. 6.3.2), but is only one part of the puzzle. One practical "bottleneck" is collecting soil samples and preparing them for subsequent analysis. For example, drying, milling and sieving soil samples is necessary for almost any kind of subsequent analysis, including for soil spectroscopy measurements, as the accuracy of on-site spectral measurements is limited (e.g. Greenberg et al., 2022a,b). However, while sensing and statistical modelling methods have rapidly advanced, soil preparation techniques have remained largely the same over the last few decades. An additional challenge is that measuring, reporting and verifying (MRV) changes in soil properties are not standardized between individual laboratories and nations (Smith et al., 2020). In summary, solving practical issues are pivotal for monitoring and mapping soils in 3D+T and the Proposal for a Directive on Soil Monitoring and Resilience all across Europe (European Commission, 2023b).

Lastly, soil surveying, mapping and monitoring is also a societal and political challenge

[1]Supplementary information (SI) of Chapter 3 is available at `https://doi.org/10.1038/s43247-024-01293-y` under "Supplementary information".

that ultimately requires investment. It is difficult to estimate the amount of public spending for obtaining soil information, but currently, only 6 % of environmental protection investments are in the domain soil and groundwater (Eurostat, 2023). However, soil is directly linked to a wide array of ecosystem services and human well-being (Fig. 6.2). Therefore, although requiring substantial investment, detailed soil surveys that can provide accurate soil information can save costs in a multitude of sectors, such as for drinking water preparation, optimizing irrigation and agricultural inputs, infrastructure planning and mitigating pollution. In Switzerland, conservative estimates based on an in-depth analysis showed that every Swiss franc invested in acquiring soil information generates an added value of 2 - 13 Swiss francs (Keller et al., 2018, Sect. 4.2). Hence, future scientific research is important, but the potential of using accurate soil information for human well-being also requires society at broad to recognize the importance of soils.

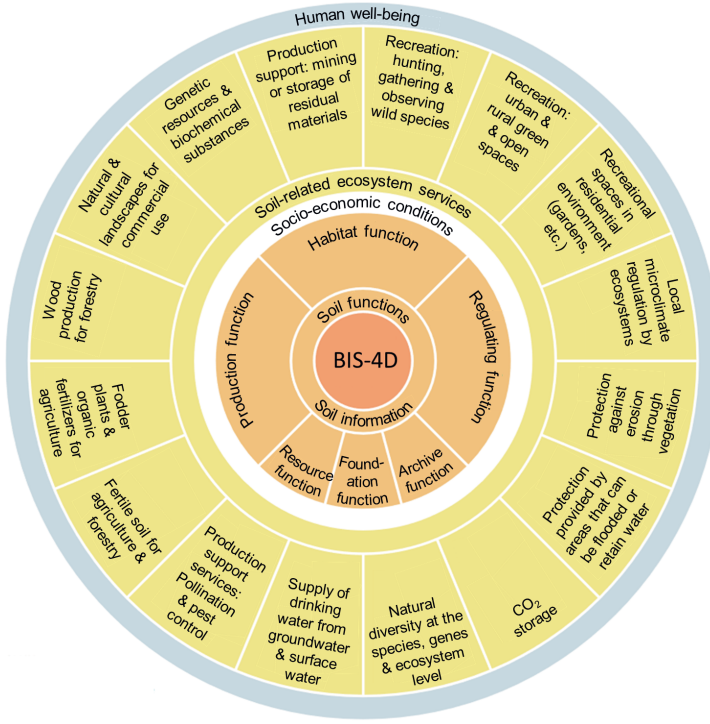


Figure 6.2: The larger context: soil properties predicted using BIS-4D contribute to the ability of soil resources to provide important functions and services for society, ecosystems and human well-being (adjusted from Keller et al., 2018, Fig. 2).

6.4 Conclusion

In this thesis, I have successfully developed BIS-4D, a high-resolution soil modelling and mapping platform for the Netherlands in 3D+T. Spatial information of soil properties obtained using BIS-4D provides the basis for deriving more complex soil information, such as soil functions, and links to soil-related ecosystem services (Fig. 6.2). Therefore, by developing BIS-4D, I have not only addressed several scientific challenges, but also produced new information, which will hopefully lead to better knowledge of the world around us and wisdom in the choices that we make.

As the skin of the earth, soil acts as a substrate for the cycling of water, gases and nutrients, all of which we depend on for a healthy living environment. Humans are altering the environment at an unprecedented rate, but soil is a precious resource that is non-renewable within human lifespans and terrestrial life without it is unimaginable. As any other resource, sustainable use of soils requires that we quantify and understand its spatial variation across the land surface, but also over depth and time. The well-being of humanity, echoing the needs of early civilizations, hinges upon the health and preservation of this fundamental resource.

References

Aarts, N., & Leeuwis, C. (2023). The Politics of Changing the Dutch Agri-Food System. *Journal of Political Sociology*, *1*. doi: 10.54195/jps.14922. Number: 1.

Abramoff, R., Xu, X., Hartman, M., O'Brien, S., Feng, W., Davidson, E., Finzi, A., Moorhead, D., Schimel, J., Torn, M., & Mayes, M. A. (2018). The Millennial model: in search of measurable pools and transformations for modeling soil carbon in the new century. *Biogeochemistry*, *137*, 51–71. doi: 10.1007/s10533-017-0409-7.

Abramoff, R. Z., Guenet, B., Zhang, H., Georgiou, K., Xu, X., Viscarra Rossel, R. A., Yuan, W., & Ciais, P. (2022). Improved global-scale predictions of soil carbon stocks with Millennial Version 2. *Soil Biology and Biochemistry*, *164*, 108466. doi: 10.1016/j.soilbio.2021.108466.

Adhikari, K., Hartemink, A. E., Minasny, B., Kheir, R. B., Greve, M. B., & Greve, M. H. (2014a). Digital Mapping of Soil Organic Carbon Contents and Stocks in Denmark. *PLOS ONE*, *9*, e105519. doi: 10.1371/journal.pone.0105519.

Adhikari, K., Kheir, R. B., Greve, M. B., Bøcher, P. K., Malone, B. P., Minasny, B., McBratney, A. B., & Greve, M. H. (2013). High-Resolution 3-D Mapping of Soil Texture in Denmark. *Soil Science Society of America Journal*, *77*, 860–876. doi: 10.2136/sssaj2012.0275. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2136/sssaj2012.0275.

Adhikari, K., Kheir, R. B., Greve, M. B., Greve, M. H., Malone, B. P., Minasny, B., & McBratney, A. B. (2014b). Mapping soil pH and bulk density at multiple soil depths in Denmark. In *GlobalSoilMap: Basis of the global spatial soil information system* (p. 6). Boca Raton: CRC Press Taylor & Francis Group. Google-Books-ID: S5ClAgAAQBAJ.

Adhikari, K., Owens, P. R., Libohova, Z., Miller, D. M., Wills, S. A., & Nemecek, J. (2019). Assessing soil organic carbon stock of Wisconsin, USA and its fate under future land use and climate change. *Science of The Total Environment*, *667*, 833–845. doi: 10.1016/j.scitotenv.2019.02.420.

AHN (2023). Actueel Hoogtebestand Nederland (AHN). Publisher: AHN.

Aitchison, J. (1986). *The statistical analysis of compositional data*. London: Chapman and Hall.

van den Akker, J., Kuikman, P., de Vries, F., Hoving, I., Pleijter, M., Hendriks, R., Wolleswinkel, R., Simões, R., & Kwakernaak, C. (2008). Emission of CO2 from agricultural peat soils in the Netherlands and ways to limit this emission. *Proceedings of the 13th International Peat Congress After Wise Use - The Future of Peatlands*, *1*, 645–648.

van den Akker, J. J. H., Hendriks, R., Pleijter, M., Wolleswinkel, R. J., Beuving, J., Hoving, I. E., & Houwelingen, K. v. (2007). Empirical relations between subsidence, CO2 emissions and water management. In *Proceedings of the First International Symposium on Carbon in Peatlands, Wageningen, The Netherlands, 15 - 18 April 2007* (pp. 102–102). Wageningen University.

van den Akker, J. J. H., & Hoogland, T. (2011). Comparison of risk assessment methods to determine the subsoil compaction risk of agricultural soils in The Netherlands. *Soil and Tillage Research*, *114*, 146–154. doi: 10.1016/j.still.2011.04.002.

van den Akker, J. J. H., de Vries, F., Vermeulen, G., Hack-ten Broeke, M., & Schouten, T. (2012). *Risico op ondergrondverdichting in het landelijk gebied in kaart*. Alterra-rapport 2409, Alterra Wageningen, the Netherlands.

Akpa, S. I. C., Odeh, I. O. A., Bishop, T. F. A., & Hartemink, A. E. (2014). Digital Mapping of Soil Particle-Size Fractions for Nigeria. *Soil Science Society of America Journal*, *78*, 1953–1966. doi: 10.2136/sssaj2014.05.0202. _eprint: https://acsess.onlinelibrary.wiley.com/doi/pdf/10.2136/sssaj2014.05.0202.

Alterra (2004). Historisch Grondgebruik Nederland (HGN).

Amelung, W. et al. (2020). Towards a global-scale soil climate mitigation strategy. *Nature Communica-*

tions, *11*, 5427. doi: 10.1038/s41467-020-18887-7. Number: 1 Publisher: Nature Publishing Group.

Amirian-Chakan, A., Minasny, B., Taghizadeh-Mehrjardi, R., Akbarifazli, R., Darvishpasand, Z., & Khordehbin, S. (2019). Some practical aspects of predicting texture data in digital soil mapping. *Soil and Tillage Research*, *194*, 104289. doi: 10.1016/j.still.2019.06.006.

Amundson, R., Berhe, A. A., Hopmans, J. W., Olson, C., Sztein, A. E., & Sparks, D. L. (2015). Soil and human security in the 21st century. *Science*, *348*, 1261071. doi: 10.1126/science.1261071. Publisher: American Association for the Advancement of Science.

Appelhans, T., Detsch, F., Reudenbach, C., & Woellauer, S. (2023). *mapview: Interactive viewing of spatial data in R*. Technical Report, . `https://github.com/r-spatial/mapview`.

Arets, E. J. M. M., Kolk, J. W. H. v. d., Hengeveld, G. M., Lesschen, J. P., Kramer, H., Kuikman, P. J., & Schelhaas, N. J. (2020). *Greenhouse gas reporting for the LULUCF sector in the Netherlands: methodological background, update 2020*. WOt-technical report 168, Statutory Research Tasks Unit for Nature & the Environment (WOT Natuur & Milieu) Wageningen, the Netherlands. `https://research.wur.nl/en/publications/greenhouse-gas-reporting-for-the-lulucf-sector-in-the-netherlands-5` publisher: WOT Natuur & Milieu DOI 10.18174/517340.

Arrouays, D. et al. (2014a). Chapter Three - GlobalSoilMap: Toward a Fine-Resolution Global Grid of Soil Properties. In D. L. Sparks (Ed.), *Advances in Agronomy* (pp. 93–134). Academic Press volume 125. doi: 10.1016/B978-0-12-800137-0.00003-0.

Arrouays, D. et al. (2017). Soil legacy data rescue via GlobalSoilMap and other international and national initiatives. *GeoResJ*, *14*, 1–19. doi: 10.1016/j.grj.2017.06.001.

Arrouays, D., McBratney, A., Bouma, J., Libohova, Z., Richer-de Forges, A. C., Morgan, C. L. S., Roudier, P., Poggio, L., & Mulder, V. L. (2020). Impressions of digital soil maps: The good, the not so good, and making them ever better. *Geoderma Regional*, (p. e00255). doi: 10.1016/j.geodrs.2020.e00255.

Arrouays, D., McBratney, A., Minasny, B., Hempel, J., Heuvelink, G. B. M., MacMillan, R. A., Hartemink, A., Lagacherie, P., & McKenzie, N. (2015). The GlobalSoilMap project specifications. In *Proceedings of the 1st GlobalSoilMap Conference* (pp. 9–12). doi: 10.1201/b16500-4.

Arrouays, D., McKenzie, N., Hempel, J., Forges, A. R. d., & McBratney, A. (2014b). Preface. In *GlobalSoilMap: Basis of the global spatial soil information system* (p. xiii). Boca Raton: CRC Press Taylor & Francis Group.

Arrouays, D., Mulder, V. L., & Richer-de Forges, A. C. (2021). Soil mapping, digital soil mapping and soil monitoring over large areas and the dimensions of soil security – A review. *Soil Security*, *5*, 100018. doi: 10.1016/j.soisec.2021.100018.

van Asselen, S., Erkens, G., Stouthamer, E., Woolderink, H. A. G., Geeraert, R. E. E., & Hefting, M. M. (2018). The relative contribution of peat compaction and oxidation to subsidence in built-up areas in the Rhine-Meuse delta, The Netherlands. *Science of The Total Environment*, *636*, 177–191. doi: 10.1016/j.scitotenv.2018.04.141.

Baake, K. (2018). *Quantifying Uncertainty of Random Forest Predictions: A Digital Soil Mapping Case Study*. Thesis Report GIRS-2017-14 Wageningen University Wageningen, the Netherlands.

de Bakker, H., & Schelling, J. (1966). *Systeem van bodemclassificatie voor Nederland: De hogere niveaus*. (1st ed.). Wageningen, the Netherlands: Centrum voor Landbouwpublikaties en Landbouwdocumentatie. First version.

de Bakker, H., & Schelling, J. (1989). *Systeem van bodemclassificatie voor Nederland: de hogere niveaus: With Engl. summary: A system of soil classification for the Netherlands*. (Second, revised edition ed.). Wageningen, the Netherlands: Centrum voor Landbouwpublikaties en Landbouwdocumentatie.

Bakker, J., Dessel, B. v., & Zadelhoff, F. v. (1989). *Natuurwaardenkaart 1988 : natuurgebieden, bossen en natte gronden in Nederland*. 266862. s-Gravenhage SDU.

Ballabio, C., Lugato, E., Fernández-Ugalde, O., Orgiazzi, A., Jones, A., Borrelli, P., Montanarella, L., &

Panagos, P. (2019). Mapping LUCAS topsoil chemical properties at European scale using Gaussian process regression. *Geoderma*, *355*, 113912. doi: 10.1016/j.geoderma.2019.113912.

Ballabio, C., Panagos, P., & Monatanarella, L. (2016). Mapping topsoil physical properties at European scale using the LUCAS database. *Geoderma*, *261*, 110–123. doi: 10.1016/j.geoderma.2015.07.006.

Ballantine, K., & Schneider, R. (2009). Fifty-five years of soil development in restored freshwater depressional wetlands. *Ecological Applications*, *19*, 1467–1480. doi: 10.1890/07-0588.1. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1890/07-0588.1.

Baltensweiler, A., Walthert, L., Hanewinkel, M., Zimmermann, S., & Nussbaum, M. (2021). Machine learning based soil maps for a wide range of soil properties for the forested area of Switzerland. *Geoderma Regional*, *27*, e00437. doi: 10.1016/j.geodrs.2021.e00437.

Baptist, M., van Hattum, T., Reinhard, S., van Buuren, M., de Rooij, B., Hu, X., van Rooij, S., Polman, N., van den Burg, S., Piet, G., & Ysebaert, T. (2019). *A nature-based future for the Netherlands in 2120*. Knowledge Base Program Nature Inclusive Transitions project number KB-36-003-004, Wageningen University & Research Wageningen, the Netherlands. `https://library.wur.nl/WebQuery/wurpubs/fulltext/512277`.

van Baren, H., Hartemink, A. E., & Tinker, P. B. (2000). 75 years The International Society of Soil Science. *Geoderma*, *96*, 1–18. doi: 10.1016/S0016-7061(99)00097-X.

Batjes, N. H. (2016). Harmonized soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks. *Geoderma*, *269*, 61–68. doi: 10.1016/j.geoderma.2016.01.034.

Baumann, P., Helfenstein, A., Gubler, A., Keller, A., Meuli, R. G., Wächter, D., Lee, J., Viscarra Rossel, R., & Six, J. (2021). Developing the Swiss mid-infrared soil spectral library for local estimation and monitoring. *SOIL*, *7*, 525–546. doi: 10.5194/soil-7-525-2021. Publisher: Copernicus GmbH.

van Beek, C. L., Pleijter, M., & Kuikman, P. J. (2011). Nitrous oxide emissions from fertilized and unfertilized grasslands on peat soil. *Nutrient Cycling in Agroecosystems*, *89*, 453–461. doi: 10.1007/s10705-010-9408-y.

Been, T. H., Kempenaar, C., van Evert, F. K., Hoving, I. E., Kessel, G. J. T., Dantuma, W., Booij, J. A., Molendijk, L. P. G., Sijbrandij, F. D., & van Boheemen, K. (2023). Akkerweb and farmmaps: Development of Open Service Platforms for Precision Agriculture. In D. Cammarano, F. K. van Evert, & C. Kempenaar (Eds.), *Precision Agriculture: Modelling* Progress in Precision Agriculture (pp. 269–293). Cham: Springer International Publishing. doi: 10.1007/978-3-031-15258-0_16.

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

Behrens, T., Schmidt, K., MacMillan, R. A., & Viscarra Rossel, R. A. (2018a). Multi-scale digital soil mapping with deep learning. *Scientific Reports*, *8*. doi: 10.1038/s41598-018-33516-6.

Behrens, T., Schmidt, K., Rossel, R. A. V., Gries, P., Scholten, T., & MacMillan, R. A. (2018b). Spatial modelling with Euclidean distance fields and machine learning. *European Journal of Soil Science*, *69*, 757–770. doi: 10.1111/ejss.12687.

Beillouin, D., Cardinael, R., Berre, D., Boyer, A., Corbeels, M., Fallot, A., Feder, F., & Demenois, J. (2022). A global overview of studies about land management, land-use change, and climate change effects on soil organic carbon. *Global Change Biology*, *28*, 1690–1702. doi: 10.1111/gcb.15998. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.15998.

Beillouin, D., Corbeels, M., Demenois, J., Berre, D., Boyer, A., Fallot, A., Feder, F., & Cardinael, R. (2023). A global meta-analysis of soil organic carbon in the Anthropocene. *Nature Communications*, *14*, 3700. doi: 10.1038/s41467-023-39338-z. Number: 1 Publisher: Nature Publishing Group.

van den Berg, F., Tiktak, A., Hoogland, T., Poot, A., Boesten, J., van der Linden, A. M. A., & Pol, J. W. (2017). *An improved soil organic matter map for GeoPEARL_NL: Model description of version 4.4.4 and consequence for the Dutch decision tree on leaching to groundwater*. Technical Report, Wageningen Environmental Research (Alterra) Wageningen. `http://library.wur.nl/WebQuery/wurpubs/`

532249. doi: 10.18174/424920.

BIJ12 (2019). Informatiemodel Natuur (IMNa).

Bishop, T. F. A., & McBratney, A. B. (2001). A comparison of prediction methods for the creation of field-extent soil property maps. *Geoderma*, *103*, 149–160. doi: 10.1016/S0016-7061(01)00074-X.

Blume, H.-P., & Leinweber, P. (2004). Plaggen Soils: landscape history, properties, and classification. *Journal of Plant Nutrition and Soil Science*, *167*, 319–327. doi: 10.1002/jpln.200420905. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jpln.200420905.

Blume, H.-P., Scheffer, F., Schachtschabel, P., Brümmer, G., Horn, R., Kandeler, E., Kögel-Knabner, I., Kretzschmar, R., Stahr, K., & Wilke, B.-M. (2010). *Scheffer/Schachtschabel: Lehrbuch der Bodenkunde*. (16th ed.). Heidelberg: Springer, Spektrum, Akademischer Verlag. OCLC: 506415938.

Boehmke, B., & Greenwell, B. (2020). *Hands-On Machine Learning with R*. Taylor & Francis.

Borrelli, P. et al. (2017). An assessment of the global impact of 21st century land use change on soil erosion. *Nature Communications*, *8*, 2013. doi: 10.1038/s41467-017-02142-7. Number: 1 Publisher: Nature Publishing Group.

Bouma, J., & Hartemink, A. E. (2003). Soil science and society in the Dutch context. *Netherlands Journal of Agricultural Science*, *50*, 133–140. Number: 2.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32. doi: 10.1023/A:1010933404324.

Breiman, L. (2002). *Manual on Setting Up, Using, and Understanding Random Forests v3.1*. Technical report ftp://ftp.stat.berkeley.edu/pub/users/breiman/Using_random_forests_v3.1.pdf, University of Berkeley Berkeley. ftp://ftp.stat.berkeley.edu/pub/users/breiman/Using_random_forests_v3.1.pdf.

Breman, B. et al. (2022). *Natuurverkenning 2050 – Scenario Natuurinclusief*. Technical Report 136, Wettelijke Onderzoekstaken Natuur & Milieu Wageningen. https://library.wur.nl/WebQuery/wurpubs/594007 iSSN: 1871-028X.

Brenning, A. (2005). Spatial prediction models for landslide hazards: review, comparison and evaluation. *Natural Hazards and Earth System Sciences*, *5*, 853–862. doi: 10.5194/nhess-5-853-2005. Publisher: Copernicus GmbH.

Brenning, A. (2012). Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The R package sperrorest. In *2012 IEEE International Geoscience and Remote Sensing Symposium* (pp. 5372–5375). doi: 10.1109/IGARSS.2012.6352393 iSSN: 2153-7003.

Hack-ten Broeke, M., van Beek, C. L., Hoogland, T., Knotters, M., Mol-Dijkstra, J. P., Schils, R., Smit, A., & de Vries, F. (2009). *Kaderrichtlijn Bodem; Basismateriaal voor eventuele prioritaire gebieden*. Alterra-rapport 2007, Alterra Wageningen, the Netherlands.

Hack-ten Broeke, M. J. D. et al. (2019). Quantitative land evaluation implemented in Dutch water management. *Geoderma*, *338*, 536–545. doi: 10.1016/j.geoderma.2018.11.002.

Brouwer, F., Assinck, F., Harkema, T., Teuling, K., & Walvoort, D. (2023). *Actualisatie van de bodemkaart in degemeente Vijfheerenlanden: herkartering van de verbreiding van veen*. WOt-rapport 151, WOT Natuur & Milieu Wageningen. https://research.wur.nl/en/publications/actualisatie-van-de-bodemkaart-in-degemeente-vijfheerenlanden-her publisher: WOT Natuur & Milieu.

Brouwer, F., Maas, G., Teuling, K., Harkema, T., & Verzandvoort, S. (2021). *Bodemkaart en Geomorfologische Kaart van Nederland: actualisatie 2020-2021 en toepassing: deelgebieden Gelderse Vallei-Zuid en -West en Veluwe-Zuid*. WOt-rapport 134, WOT Natuur & Milieu Wageningen. https://doi.org/10.18174/557455 publisher: WOT Natuur & Milieu.

Brouwer, F., de Vries, F., & Walvoort, D. J. J. (2018). *Basisregistratie Ondergrond (BRO) actualisatie bodemkaart : Herkartering van de bodem in Flevoland*. WOt technical report 143, WOT Natuur & Milieu, WUR Wageningen. https://library.wur.nl/WebQuery/wurpubs/549064 iSSN: 2352-2739.

Brouwer, F., & Walvoort, D. (2019). *Basisregistratie Ondergrond (BRO) - actualisatie bodemkaart : Herkartering van de bodem in Eemland*. WOt-technical report 155, WOT Natuur & Milieu Wagenin-

gen. https://doi.org/10.18174/494728. doi: 10.18174/494728 iSSN (Print): 2352-2739.

Brouwer, F., & Walvoort, D. (2020). *Basisregistratie Ondergrond (BRO) Actualisatie bodemkaart : Herkartering van de veengebieden aan de flanken van de Utrechtse Heuvelrug*. WOt-technical report 177, WOT Natuur & Milieu Wageningen. https://doi.org/10.18174/521574 iSSN (Print): 2352-2739.

Brouwer, F., & van der Werff, M. M. (2012). *Vergraven gronden : inventarisatie van 'diepe' grondbewerkingen, ophogingen en afgravingen*. Alterra-rapport 2336, Alterra Wageningen. https://edepot.wur.nl/217669 publisher: Alterra.

BRT (2020). *Basisregistratie Topografie (BRT): Catalogus en Productspecificaties*. Kadaster Versie 1.2.0.3, BRT.

de Bruin, S., Bregt, A., & Ven, M. v. d. (2001). Assessing fitness for use: the expected value of spatial data sets. *International Journal of Geographical Information Science*, *15*, 457–471. doi: 10.1080/13658810110053116.

de Bruin, S., Brus, D. J., Heuvelink, G. B. M., van Ebbenhorst Tengbergen, T., & Wadoux, A. M. J.-C. (2022). Dealing with clustered samples for assessing map accuracy by cross-validation. *Ecological Informatics*, *69*, 101665. doi: 10.1016/j.ecoinf.2022.101665.

Brus, D., Hengl, T., Heuvelink, G., Kempen, B., Mulder, T. V., Olmedo, G. F., Poggio, L., Ribeiro, E., & Omuto, C. T. (2017). *Soil Organic Carbon Mapping Cookbook*. (1st ed.). Rome: FAO.

Brus, D. J. (2014). Statistical sampling approaches for soil monitoring. *European Journal of Soil Science*, *65*, 779–791. doi: 10.1111/ejss.12176. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejss.12176.

Brus, D. J. (2019). Sampling for digital soil mapping: A tutorial supported by R scripts. *Geoderma*, *338*, 464–480. doi: 10.1016/j.geoderma.2018.07.036.

Brus, D. J. (2022). *Spatial sampling with R*. The R Series. CRC Press.

Brus, D. J., & Heuvelink, G. B. M. (2007). *Towards a Soil Information System with quantified accuracy: Three approaches for stochastic simulation of soil maps*. Statutory Research Tasks Unit for Nature and the Environment 58, Alterra Wageningen.

Brus, D. J., Kempen, B., & Heuvelink, G. B. M. (2011). Sampling for validation of digital soil maps. *European Journal of Soil Science*, *62*, 394–407. doi: https://doi.org/10.1111/j.1365-2389.2011.01364.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2389.2011.01364.x.

Brus, D. J., Vašát, R., Heuvelink, G. B. M., Knotters, M., Vries, F. d., & Walvoort, D. J. J. (2009). *Towards a Soil Information System with quantified accuracy. A prototype for mapping continuous soil properties*. Statutory Research Tasks Unit for Nature and the Environment 197, Alterra Wageningen.

Brändli, G., Zihlmann, U., & Chervet, A. (2016). Erbsen auf drainiertem, tiefgepfluegtem Moorboden. Agroscope & Amt für Landwirtschaft und Natur des Kantons Bern (LANAT).

Buchanan, S., Triantafilis, J., Odeh, I. O. A., & Subansinghe, R. (2012). Digital soil mapping of compositional particle-size fractions using proximal and remotely sensed ancillary data. *GEOPHYSICS*, *77*, WB201–WB211. doi: 10.1190/geo2012-0053.1. Publisher: Society of Exploration Geophysicists.

Burgess, T. M., & Webster, R. (1980a). Optimal Interpolation and Isarithmic Mapping of Soil Properties. I. *Journal of Soil Science*, *31*, 315–331. doi: 10.1111/j.1365-2389.1980.tb02084.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2389.1980.tb02084.x.

Burgess, T. M., & Webster, R. (1980b). Optimal Interpolation and Isarithmic Mapping of Soil Properties. II. *Journal of Soil Science*, *31*, 333–341. doi: 10.1111/j.1365-2389.1980.tb02085.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2389.1980.tb02085.x.

Buringh, P., Stuer, G. G. L., & Vink, P. (1962). Some techniques and methods of soil survey in the Netherlands. *Netherlands Journal of Agricultural Science*, *10*, 17.

Burrough, P. A., McDonnell, R. A., & Lloyd, C. D. (2015). *Principles of Geographical Information Systems*. OUP Oxford. Google-Books-ID: kvoJCAAAQBAJ.

BZK (2022). Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK): Uitvoeringsregeling Mest-

stoffenwet. Last Modified: 2024-01-17.

ten Cate, J., van Holst, A., Kleijer, H., & Stolp, J. (1995). *Handleiding bodemgeografisch onderzoek: richtlijnen en voorschriften. Deel A: Bodem*. Technisch Document 19A, DLO-Staring Centrum Wageningen. https://edepot.wur.nl/380178.

Cavero Panez, C. (2021). *The Impact of Temporal Land Use Variability on Soil Organic Carbon Changes: A case study at point level in The Netherlands from 1998-2018*. Master Thesis Report GIRS-2021-49 Wageningen University Wageningen.

CBS (2015). Bestand Bodemgebruik (BBG): 1993, 1996, 2000, 2003, 2006, 2008, 2010, 2012, 2015. Centraal Bureau voor de Statistiek (CBS). Last Modified: 28-03-2019T14:09:28.

Chabbi, A., Lehmann, J., Ciais, P., Loescher, H. W., Cotrufo, M. F., Don, A., SanClements, M., Schipper, L., Six, J., Smith, P., & Rumpel, C. (2017). Aligning agriculture and climate policy. *Nature Climate Change*, *7*, 307–309. doi: 10.1038/nclimate3286. Number: 5 Publisher: Nature Publishing Group.

Chagas, C. d. S., de Carvalho Junior, W., Bhering, S. B., & Calderano Filho, B. (2016). Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. *CATENA*, *139*, 232–240. doi: 10.1016/j.catena.2016.01.001.

Chaney, N. W., Minasny, B., Herman, J. D., Nauman, T. W., Brungard, C. W., Morgan, C. L. S., McBratney, A. B., Wood, E. F., & Yimam, Y. (2019). POLARIS Soil Properties: 30-m Probabilistic Maps of Soil Properties Over the Contiguous United States. *Water Resources Research*, *55*, 2916–2938. doi: 10.1029/2018WR022797. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2018WR022797.

Chardon, W. J., Heesmans, H., & Kuikman, P. J. (2009). *Trends in carbon stocks in Dutch soils: datasets and modeling results*. Alterra-rapport 1869, Alterra Wageningen. https://edepot.wur.nl/51665.

Charman, D. (2002). *Peatlands and environmental change*. Chichester: John Wiley & Sons Ltd. Publisher: John Wiley & Sons Ltd.

Chatfield, C. (1995). Model Uncertainty, Data Mining and Statistical Inference. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *158*, 419–466. doi: 10.2307/2983440. Publisher: [Wiley, Royal Statistical Society].

Chen, S. et al. (2022). Digital mapping of GlobalSoilMap soil properties at a broad scale: A review. *Geoderma*, *409*, 115567. doi: 10.1016/j.geoderma.2021.115567.

Chen, S., Liang, Z., Webster, R., Zhang, G., Zhou, Y., Teng, H., Hu, B., Arrouays, D., & Shi, Z. (2019). A high-resolution map of soil pH in China made by hybrid modelling of sparse soil data and environmental covariates and its implications for pollution. *Science of The Total Environment*, *655*, 273–283. doi: 10.1016/j.scitotenv.2018.11.230.

Chen, S., Saby, N. P. A., Martin, M. P., Barthès, B. G., Gomez, C., Shi, Z., & Arrouays, D. (2023). Integrating additional spectroscopically inferred soil data improves the accuracy of digital soil mapping. *Geoderma*, *433*, 116467. doi: 10.1016/j.geoderma.2023.116467.

Chinilin, A., & Savin, I. Y. (2023). Combining machine learning and environmental covariates for mapping of organic carbon in soils of Russia. *The Egyptian Journal of Remote Sensing and Space Sciences*, *26*, 666–675. doi: 10.1016/j.ejrs.2023.07.007.

Clement, J. (2001). *GIS Vierde Bosstatistiek: Gebruikersdocumentatie, Documentatie van bestanden*. Technical Report, Research Instituut voor de Groene Ruimte, Alterra Wageningen.

Cochran, W. G. (1977). *Sampling techniques*. (3rd ed.). New York: John Wiley & Sons.

Coleman, K., & Jenkinson, D. (1996). RothC-26.3 - A Model for the turnover of carbon in soil. In *Evaluation of Soil Organic Matter Models Using Existing, Long-Term Datasets* (pp. 237–246). Harpenden: Rothamsted Research volume 38. doi: 10.1007/978-3-642-61094-3_17 journal Abbreviation: Evaluation of Soil Organic Matter Models Using Existing, Long-Term Datasets.

Conijn, J. G., & Lesschen, J. P. (2015). *Soil organic matter in the Netherlands : Quantification of stocks and flows in the top soil*. Technical Report 619, Plant Research International, Business Unit Agrosystems Research Wageningen. https://library.wur.nl/WebQuery/wurpubs/498774.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., &

Böhner, J. (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, *8*, 1991–2007. doi: 10.5194/gmd-8-1991-2015. Publisher: Copernicus GmbH.

Craft, C. (2022). 1 - Introduction. In C. Craft (Ed.), *Creating and Restoring Wetlands (Second Edition)* (pp. 1–24). Elsevier. doi: 10.1016/B978-0-12-823981-0.00013-7.

Creamer, R. E., Barel, J. M., Bongiorno, G., & Zwetsloot, M. J. (2022). The life of soils: Integrating the who and how of multifunctionality. *Soil Biology and Biochemistry*, *166*, 108561. doi: 10.1016/j.soilbio.2022.108561.

Crews, T. E., & Rumsey, B. E. (2017). What Agriculture Can Learn from Native Ecosystems in Building Soil Organic Matter: A Review. *Sustainability*, *9*, 578. doi: 10.3390/su9040578. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

van Dam, J. C., Huygen, J., Wesseling, J. G., Feddes, R. A., & Kabat, P. (1997). *Theory of SWAP version 2.0; Simulation of waterflow, solute transport and plant growth in the Soil-Water-Atmosphere-Plant environment*. Technical document 45, Wageningen Agricultural University and DLO Winand Staring Centre Wageningen, the Netherlands. https://library.wur.nl/WebQuery/wurpubs/fulltext/222782.

Debonne, N., Bürgi, M., Diogo, V., Helfenstein, J., Herzog, F., Levers, C., Mohr, F., Swart, R., & Verburg, P. (2022). The geography of megatrends affecting European agriculture. *Global Environmental Change*, *75*, 102551. doi: 10.1016/j.gloenvcha.2022.102551.

van Delft, B., & Maas, G. (2023). *Landschappelijke Bodemkartering (LBK) : Achtergronden, toepassingen en technische documentatie*. WOt-technical report 248, Wettelijke Onderzoekstaken Natuur & Milieu Wageningen. https://doi.org/10.18174/641887. doi: 10.18174/641887 iSSN: 2352-2739.

van Delft, S. P. J., & Maas, G. J. (2022). De Landschappelijke Bodemkaart van Nederland; versie 2022.

Delta Programme (2023). *National Delta Programme 2024: Now for the Future*. Technical Report, Ministry of Infrastructure and Water Management, the Ministry of Agriculture, Nature and Food Quality, and the Ministry of the Interior and Kingdom Relations Den Haag. https://english.deltaprogramma.nl/.

Dharumarajan, S., Hegde, R., Janani, N., & Singh, S. K. (2019). The need for digital soil mapping in India. *Geoderma Regional*, *16*, e00204. doi: 10.1016/j.geodrs.2019.e00204.

Dharumarajan, S., Kalaiselvi, B., Suputhra, A., Lalitha, M., Hegde, R., Singh, S. K., & Lagacherie, P. (2020). Digital soil mapping of key GlobalSoilMap properties in Northern Karnataka Plateau. *Geoderma Regional*, *20*, e00250. doi: 10.1016/j.geodrs.2019.e00250.

Dohong, A., Aziz, A. A., & Dargusch, P. (2017). A review of the drivers of tropical peatland degradation in South-East Asia. *Land Use Policy*, *69*, 349–360. doi: 10.1016/j.landusepol.2017.09.035.

Dokuchaev, V. (1899). *Report to the Transcaucasian Statistical Committee on Land Evaluation in General and Especially for the Transcaucasia. Horizontal and Vertical Soil Zones. (In Russian.)*. Tiflis, Russia: Off. Press Civ, Affairs Commander-in-Chief Cacasus.

Domburg, P., de Gruijter, J. J., & van Beek, P. (1997). Designing efficient soil survey schemes with a knowledge-based system using dynamic programming. *Geoderma*, *75*, 183–201. doi: 10.1016/S0016-7061(96)00090-0.

Doorn, A. v., Melman, D., Westerink, J., Polman, N., Vogelzang, T., & Korevaar, H. (2016). *Food-for-thought : natuurinclusieve landbouw*. Technical Report, Wageningen University & Research Wageningen. https://doi.org/10.18174/401503 publisher: Wageningen University & Research.

van Doorn, M., Helfenstein, A., Ros, G. H., Heuvelink, G. B., van Rotterdam-Los, D. A., Verweij, S. E., & de Vries, W. (2024). High-Resolution Digital Soil Mapping of Amorphous Iron- and Aluminium-(hydr)oxides to Guide Sustainable Phosphorus and Carbon Management. . Under review for Geoderma.

van Dorland, R. et al. (2023). *KNMI National Climate Scenarios 2023 for the Netherlands*. Scientific report WR-23-02, Royal Netherlands Meteorological Institute (KNMI); Ministry of Infrastructure and Water Management (I & W) De Bilt, the Netherlands. https://cdn.knmi.nl/system/data_center_publications/files/000/071/902/original/

KNMI23_climate_scenarios_scientific_report_WR23-02.pdf.

Dvorakova, K., Heiden, U., Pepers, K., Staats, G., van Os, G., & van Wesemael, B. (2023). Improving soil organic carbon predictions from a Sentinel–2 soil composite by assessing surface conditions and uncertainties. *Geoderma*, *429*, 116128. doi: 10.1016/j.geoderma.2022.116128.

Dvorakova, K., Heiden, U., & van Wesemael, B. (2021). Sentinel-2 Exposed Soil Composite for Soil Organic Carbon Prediction. *Remote Sensing*, *13*, 1791. doi: 10.3390/rs13091791. Number: 9 Publisher: Multidisciplinary Digital Publishing Institute.

van Ebbenhorst Tengbergen, T. (2021). *Critical evaluation and improvement of cross-validation strategies for accuracy assessment of digital soil maps*. Thesis Report GIRS-2021-13 Wageningen University Wageningen.

Edelmann, C. (1950). *Soils of the Netherlands* volume 53. Amsterdam: North-Holland Publishing Company. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jpln.19510530307.

Edmonds, D. A., Caldwell, R. L., Brondizio, E. S., & Siani, S. M. O. (2020). Coastal flooding will disproportionately impact people on river deltas. *Nature Communications*, *11*, 4741. doi: 10.1038/s41467-020-18531-4. Number: 1 Publisher: Nature Publishing Group.

EEA (2007). *CLC2006 technical guidelines*. EEA Technical Report 17, European Environment Agency (EEA) Copenhagen. https://data.europa.eu/doi/10.2800/12134.

EEA (2018). CORINE Land Cover — Copernicus Land Monitoring Service: 1986, 2000, 2006, 2012, 2018. European Environment Agency (EEA).

van den Elsen, E., van Tol-Leenders, D., Teuling, K., Römkens, P., de Haan, J., Korthals, G., & Reijneveld, A. (2020). *De staat van de Nederlandse landbouwbodems in 2018 : Op basis van beschikbare landsdekkende dataset (CC-NL) en bodem-indicatorenlijst (BLN)*. Technical Report 3048, Wageningen Environmental Research Wageningen. https://library.wur.nl/WebQuery/wurpubs/574884 iSSN: 1566-7197.

Ericson, J. P., Vörösmarty, C. J., Dingman, S. L., Ward, L. G., & Meybeck, M. (2006). Effective sea-level rise and deltas: Causes of change and human dimension implications. *Global and Planetary Change*, *50*, 63–82. doi: 10.1016/j.gloplacha.2005.07.004.

Erisman, J. W. (2021). Setting ambitious goals for agriculture to meet environmental targets. *One Earth*, *4*, 15–18. doi: 10.1016/j.oneear.2020.12.007.

Erkens, G., van der Meulen, M. J., & Middelkoop, H. (2016). Double trouble: subsidence and CO2 respiration due to 1,000 years of Dutch coastal peatlands cultivation. *Hydrogeology Journal*, *24*, 551–568. doi: 10.1007/s10040-016-1380-4.

Eurofins Agro (2024a). Bemesting Wetgeving.

Eurofins Agro (2024b). BemestingsWijzer.

European Commission (2021). *A Soil Deal for Europe: 100 living labs and lighthouses to lead the transition towards healthy soils by 2030*. Implementation Plan, European Commission. https://ec.europa.eu/info/sites/default/files/research_and_innovation/funding/documents/soil_mission_implementation_plan_final_for_publication.pdf.

European Commission (2023a). The European Green Deal.

European Commission (2023b). *Proposal for a Directive on Soil Monitoring and Resilience*. Technical Report, European Commission. https://environment.ec.europa.eu/publications/proposal-directive-soil-monitoring-and-resilience_en.

Eurostat (2022). Key figures on the European food chain.

Eurostat (2023). *Environmental protection investments*. Technical Report, . https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Environmental_protection_expenditure_accounts.

EZK (2013). Fysisch Geografische Regio's 2013; Ministerie van Economische Zaken en Klimaat (EZK; Ministry of Economic Affairs and Climate).

EZK (2019). Basisregistratie Gewaspercelen (BRP): 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013,

2014, 2015, 2016, 2017, 2018, 2019. Ministerie van Economische Zaken en Klimaat (EZK; Ministry of Economic Affairs and Climate), Agrarische Areaal Nederland.

FAO (2017). Global Soil Organic Carbon (GSOC) Map.

FAO (2018). *Soil Organic Carbon Mapping Cookbook*. (2nd ed.). Rome, Italy: FAO.

Felix, R. (1995). Bodemkartering voor 1943—het geologisch perspectief. In P. Buurman, & J. Sevink (Eds.), *Van bodemkaart tot informatiesysteem* (pp. 1–17). Wageningen: Wageningen Pers.

Ferguson, J., Littman, R., Christensen, G., Paluck, E. L., Swanson, N., Wang, Z., Miguel, E., Birke, D., & Pezzuto, J.-H. (2023). Survey of open science practices and attitudes in the social sciences. *Nature Communications*, *14*, 5401. doi: 10.1038/s41467-023-41111-1. Number: 1 Publisher: Nature Publishing Group.

Fernández-Ugalde, O., Ballabio, C., Lugato, E., Scarpa, S., & Jones, A. (2020). *Assessment of changes in topsoil properties in LUCAS samples between 2009/2012 and 2015 surveys*. Technical Report, Publications Office of the European Union Joint Research Centre (European Commission). `https://data.europa.eu/doi/10.2760/5503`.

Filippi, P., Jones, E. J., & Bishop, T. F. A. (2020). Catchment-scale 3D mapping of depth to soil sodicity constraints through combining public and on-farm soil databases – A potential tool for on-farm management. *Geoderma*, *374*, 114396. doi: 10.1016/j.geoderma.2020.114396.

Filippi, P., Jones, E. J., Ginns, B. J., Whelan, B. M., Roth, G. W., & Bishop, T. F. A. (2019). Mapping the Depth-to-Soil pH Constraint, and the Relationship with Cotton and Grain Yield at the Within-Field Scale. *Agronomy*, *9*, 251. doi: 10.3390/agronomy9050251. Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.

Finke, P. A. (2012). On digital soil assessment with models and the Pedometrics agenda. *Geoderma*, *171-172*, 3–15. doi: 10.1016/j.geoderma.2011.01.001.

Finke, P. A., de Gruijter, J. J., & Visschers, R. (2001). *Status 2001 Landelijke Steekproef Kaarteenheden en toepassingen; Gestructureerde bemonstering en karakterisering Nederlandse bodems*. Alterra-rapport 389, Alterra, Research Instituut voor de Groene Ruimte Wageningen.

Francos, N., Heller-Pearlshtien, D., Demattê, J. A. M., Van Wesemael, B., Milewski, R., Chabrillat, S., Tziolas, N., Sanz Diaz, A., Yagüe Ballester, M. J., Gholizadeh, A., & Ben-Dor, E. (2023). A Spectral Transfer Function to Harmonize Existing Soil Spectral Libraries Generated by Different Protocols. *Applied and Environmental Soil Science*, *2023*, e4155390. doi: 10.1155/2023/4155390. Publisher: Hindawi.

Fuller, W. A. (1999). Environmental Surveys over Time. *Journal of Agricultural, Biological, and Environmental Statistics*, *4*, 331–345. doi: 10.2307/1400493. Publisher: [International Biometric Society, Springer].

Gallant, J. C., & Dowling, T. I. (2003). A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resources Research*, *39*. doi: https://doi.org/10.1029/2002WR001426. _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2002WR001426.

Gasch, C. K., Hengl, T., Gräler, B., Meyer, H., Magney, T. S., & Brown, D. J. (2015). Spatio-temporal interpolation of soil water, temperature, and electrical conductivity in 3D + T: The Cook Agronomy Farm data set. *Spatial Statistics*, *14*, 70–90. doi: 10.1016/j.spasta.2015.04.001.

GDAL/OGR contributors (2023). GDAL/OGR geospatial data abstraction software library.

Global Soil Data Task Group (2000). *Global Gridded Surfaces of Selected Soil Characteristics (IGBP-DIS)*. Oak Ridge, Tennessee, USA: ORNL DAAC.

Gomes, L. C., Faria, R. M., de Souza, E., Veloso, G. V., Schaefer, C. E. G. R., & Filho, E. I. F. (2019). Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma*, *340*, 337–350. doi: 10.1016/j.geoderma.2019.01.007.

Gong, Z., Zhang, X., Chen, J., & Zhang, G. (2003). Origin and development of soil science in ancient China. *Geoderma*, *115*, 3–13. doi: 10.1016/S0016-7061(03)00071-5.

Goovaerts, P. (1999). Using elevation to aid the geostatistical mapping of rainfall erosivity. *CATENA*,

34, 227–242. doi: 10.1016/S0341-8162(98)00116-7.

Goovaerts, P. (2001). Geostatistical modelling of uncertainty in soil science. *Geoderma*, *103*, 3–26. doi: 10.1016/S0016-7061(01)00067-2.

Government of the Netherlands (2019). *Climate Agreement*. Technical Report, Dutch government The Hague.

Gray, J. M., & Bishop, T. F. A. (2016). Change in Soil Organic Carbon Stocks under 12 Climate Change Projections over New South Wales, Australia. *Soil Science Society of America Journal*, *80*, 1296–1307. doi: 10.2136/sssaj2016.02.0038. _eprint: https://acsess.onlinelibrary.wiley.com/doi/pdf/10.2136/sssaj2016.02.0038.

Gray, J. M., & Bishop, T. F. A. (2019). Corrigendum to: Mapping change in key soil properties due to climate change over south-eastern Australia. *Soil Research*, *57*, 805–805. doi: 10.1071/sr18139_co. Publisher: CSIRO PUBLISHING.

Greenberg, I., Seidel, M., Vohland, M., Koch, H.-J., & Ludwig, B. (2022a). Performance of in situ vs laboratory mid-infrared soil spectroscopy using local and regional calibration strategies. *Geoderma*, *409*, 115614. doi: 10.1016/j.geoderma.2021.115614.

Greenberg, I., Seidel, M., Vohland, M., & Ludwig, B. (2022b). Performance of field-scale lab vs in situ visible/near- and mid-infrared spectroscopy for estimation of soil properties. *European Journal of Soil Science*, *73*, e13180. doi: 10.1111/ejss.13180. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejss.13180.

Gregoire, T. G., & Valentine, H. T. (2007). *Sampling Strategies for Natural Resources and the Environment*. Boca Raton, USA: CRC Press. Google-Books-ID: fUvLBQAAQBAJ.

Grigelis, A., Wójcik, Z., Narebski, W., Gelumbauskaitė, L. , & Kozák, J. (2011). Stanisław Staszic: An Early Surveyor of the Geology of Central and Eastern Europe. *Annals of Science*, *68*, 199–228. doi: 10.1080/00033790.2010.511263. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00033790.2010.511263.

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

Groenendijk, P., Renaud, L. V., Boekel, E. M. P. M. v., Salm, C. v. d., & Schoumans, O. F. (2013). *Voorbereiding STONE2.4 op berekeningen voor de evaluatie Meststoffenwet 2012*. Technical Report 2462, Alterra, Wageningen-UR Wageningen. `https://library.wur.nl/WebQuery/wurpubs/451011` iSSN: 1566-7197.

de Gruijter, J. J., Brus, D., Bierkens, M., & Knotters, M. (2006). *Sampling for Natural Resource Monitoring*. The Netherlands: Springer.

de Gruijter, J. J., van der Horst, J. B. F., Heuvelink, G. B. M., Knotters, M., & Hoogland, T. (2004). *Grondwater opnieuw op de kaart; methodiek voor de actualisering van grondwaterstandsinformatie en perceelsclassificatie naar uitspoelingsgevoeligheid voor nitraat*. Alterra-rapport 915, Alterra Wageningen. `https://edepot.wur.nl/26169` publisher: Alterra.

de Gruijter, J. J., Walvoort, D. J. J., & van Gams, P. F. M. (1997). Continuous soil maps — a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models. *Geoderma*, *77*, 169–195. doi: 10.1016/S0016-7061(97)00021-9.

de Gruijter, J. J., Webster, R., & Myers, D. E. (1994). Preface. *Geoderma*, *62*, vii–viii. doi: 10.1016/0016-7061(94)90021-3.

Grundy, M. J., Rossel, R. A. V., Searle, R. D., Wilson, P. L., Chen, C., & Gregory, L. J. (2015). Soil and Landscape Grid of Australia. *Soil Research*, *53*, 835–844. doi: 10.1071/SR15191. Publisher: CSIRO PUBLISHING.

Gupta, S., Hengl, T., Lehmann, P., Bonetti, S., & Or, D. (2021). SoilKsatDB: global database of soil saturated hydraulic conductivity measurements for geoscience applications. *Earth System Science Data*, *13*, 1593–1612. doi: 10.5194/essd-13-1593-2021. Publisher: Copernicus GmbH.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning*, *46*, 389–422. doi: 10.1023/A:1012487302797.

Hartemink, A. E., Hempel, J., Lagacherie, P., McBratney, A., McKenzie, N., MacMillan, R. A., Minasny, B., Montanarella, L., de Mendonça Santos, M. L., Sanchez, P., Walsh, M., & Zhang, G.-L. (2010). GlobalSoilMap.net – A New Digital Soil Map of the World. In J. L. Boettinger, D. W. Howell, A. C. Moore, A. E. Hartemink, & S. Kienast-Brown (Eds.), *Digital Soil Mapping: Bridging Research, Environmental Application, and Operation* Progress in Soil Science (pp. 423–428). Dordrecht: Springer Netherlands. doi: 10.1007/978-90-481-8863-5_33.

Hartemink, A. E., Krasilnikov, P., & Bockheim, J. G. (2013). Soil maps of the world. *Geoderma*, *207-208*, 256–267. doi: 10.1016/j.geoderma.2013.05.003.

Hartemink, A. E., & McBratney, A. (2008). A soil science renaissance. *Geoderma*, *148*, 123–129. doi: 10.1016/j.geoderma.2008.10.006.

Hartemink, A. E., & Sonneveld, M. P. W. (2013). Soil maps of The Netherlands. *Geoderma*, *204-205*, 1–9. doi: 10.1016/j.geoderma.2013.03.022.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer series in statistics (2nd ed.). New York, NY: Springer.

Hazeu, G., Schuiling, R., Thomas, D., Vittek, M., Storm, M., & Bulens, J. D. (2023). *Landelijk Grondgebruiksbestand Nederland 2021 (LGN2021) : achtergronden, methodiek en validatie*. Rapport 3235, Wageningen Environmental Research Wageningen. `https://research.wur.nl/en/publications/landelijk-grondgebruiksbestand-nederland-2021-lgn2021-achtergrond` publisher: Wageningen Environmental Research.

Hazeu, G. W. (2014). Operational Land Cover and Land Use Mapping in the Netherlands. In I. Manakos, & M. Braun (Eds.), *Land Use and Land Cover Mapping in Europe: Practices & Trends* Remote Sensing and Digital Image Processing (pp. 283–296). Dordrecht: Springer Netherlands. doi: 10.1007/978-94-007-7969-3_18.

Hazeu, G. W., Vittek, M., Schuiling, R., Bulens, J. D., Storm, M. H., Roerink, G. J., & Meijninger, W. M. L. (2020). *LGN2018: een nieuwe weergave van het grondgebruik in Nederland*. Technical Report 3010, Wageningen Environmental Research Wageningen. `https://library.wur.nl/WebQuery/wurpubs/565896` iSSN: 1566-7197.

Hazeu, G. W., & Wit, A. J. W. d. (2004). CORINE land cover database of the Netherlands: monitoring land cover changes between 1986 and 2000. *EARSeL eProceedings*, *3*, 382–387. Number: 3.

Heinen, M., Mulder, H. M., Bakker, G., Wösten, J. H. M., Brouwer, F., Teuling, K., & Walvoort, D. J. J. (2022). The Dutch soil physical units map: BOFEK. *Geoderma*, *427*, 116123. doi: 10.1016/j.geoderma.2022.116123.

Helfenstein, A., Baumann, P., Viscarra Rossel, R., Gubler, A., Oechslin, S., & Six, J. (2021). Quantifying soil carbon in temperate peatlands using a mid-IR soil spectral library. *SOIL*, *7*, 193–215. doi: 10.5194/soil-7-193-2021. Publisher: Copernicus GmbH.

Helfenstein, A., Mulder, V. L., Hack-ten Broeke, M. J., van Doorn, M., Teuling, K., Walvoort, D. J., & Heuvelink, G. B. (2024a). BIS-4D: Maps of soil properties and their uncertainties at 25 m resolution in the Netherlands. doi: 10.4121/0C934AC6-2E95-4422-8360-D3A802766C71.V1.

Helfenstein, A., Mulder, V. L., Hack-ten Broeke, M. J., van Doorn, M., Teuling, K., Walvoort, D. J., & Heuvelink, G. B. (2024b). Spatially explicit environmental variables at 25m resolution for spatial modelling in the Netherlands. doi: 10.4121/6AF610ED-9006-4AC5-B399-4795C2AC01EC.V1.

Helfenstein, A., Sekulić, A., Mulder, V. L., & Heuvelink, G. B. (2022). Random Forest Spatial Interpolation for 3D digital soil mapping. In *Digital Soil Mapping: advances towards Digital Soil Assessment*. Glasgow, Scotland. 22nd World Congress of Soil Science (WCSS).

Helfenstein, A., Teuling, K., Walvoort, D. J., Hack-ten Broeke, M. J., Mulder, V. L., van Doorn, M., & Heuvelink, G. B. (2024c). Georeferenced point data of soil properties in the Netherlands. doi: 10.4121/C90215B3-BDC6-4633-B721-4C4A0259D6DC.V2.

Hempel, J., Libohova, Z., Thompson, J., Odgers, N., Smith, C., Lelyk, G., & Geraldo, G. (2014a). GlobalSoilMap North American Node progress. In D. Arrouays, N. McKenzie, J. Hempel, A. de Forges, & A. McBratney (Eds.), *GlobalSoilMap* (pp. 41–45). CRC Press. doi: 10.1201/b16500-11.

Hempel, J., McBratney, A., Arrouays, D., McKenzie, N., & Hartemink, A. (2014b). GlobalSoilMap project history. In D. Arrouays, N. McKenzie, J. Hempel, A. C. Richer de Forges, & A. McBratney (Eds.), *GlobalSoilMap: Basis of the global spatial soil information system*. Boca Raton: CRC Press Taylor & Francis Group.

Hengl, T., Heuvelink, G., Sanderman, J., & MacMillan, R. (2017a). Spatiotemporal models of global soil organic carbon stock to support land degradation assessments at regional and global scales: limitations, challenges and opportunities. In *Geophysical Research Abstracts* (p. 1). volume 19.

Hengl, T., Heuvelink, G. B. M., Kempen, B., Leenaars, J. G. B., Walsh, M. G., Shepherd, K. D., Sila, A., MacMillan, R. A., Mendes de Jesus, J., Tamene, L., & Tondoh, J. E. (2015). Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions. *PLOS ONE*, *10*, e0125814. doi: 10.1371/journal.pone.0125814.

Hengl, T., Heuvelink, G. B. M., & Stein, A. (2004). A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, *120*, 75–93. doi: 10.1016/j.geoderma.2003.08.018.

Hengl, T. et al. (2017b). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, *12*, e0169748. doi: 10.1371/journal.pone.0169748.

Hengl, T., Jesus, J. M. d., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B. M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J. G. B., Walsh, M. G., & Gonzalez, M. R. (2014). SoilGrids1km — Global Soil Information Based on Automated Mapping. *PLOS ONE*, *9*, e105992. doi: 10.1371/journal.pone.0105992.

Hengl, T., Leenaars, J. G. B., Shepherd, K. D., Walsh, M. G., Heuvelink, G. B. M., Mamo, T., Tilahun, H., Berkhout, E., Cooper, M., Fegraus, E., Wheeler, I., & Kwabena, N. A. (2017c). Soil nutrient maps of Sub-Saharan Africa: assessment of soil nutrient content at 250 m spatial resolution using machine learning. *Nutrient Cycling in Agroecosystems*, *109*, 77–102. doi: 10.1007/s10705-017-9870-x.

Hengl, T., & MacMillan, R. A. (2019). *Predictive Soil Mapping with R*. Wageningen, the Netherlands: OpenGeoHub foundation.

Hengl, T. et al. (2021). African soil properties and nutrients mapped at 30 m spatial resolution using two-scale ensemble machine learning. *Scientific Reports*, *11*, 6130. doi: 10.1038/s41598-021-85639-y. Number: 1 Publisher: Nature Publishing Group.

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, *6*, e5518. doi: 10.7717/peerj.5518.

Hessel, R., Stolte, J., & Riksen, M. J. P. M. (2011). Huidige maatregelen tegen water- en winderosie in Nederland. *Bodem*, *21*, 11–12. Publisher: Wolters Kluwer.

Heuvelink, G., Brus, D., De Vries, F., Kempen, B., Knotters, M., Vasat, R., & Walvoort, D. (2010). Implications of digital soil mapping for soil information systems. In *4th Global Workshop on Digital Soil Mapping* (p. 6). Rome, Italy. doi: https://edepot.wur.nl/160764.

Heuvelink, G. B. M. (1998). *Error Propagation in Environmental Modelling with GIS*. CRC Press. Google-Books-ID: C_XWjSsboeUC.

Heuvelink, G. B. M. (2014). Uncertainty quantification of GlobalSoilMap products. In *GlobalSoilMap: Basis of the global spatial soil information system* (pp. 335–340). CRC Press. Google-Books-ID: 3OPKBQAAQBAJ.

Heuvelink, G. B. M. (2018). Uncertainty and Uncertainty Propagation in Soil Mapping and Modelling. In A. B. McBratney, B. Minasny, & U. Stockmann (Eds.), *Pedometrics* Progress in Soil Science (pp. 439–461). Cham: Springer International Publishing. doi: 10.1007/978-3-319-63439-5_14.

Heuvelink, G. B. M., Angelini, M. E., Poggio, L., Bai, Z., Batjes, N. H., Bosch, R. v. d., Bossio, D., Estella, S., Lehmann, J., Olmedo, G. F., & Sanderman, J. (2020). Machine learning in space and

time for modelling soil organic carbon change. *European Journal of Soil Science*, (pp. 1–17). doi: 10.1111/ejss.12998. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejss.12998.

Hijmans, R. J. (2020). Geographic Data Analysis and Modeling: package 'raster'.

Hijmans, R. J. (2023). Spatial Data Analysis: package 'terra'.

Hinsberg, A. v., Egmond, P. v., Pouwels, R., Dirkx, G. H. P., & Breman, B. C. (2020). *Referentiescenario's Natuur : Tussenrapportage Natuurverkenning 2050*. Technical Report 3574, Planbureau voor de Leefomgeving (PBL) Den Haag. `https://library.wur.nl/WebQuery/wurpubs/569486`.

von Hippel, P. T. (2005). Mean, Median, and Skew: Correcting a Textbook Rule. *Journal of Statistics Education*, *13*. doi: 10.1080/10691898.2005.11910556. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10691898.2005.11910556.

Hoogland, T., van den Akker, J. J. H., & Brus, D. J. (2012). Modeling the subsidence of peat soils in the Dutch coastal area. *Geoderma*, *171-172*, 92–97. doi: 10.1016/j.geoderma.2011.02.013.

Hoogland, T., Knotters, M., Pleijter, M., & Walvoort, D. J. J. (2014). *Actualisatie van de grondwatertrappenkaart van holoceen Nederland: resultaten van het veldonderzoek*. Alterra-rapport 2612, Alterra Wageningen UR Wageningen. `https://edepot.wur.nl/339780` publisher: Alterra.

Huang, J., Hartemink, A. E., & Zhang, Y. (2019). Climate and Land-Use Change Effects on Soil Carbon Stocks over 150 Years in Wisconsin, USA. *Remote Sensing*, *11*, 1504. doi: 10.3390/rs11121504. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute.

Huang, J., Malone, B. P., Minasny, B., McBratney, A. B., & Triantafilis, J. (2017). Evaluating a Bayesian modelling approach (INLA-SPDE) for environmental mapping. *Science of The Total Environment*, *609*, 621–632. doi: 10.1016/j.scitotenv.2017.07.201.

Höper, H., Augustin, J., Cagampan, J. P., Drösler, M., Lundin, L., Moors, E. J., Vasander, H., Waddington, J. M., & Wilson, D. (2008). Restoration of peatlands and greenhouse gas balances. In *Peatlands and Climate Change* (pp. 182–210). Jyvaskyla: International Peat Society.

IenM, & TNO (2017). *Basisregistratie Ondergrond (BRO) Catalogus: Booronderzoek, Bodemkundige boormonsterbeschrijving*. Technical Report Versie 1.0, Ministrie van Infrastructuur en Milieu (IenM), Geologische Dienst Nederland (TNO) Den Haag. `https://basisregistratieondergrond.nl/inhoud-bro/registratieobjecten/bodem-grondonderzoek/booronderzoek-bhr/bodemkundig-booronderzoek-bhr/`.

IPBES (2019). *Summary for policymakers of the global assessment report on biodiversity and ecosystem services*. Technical Report, Zenodo. `https://zenodo.org/records/3553579`. doi: 10.5281/zenodo.3553579.

IUSS Working Group WRB (2014). *World Reference Base for Soil Resources 2014. International soil classification system for naming soils and creating legends for soil maps.*. World Soil Resources Reports No. 106. Rome: FAO. OCLC: 979061096.

Iwema, J. (2023). SciML - WUR Scientific Machine Learning Network.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics. New York, NY: Springer US. doi: 10.1007/978-1-0716-1418-1.

Janitza, S., Binder, H., & Boulesteix, A.-L. (2016). Pitfalls of hypothesis tests and model selection on bootstrap samples: Causes and consequences in biometrical applications. *Biometrical Journal*, *58*, 447–473. doi: https://doi.org/10.1002/bimj.201400246. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bimj.201400246.

Jenny, H. (1941). *Factors of Soil Formation: A System of Quantitative Pedology*. New York: McGraw-Hill.

Jones, M. J. (1973). The Organic Matter Content of the Savanna Soils of West Africa. *Journal of Soil Science*, *24*, 42–53. doi: 10.1111/j.1365-2389.1973.tb00740.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2389.1973.tb00740.x.

Jongmans, A. G., Berg, M. W. v. d., Sonneveld, M. P. W., Peek, G. J. W. C., & Saparoea, R. M. v. d.

B. v. (2013). *Landschappen van Nederland, geologie, bodem en landgebruik*. Wageningen Academic Publishers.

Joosten, H., & Clarke, D. (2002). *Wise use of mires and peatlands: background and principles including a framework for decision-making*. Totnes: Internat. Mire Conservation Group [u.a.].

Jukema, G., Ramaekers, P., & Berkhout, P. (2023). *De Nederlandse agrarische sector in internationaal verband: Editie 2023*. Wageningen Economic Research. doi: 10.18174/584222.

Jónsson, J. G., Davísdóttir, B., Jónsdóttir, E. M., Kristinsdóttir, S. M., & Ragnarsdóttir, K. V. (2016). Soil indicators for sustainable development: A transdisciplinary approach for indicator development using expert stakeholders. *Agriculture, Ecosystems & Environment*, *232*, 179–189. doi: 10.1016/j.agee.2016.08.009.

Keesstra, S., Nunes, J., Novara, A., Finger, D., Avelar, D., Kalantari, Z., & Cerdà, A. (2018). The superior effect of nature based solutions in land management for enhancing ecosystem services. *Science of The Total Environment*, *610-611*, 997–1009. doi: 10.1016/j.scitotenv.2017.08.077.

Keesstra, S. D. et al. (2016). The significance of soils and soil science towards realization of the United Nations Sustainable Development Goals. *SOIL*, *2*, 111–128. doi: 10.5194/soil-2-111-2016. Publisher: Copernicus GmbH.

Keesstra, S. D. et al. (2021). *Deliverable 2.4: Roadmap for the European Joint Programme SOIL*. WP2 D2.4, European Joint Project COFUND. `https://ejpsoil.eu/fileadmin/projects/ejpsoil/WP2/Deliverable_2.4_Roadmap_for_the_European_Joint_Programme_SOIL.pdf`.

van Kekem, A. J., Hoogland, T., & van der Horst, J. B. F. (2005). *Uitspoelingsgevoelige gronden op de kaart; werkwijze en resultaten*. Technical Report 1080, Alterra Wageningen. `https://library.wur.nl/WebQuery/wurpubs/338759` iSSN: 1566-7197.

Keller, A., Franzen, J., Knüsel, P., Papritz, A., & Zürrer, M. (2018). *Bodeninformations-Plattform Schweiz (BIP-CH)*. Thematische Synthese TS4 des Nationalen Forschungsprogramms «Nachhaltige Nutzung der Ressource Boden» (NFP 68), Schweizerischer Nationalfonds zur Förderung der wissenschaftlichen Forschung (SNF) Bern.

Kempen, B., Brus, D. J., & Heuvelink, G. B. M. (2012a). Soil type mapping using the generalised linear geostatistical model: A case study in a Dutch cultivated peatland. *Geoderma*, *189-190*, 540–553. doi: 10.1016/j.geoderma.2012.05.028.

Kempen, B., Brus, D. J., Heuvelink, G. B. M., & Stoorvogel, J. J. (2009). Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. *Geoderma*, *151*, 311–326. doi: 10.1016/j.geoderma.2009.04.023.

Kempen, B., Brus, D. J., & Stoorvogel, J. J. (2011). Three-dimensional mapping of soil organic matter content using soil type–specific depth functions. *Geoderma*, *162*, 107–123. doi: 10.1016/j.geoderma.2011.01.010.

Kempen, B., Brus, D. J., Stoorvogel, J. J., Heuvelink, G. B. M., & de Vries, F. (2012b). Efficiency Comparison of Conventional and Digital Soil Mapping for Updating Soil Maps. *Soil Science Society of America Journal*, *76*, 2097–2115. doi: 10.2136/sssaj2011.0424.

Kempen, B., Brus, D. J., & de Vries, F. (2015). Operationalizing digital soil mapping for nationwide updating of the 1:50,000 soil map of the Netherlands. *Geoderma*, *241-242*, 313–329. doi: 10.1016/j.geoderma.2014.11.030.

Kempen, B., Heuvelink, G. B. M., Brus, D., & Walvoort, D. (2014). Towards GlobalSoilMap.net products for The Netherlands. *GlobalSoilMap: Basis of the Global Spatial Soil Information System - Proceedings of the 1st GlobalSoilMap Conference*, (pp. 85–90). doi: 10.1201/b16500-19.

Keskin, H., & Grunwald, S. (2018). Regression kriging as a workhorse in the digital soil mapper's toolbox. *Geoderma*, *326*, 22–41. doi: 10.1016/j.geoderma.2018.04.004.

Keskin, H., Grunwald, S., & Harris, W. G. (2019). Digital mapping of soil carbon fractions with machine learning. *Geoderma*, *339*, 40–58. doi: 10.1016/j.geoderma.2018.12.037.

Khaledian, Y., & Miller, B. A. (2020). Selecting appropriate machine learning methods for digital soil

mapping. *Applied Mathematical Modelling*, *81*, 401–418. doi: 10.1016/j.apm.2019.12.016.

KNMI (2020). Koninklijk Nederlands Meteorologisch Instituut (KNMI) Dataplatform. Last Modified: 2021-03-12T09:43 Publisher: Koninklijk Nederlands Meteorologisch Instituut.

KNMI (2023). KNMI'23-klimaatscenario's voor Nederland.

Knotters, M., Broeke, M. J. D. H.-t., Hinssen, P. J. W., Kolk, J. W. H. v. d., & Okx, J. P. (2015a). *Betekenis van BRO/BIS Nederland voor WOT Natuur & Milieu: Een risicoanalyse*. WOt-interne notitie 121, WOT Natuur & Milieu Wageningen. `https://research.wur.nl/en/publications/betekenis-van-brobis-nederland-voor-wot-natuur-amp-milieu-een-ris` publisher: WOT Natuur & Milieu.

Knotters, M., Brus, D. J., & Oude Voshaar, J. H. (1995). A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. *Geoderma*, *67*, 227–246. doi: 10.1016/0016-7061(95)00011-C.

Knotters, M., Okx, J., Hack-ten Broeke, M., & de Vries, F. (2015b). Bodem in beweging: BIS NEderland informeert. *Bodem*, *25*, 11–13.

Knotters, M., Teuling, K., Reijneveld, A., Lesschen, J. P., & Kuikman, P. (2022). Changes in organic matter contents and carbon stocks in Dutch soils, 1998–2018. *Geoderma*, *414*, 115751. doi: 10.1016/j.geoderma.2022.115751.

Knotters, M., & Vroon, H. R. J. (2015). The economic value of detailed soil survey in a drinking water collection area in the Netherlands. *Geoderma Regional*, *5*, 44–53. doi: 10.1016/j.geodrs.2015.03.002.

Knotters, M., & Walvoort, D. (2020). Hoge resolutie, nauwkeurige kaarten? Last Modified: 2020-10-16.

Knotters, M., Walvoort, D. J. J., Brouwer, F., Stuyt, L. C. P. M., & Okx, J. P. (2018). Landsdekkende, actuele informatie over grondwatertrappen digitaal beschikbaar. *H2O online*, (pp. –).

Kombrink, H., Doornenbal, J. C., Duin, E. J. T., Dulk, M. d., Veen, J. H. t., & Witmans, N. (2012). New insights into the geological structure of the Netherlands; results of a detailed mapping project. *Netherlands Journal of Geosciences*, *91*, 419–446. doi: 10.1017/S0016774600000329. Publisher: Cambridge University Press.

Kooistra, L., & Kuikman, P. J. (2002). *Soil carbon sequestration in the Netherlands: inventory of long term experiments to validate effectiveness of soil carbon management in agriculture and land use change*. Technical Report 650, Alterra Wageningen. `https://library.wur.nl/WebQuery/wurpubs/320604`.

Koomen, A., & Maas, G. (2004). *Geomorfologische Kaart Nederland (GKN); Achtergronddocument bij het landsdekkende digitale bestand*. Altera-rapport 1039, Alterra Wageningen.

Kopittke, P. M., Menzies, N. W., Wang, P., McKenna, B. A., & Lombi, E. (2019). Soil and the intensification of agriculture for global food security. *Environment International*, *132*, 105078. doi: 10.1016/j.envint.2019.105078.

Kortleve, A. J., Mogollón, J. M., Heimovaara, T. J., & Gebert, J. (2023). Topsoil Carbon Stocks in Urban Greenspaces of The Hague, the Netherlands. *Urban Ecosystems*, *26*, 725–742. doi: 10.1007/s11252-022-01315-7.

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

Kramer, H., & Clement, J. (2015). *Basiskaart Natuur 2013: Een landsdekkend basisbestand voor de terrestrische natuur in Nederland*. WOt-technical report 41, Wettelijke Onderzoekstaken Natuur & Milieu Wageningen.

Kramer, H., van Dorland, G., & Gijsbertse, H. (2010). Historisch grondgebruik Nederland. In *Tijd en Ruimte. Nieuwe toepassingen van GIS in de alfawetenschappen* (pp. 142–153). Alterra Wageningen UR: Uitgeverij Matrijs.

Kroes, J., Van Dam, J., Bartholomeus, R., Groenendijk, P., Heinen, M., Hendriks, R., Mulder, H., Supit, I., & Van Walsum, P. (2017). *SWAP version 4; Theory description and user manual*. Report 2780, Wageningen Environmental Research Wageningen. `https://research.wur.nl/en/publications/`

9340d4ce-b26d-43b2-b20c-14b7f91c8f21. doi: 10.18174/416321.

KRW (2004). Kaderrichtlijn Water (KRW) Grote grondwaterlichamen 2004.

Kuhn, M. (2008). Building Predictive Models in R Using the Caret Package. *Journal of Statistical Software*, *28*. doi: 10.18637/jss.v028.i05.

Kuhn, M. (2019). The caret Package.

Kuhn, M. (2022). Classification and Regression Training: package 'caret'.

Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. New York, NY: Springer New York. doi: 10.1007/978-1-4614-6849-3.

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

Lagacherie, P., Arrouays, D., Bourennane, H., Gomez, C., Martin, M., & Saby, N. P. A. (2019). How far can the uncertainty on a Digital Soil Map be known?: A numerical experiment using pseudo values of clay content obtained from Vis-SWIR hyperspectral imagery. *Geoderma*, *337*, 1320–1328. doi: 10.1016/j.geoderma.2018.08.024.

Lagacherie, P., Arrouays, D., Bourennane, H., Gomez, C., & Nkuba-Kasanda, L. (2020). Analysing the impact of soil spatial sampling on the performances of Digital Soil Mapping models and their evaluation: A numerical experiment on Quantile Random Forest using clay contents obtained from Vis-NIR-SWIR hyperspectral imagery. *Geoderma*, *375*, 114503. doi: 10.1016/j.geoderma.2020.114503.

Lamigueiro, O. P., & Hijmans, R. J. (2023). Visualization Methods for Raster Data: package 'rasterVis'.

Lark, R. M., Ander, E. L., Cave, M. R., Knights, K. V., Glennon, M. M., & Scanlon, R. P. (2014). Mapping trace element deficiency by cokriging from regional geochemical soil data: A case study on cobalt for grazing sheep in Ireland. *Geoderma*, *226-227*, 64–78. doi: 10.1016/j.geoderma.2014.03.002.

Lark, R. M., & Bishop, T. F. A. (2007). Cokriging particle size fractions of the soil. *European Journal of Soil Science*, *58*, 763–774. doi: 10.1111/j.1365-2389.2006.00866.x.

Larrosa, C., Carrasco, L. R., & Milner-Gulland, E. J. (2016). Unintended Feedbacks: Challenges and Opportunities for Improving Conservation Effectiveness. *Conservation Letters*, *9*, 316–326. doi: 10.1111/conl.12240. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/conl.12240.

Le Rest, K., Pinaud, D., Monestiez, P., Chadoeuf, J., & Bretagnolle, V. (2014). Spatial leave-one-out cross-validation for variable selection in the presence of spatial autocorrelation. *Global Ecology and Biogeography*, *23*, 811–820. doi: 10.1111/geb.12161. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/geb.12161.

van Leeuwen, C. C. E., Mulder, V. L., Batjes, N. H., & Heuvelink, G. B. M. (2021). Statistical modelling of measurement error in wet chemistry soil data. *European Journal of Soil Science*, *n/a*, 1–17. doi: 10.1111/ejss.13137. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejss.13137.

Lehmann, J., Bossio, D. A., Kögel-Knabner, I., & Rillig, M. C. (2020). The concept and future prospects of soil health. *Nature Reviews Earth & Environment*, *1*, 544–553. doi: 10.1038/s43017-020-0080-8. Number: 10 Publisher: Nature Publishing Group.

Leifeld, J., & Menichetti, L. (2018). The underappreciated potential of peatlands in global climate change mitigation strategies. *Nature Communications*, *9*. doi: 10.1038/s41467-018-03406-6.

Lemercier, B., Lagacherie, P., Amelin, J., Sauter, J., Pichelin, P., Richer-de Forges, A. C., & Arrouays, D. (2022). Multiscale evaluations of global, national and regional digital soil mapping products in France. *Geoderma*, *425*, 116052. doi: 10.1016/j.geoderma.2022.116052.

Li, J., Heap, A. D., Potter, A., & Daniell, J. J. (2011). Application of machine learning methods to spatial interpolation of environmental variables. *Environmental Modelling & Software*, *26*, 1647–1659. doi: https://doi.org/10.1016/j.envsoft.2011.07.004.

Liang, Z., Chen, S., Yang, Y., Zhou, Y., & Shi, Z. (2019). High-resolution three-dimensional mapping of soil organic carbon in China: Effects of SoilGrids products on national modeling. *Science of The Total Environment*, *685*, 480–489. doi: 10.1016/j.scitotenv.2019.05.332.

Libohova, Z., Seybold, C., Adhikari, K., Wills, S., Beaudette, D., Peaslee, S., Lindbo, D., & Owens, P. R. (2019). The anatomy of uncertainty for soil pH measurements and predictions: Implications for modellers and practitioners. *European Journal of Soil Science*, *70*, 185–199. doi: 10.1111/ejss.12770. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejss.12770.

Lilburne, L., Heuvelink, G., & Helfenstein, A. (2023). Interpreting and evaluating digital soil mapping prediction uncertainties. In *EGU23-17024*. Vienna, Austria: Copernicus Meetings. doi: 10.5194/egusphere-egu23-17024 eGU General Assembly 2023.

Liu, F., Rossiter, D. G., Song, X.-D., Zhang, G.-L., Yang, R.-M., Zhao, Y.-G., Li, D.-C., & Ju, B. (2016). A similarity-based method for three-dimensional prediction of soil organic matter concentration. *Geoderma*, *263*, 254–263. doi: 10.1016/j.geoderma.2015.05.013.

Liu, F., Zhang, G.-L., Song, X., Li, D., Zhao, Y., Yang, J., Wu, H., & Yang, F. (2020). High-resolution and three-dimensional mapping of soil texture of China. *Geoderma*, *361*, 114061. doi: 10.1016/j.geoderma.2019.114061.

Liu, L., Ji, M., & Buchroithner, M. (2018). Transfer Learning for Soil Spectroscopy Based on Convolutional Neural Networks and Its Application in Soil Clay Content Mapping Using Hyperspectral Imagery. *Sensors*, *18*, 3169. doi: 10.3390/s18093169. Number: 9 Publisher: Multidisciplinary Digital Publishing Institute.

Lobsey, C. R., Viscarra Rossel, R. A., Roudier, P., & Hedley, C. B. (2017). rs-local data-mines information from spectral libraries to improve local calibrations. *European Journal of Soil Science*, *68*, 840–852. doi: 10.1111/ejss.12490.

Loiseau, T. et al. (2019). Satellite data integration for soil clay content modelling at a national scale. *International Journal of Applied Earth Observation and Geoinformation*, *82*, 101905. doi: 10.1016/j.jag.2019.101905.

Lookman, R., Vandeweert, N., Merckx, R., & Vlassak, K. (1995). Geostatistical assessment of the regional distribution of phosphate sorption capacity parameters (FeOX and AlOX) in northern Belgium. *Geoderma*, *66*, 285–296. doi: 10.1016/0016-7061(94)00084-N.

Lorenz, K., Lal, R., & Ehlers, K. (2019). Soil organic carbon stock as an indicator for monitoring land and soil degradation in relation to United Nations' Sustainable Development Goals. *Land Degradation & Development*, *30*, 824–838. doi: 10.1002/ldr.3270. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ldr.3270.

Lowndes, J. S. S., Best, B. D., Scarborough, C., Afflerbach, J. C., Frazier, M. R., O'Hara, C. C., Jiang, N., & Halpern, B. S. (2017). Our path to better science in less time using open data science tools. *Nature Ecology & Evolution*, *1*, 0160. doi: 10.1038/s41559-017-0160.

Ma, Y., Minasny, B., McBratney, A., Poggio, L., & Fajardo, M. (2021). Predicting soil properties in 3D: Should depth be a covariate? *Geoderma*, *383*, 114794. doi: 10.1016/j.geoderma.2020.114794.

Maas, G., van der Meij, M., Delft, S., & Heidema, A. (2019). *Toelichting bij de legenda Geomorfologische kaart van Nederland 1:50 000 (2019)*. Wageningen: Wageningen Environmental Research.

Mallavan, B., Minasny, B., & McBratney, A. (2010). Homosoil, a Methodology for Quantitative Extrapolation of Soil Information Across the Globe. In J. L. Boettinger, D. W. Howell, A. C. Moore, A. E. Hartemink, & S. Kienast-Brown (Eds.), *Digital Soil Mapping: Bridging Research, Environmental Application, and Operation* Progress in Soil Science (pp. 137–150). Dordrecht: Springer Netherlands. doi: 10.1007/978-90-481-8863-5_12.

Malone, B., Searle, R., Malone, B., & Searle, R. (2021). Updating the Australian digital soil texture mapping (Part 2*): spatial modelling of merged field and lab measurements. *Soil Research*, *59*, 435–451. doi: 10.1071/SR20284. Publisher: CSIRO PUBLISHING.

Malone, B. P., de Gruijter, J. J., McBratney, A. B., Minasny, B., & Brus, D. J. (2011). Using Additional Criteria for Measuring the Quality of Predictions and Their Uncertainties in a Digital Soil Mapping Framework. *Soil Science Society of America Journal*, *75*, 1032–1043. doi: 10.2136/sssaj2010.0280. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2136/sssaj2010.0280.

Malone, B. P., McBratney, A. B., & Minasny, B. (2013). Spatial Scaling for Digital Soil Mapping. *Soil Science Society of America Journal*, *77*, 890–902. doi: 10.2136/sssaj2012.0419. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2136/sssaj2012.0419.

Malone, B. P., Minasny, B., & McBratney, A. B. (2017). *Using R for Digital Soil Mapping*. Progress in Soil Science. Cham: Springer International Publishing. doi: 10.1007/978-3-319-44327-0.

Maring, L., Vries, F. d., Brouwer, F., Groot, H., Kiden, P., Leeters, E. E. J. M., & Mol, G. (2009). *IMBOD deelactiviteit 5 : inhoudelijke afstemming*. Alterra-rapport 1817, Alterra Wageningen. `https://research.wur.nl/en/publications/imbod-deelactiviteit-5-inhoudelijke-afstemming` publisher: Alterra.

Martín-Antón, M., Negro, V., del Campo, J. M., López-Gutiérrez, J. S., & Esteban, M. D. (2016). Review of coastal Land Reclamation situation in the World. *Journal of Coastal Research*, (pp. 667–671). doi: 10.2112/SI75-133.1.

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

McBratney, A. B., Minasny, B., & Stockmann, U. (Eds.) (2018). *Pedometrics*. Progress in Soil Science. Cham: Springer International Publishing. doi: 10.1007/978-3-319-63439-5.

McBratney, A. B., & Odeh, I. O. A. (1997). Application of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions. *Geoderma*, *77*, 85–113. doi: 10.1016/S0016-7061(97)00017-7.

McKeague, J. A. (1967). An evaluation of 0.1m pyrophosphate and pyrophosphate-dithionite in comparison with oxalate as extractants of the accumulation products in podzols and some other soils. *Canadian Journal of Soil Science*, *47*, 95–99. doi: 10.4141/cjss67-017. Publisher: NRC Research Press.

McKeague, J. A., Brydon, J. E., & Miles, N. M. (1971). Differentiation of Forms of Extractable Iron and Aluminum in Soils. *Soil Science Society of America Journal*, *35*, 33–38. doi: 10.2136/sssaj1971.03615995003500010016x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2136/sssaj1971.03615995003500010016x.

Meersmans, J., Arrouays, D., Van Rompaey, A., Pagé, C., De Baets, S., & Quine, T. (2016). Future C loss in mid-latitude mineral soils: Climate change exceeds land use mitigation potential in France. *Scientific Reports*, *6*. doi: 10.1038/srep35798.

Meersmans, J., Van Wesemael, B., De Ridder, F., Fallas Dotti, M., De Baets, S., & Van Molle, M. (2009a). Changes in organic carbon distribution with depth in agricultural soils in northern Belgium, 1960–2006. *Global Change Biology*, *15*, 2739–2750. doi: 10.1111/j.1365-2486.2009.01855.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2486.2009.01855.x.

Meersmans, J., Van WESEMAEL, B., Goidts, E., Van MOLLE, M., De BAETS, S., & De RIDDER, F. (2011). Spatial analysis of soil organic carbon evolution in Belgian croplands and grasslands, 1960–2006. *Global Change Biology*, *17*, 466–479. doi: 10.1111/j.1365-2486.2010.02183.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2486.2010.02183.x.

Meersmans, J., van Wesemael, B., De Ridder, F., & Van Molle, M. (2009b). Modelling the three-dimensional spatial distribution of soil organic carbon (SOC) at the regional scale (Flanders, Belgium). *Geoderma*, *152*, 43–52. doi: 10.1016/j.geoderma.2009.05.015.

Meinshausen, N. (2006). Quantile Regression Forests. *Journal of Machine Learning Research*, *7*, 17.

van der Meulen, M. et al. (2013). 3D geology in a 2D country: perspectives for geological surveying in the Netherlands. *Netherlands Journal of Geosciences - Geologie en Mijnbouw*, *92*, 217–241. doi: 10.1017/S0016774600000184.

Meyer, H. (2023). 'caret' Applications for Spatial-Temporal Models: package 'CAST'.

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., & Nauss, T. (2018). Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environmental Modelling & Software*, *101*, 1–9. doi: 10.1016/j.envsoft.2017.12.001.

Miller, B. A., Koszinski, S., Wehrhan, M., & Sommer, M. (2015). Impact of multi-scale predictor selection for modeling soil properties. *Geoderma*, *239-240*, 97–106. doi: 10.1016/j.geoderma.2014.09.018.

Miller, R. O., & Kissel, D. E. (2010). Comparison of Soil pH Methods on Soils of North America. *Soil Science Society of America Journal*, *74*, 310–316. doi: 10.2136/sssaj2008.0047. _eprint: https://acsess.onlinelibrary.wiley.com/doi/pdf/10.2136/sssaj2008.0047.

Minasny, B., Hong, S. Y., Hartemink, A. E., Kim, Y. H., & Kang, S. S. (2016). Soil pH increase under paddy in South Korea between 2000 and 2012. *Agriculture, Ecosystems & Environment*, *221*, 205–213. doi: 10.1016/j.agee.2016.01.042.

Minasny, B. et al. (2017). Soil carbon 4 per mille. *Geoderma*, *292*, 59–86. doi: 10.1016/j.geoderma.2017.01.002.

Moinet, G. Y. K., Hijbeek, R., van Vuuren, D. P., & Giller, K. E. (2023). Carbon for soils, not soils for carbon. *Global Change Biology*, *29*, 2384–2398. doi: 10.1111/gcb.16570. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.16570.

Mukhortova, L., Schepaschenko, D., Moltchanova, E., Shvidenko, A., Khabarov, N., & See, L. (2021). Respiration of Russian soils: Climatic drivers and response to climate change. *Science of The Total Environment*, *785*, 147314. doi: 10.1016/j.scitotenv.2021.147314.

Mulder, M., Walvoort, D., Brouwer, F., Tol-Leenders, D. v., & Verzandvoort, S. (2022). *Bodemgeschiktheidskaarten voor landbouw in de provincie Noord-Brabant: Een toepassing van Waterwijzer Landbouw*. Rapport 3206, Wageningen Environmental Research Wageningen, the Netherlands. `https://research.wur.nl/en/publications/bodemgeschiktheidskaarten-voor-landbouw-in-de-provincie-noord-bra` publisher: Wageningen Environmental Research.

Mulder, V. L., Lacoste, M., Richer-de Forges, A. C., & Arrouays, D. (2016a). GlobalSoilMap France: High-resolution spatial modelling the soils of France up to two meter depth. *Science of The Total Environment*, *573*, 1352–1369. doi: 10.1016/j.scitotenv.2016.07.066. Citation Key Alias: mulderGlobalSoilMapFranceHighresolution2016a.

Mulder, V. L., Lacoste, M., Richer-de Forges, A. C., Martin, M. P., & Arrouays, D. (2016b). National versus global modelling the 3D distribution of soil organic carbon in mainland France. *Geoderma*, *263*, 16–34. doi: 10.1016/j.geoderma.2015.08.035. Citation Key Alias: mulderNationalGlobalModelling2016a.

Murdiyarso, D., Hergoualc'h, K., & Verchot, L. V. (2010). Opportunities for reducing greenhouse gas emissions in tropical peatlands. *Proceedings of the National Academy of Sciences*, *107*, 19655–19660. doi: 10.1073/pnas.0911966107. Publisher: Proceedings of the National Academy of Sciences.

Møller, A. B., Beucher, A. M., Pouladi, N., & Greve, M. H. (2020). Oblique geographic coordinates as covariates for digital soil mapping. *SOIL*, *6*, 269–289. doi: https://doi.org/10.5194/soil-6-269-2020. Publisher: Copernicus GmbH.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, *10*, 282–290. doi: 10.1016/0022-1694(70)90255-6.

Nauman, T. W., & Duniway, M. C. (2019). Relative prediction intervals reveal larger uncertainty in 3D approaches to predictive digital soil mapping of soil properties with legacy data. *Geoderma*, *347*, 170–184. doi: 10.1016/j.geoderma.2019.03.037.

Negassa, W., Acksel, A., Eckhardt, K.-U., Regier, T., & Leinweber, P. (2019). Soil organic matter characteristics in drained and rewetted peatlands of northern Germany: Chemical and spectroscopic analyses. *Geoderma*, *353*, 468–481. doi: 10.1016/j.geoderma.2019.07.002.

NEN 5753 (2020). *Soil – Determination of clay content and particle size distribution in soil and sediment by sieve and pipet*. Standards ICS codes: 13.080.20 (2018th ed.). Available at www.nen.nl (in Dutch).: Anorganische parameters (Milieukwaliteit).

Nenkam, A. M., Wadoux, A. M. J.-C., Minasny, B., McBratney, A. B., Traore, P. C. S., Falconnier, G. N., & Whitbread, A. M. (2022). Using homosoils for quantitative extrapolation of soil mapping models. *European Journal of Soil Science*, *73*, e13285. doi: 10.1111/ejss.13285. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejss.13285.

Neyroud, J.-A., & Lischer, P. (2003). Do different methods used to estimate soil phosphorus availability across Europe give comparable results? *Journal of Plant Nutrition and Soil Science*, *166*, 422–431. doi: 10.1002/jpln.200321152. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jpln.200321152.

NHI (2023). Nederlands Hydrologisch Instrumentarium (NHI).

Noellemeyer, E., & Six, J. (2015). Basic Principles of Soil Carbon Management for Multiple Ecosystem Benefits. In *Soil carbon: science, management, and policy for multiple benefits* number volume 71 in SCOPE series. Wallingford, Oxfordshire, UK ; Boston, MA, USA: CAB International.

Nol, L., Heuvelink, G. B. M., Veldkamp, A., de Vries, W., & Kros, J. (2010). Uncertainty propagation analysis of an N2O emission model at the plot and landscape scale. *Geoderma*, *159*, 9–23. doi: 10.1016/j.geoderma.2010.06.009.

Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., Schaepman, M. E., & Papritz, A. (2018). Evaluation of digital soil mapping approaches with large sets of environmental covariates. *SOIL*, *4*, 1–22. doi: 10.5194/soil-4-1-2018.

Nussbaum, M., Walthert, L., Fraefel, M., Greiner, L., & Papritz, A. (2017). Mapping of soil properties at high resolution in Switzerland using boosted geoadditive models. *SOIL*, *3*, 191–210. doi: 10.5194/soil-3-191-2017. Citation Key Alias: nussbaumMappingSoilProperties2017a.

Nussbaum, M., Zimmermann, S., Walthert, L., & Baltensweiler, A. (2023). Benefits of hierarchical predictions for digital soil mapping—An approach to map bimodal soil pH. *Geoderma*, *437*, 116579. doi: 10.1016/j.geoderma.2023.116579.

Odeh, I. O. A., McBratney, A. B., & Chittleborough, D. J. (1994). Spatial prediction of soil properties from landform attributes derived from a digital elevation model. *Geoderma*, *63*, 197–214. doi: 10.1016/0016-7061(94)90063-9.

Odeh, I. O. A., McBratney, A. B., & Chittleborough, D. J. (1995). Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. *Geoderma*, *67*, 215–226. doi: 10.1016/0016-7061(95)00007-B.

Odeh, I. O. A., Todd, A. J., & Triantafilis, J. (2003). Spatial prediction of soil particle-size fractions as compositional data. *Soil Science*, *168*, 501. doi: 10.1097/01.ss.0000080335.10341.23.

OpenGeoHub, Hengl, T., Parente, L., Wheeler, I., & Bonannella, C. (2022). Dynamic soil information at farm scale based on Machine Learning and EO data: building an Open Soil Data Cube for Europe.

OpenGeoHub, Parente, L., Witjes, M., Hengl, T., Ilie, C. M., & Landa, M. (2021). Europe From Above: Space-Time Machine Learning Reveals Our Changing Environment.

Orton, T. G., Pringle, M. J., & Bishop, T. F. A. (2016). A one-step approach for modelling and mapping soil properties based on profile data sampled over varying depth intervals. *Geoderma*, *262*, 174–186. doi: 10.1016/j.geoderma.2015.08.013.

Orton, T. G., Pringle, M. J., Bishop, T. F. A., Menzies, N. W., & Dang, Y. P. (2020). Increment-averaged kriging for 3-D modelling and mapping soil properties: Combining machine learning and geostatistical methods. *Geoderma*, *361*, 114094. doi: 10.1016/j.geoderma.2019.114094.

Ottoy, S., De Vos, B., Sindayihebura, A., Hermy, M., & Van Orshoven, J. (2017). Assessing soil organic carbon stocks under current and potential forest cover using digital soil mapping and spatial generalisation. *Ecological Indicators*, *77*, 139–150. doi: 10.1016/j.ecolind.2017.02.010.

Padarian, J., Minasny, B., & McBratney, A. B. (2017). Chile and the Chilean soil grid: A contribution to GlobalSoilMap. *Geoderma Regional*, *9*, 17–28. doi: 10.1016/j.geodrs.2016.12.001.

Padarian, J., Minasny, B., & McBratney, A. B. (2019a). Transfer learning to localise a continental soil vis-NIR calibration model. *Geoderma*, *340*, 279–288. doi: 10.1016/j.geoderma.2019.01.009.

Padarian, J., Minasny, B., & McBratney, A. B. (2019b). Using deep learning to predict soil properties from regional spectral data. *Geoderma Regional*, *16*, e00198. doi: 10.1016/j.geodrs.2018.e00198.

Pahlavan-Rad, M. R., & Akbarimoghaddam, A. (2018). Spatial variability of soil texture fractions and pH in a flood plain (case study from eastern Iran). *CATENA*, *160*, 275–281. doi: 10.1016/j.catena.2017.10.002.

Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*, 1345–1359. doi: 10.1109/TKDE.2009.191.

Panagos, P., Köningner, J., Ballabio, C., Liakos, L., Muntwyler, A., Borrelli, P., & Lugato, E. (2022a). Improving the phosphorus budget of European agricultural soils. *Science of The Total Environment*, *853*, 158706. doi: 10.1016/j.scitotenv.2022.158706.

Panagos, P., Montanarella, L., Barbero, M., Schneegans, A., Aguglia, L., & Jones, A. (2022b). Soil priorities in the European Union. *Geoderma Regional*, *29*, e00510. doi: 10.1016/j.geodrs.2022.e00510.

Panagos, P., Van Liedekerke, M., Borrelli, P., Köninger, J., Ballabio, C., Orgiazzi, A., Lugato, E., Liakos, L., Hervas, J., Jones, A., & Montanarella, L. (2022c). European Soil Data Centre 2.0: Soil data and knowledge in support of the EU policies. *European Journal of Soil Science*, *73*, e13315. doi: 10.1111/ejss.13315. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejss.13315.

Papadopoulos, G., Edwards, P., & Murray, A. (2001). Confidence estimation methods for neural networks: a practical comparison. *IEEE Transactions on Neural Networks*, *12*, 1278–1287. doi: 10.1109/72.963764. Conference Name: IEEE Transactions on Neural Networks.

Parton, W. J., Schimel, D. S., Cole, C. V., & Ojima, D. S. (1987). Analysis of Factors Controlling Soil Organic Matter Levels in Great Plains Grasslands 1. *Soil Science Society of America Journal*, *51*, 1173–1179. doi: 10.2136/sssaj1987.03615995005100050015x.

Paul, C., Bartkowski, B., Dönmez, C., Don, A., Mayer, S., Steffens, M., Weigl, S., Wiesmeier, M., Wolf, A., & Helming, K. (2023). Carbon farming: Are soil carbon certificates a suitable tool for climate change mitigation? *Journal of Environmental Management*, *330*, 117142. doi: 10.1016/j.jenvman.2022.117142.

Pawlowsky-Glahn, V., & Buccianti, A. (2011). *Compositional Data Analysis: Theory and Applications*. West Sussex: John Wiley & Sons, Ltd. doi: 10.1002/9781119976462.

Pawlowsky-Glahn, V., Egozcue, J. J., & Tolosana-Delgado, R. (2015). *Modeling and Analysis of Compositional Data*. West Sussex: John Wiley & Sons. Google-Books-ID: KS0IBgAAQBAJ.

Piikki, K., Wetterlind, J., Söderström, M., & Stenberg, B. (2021). Perspectives on validation in digital soil mapping of continuous attributes—A review. *Soil Use and Management*, *37*, 7–21. doi: 10.1111/sum.12694. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/sum.12694.

Ploton, P. et al. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature Communications*, *11*, 4540. doi: 10.1038/s41467-020-18321-y. Number: 1 Publisher: Nature Publishing Group.

Poggio, L., & Gimona, A. (2014). National scale 3D modelling of soil organic carbon stocks with uncertainty propagation — An example from Scotland. *Geoderma*, *232-234*, 284–299. doi: 10.1016/j.geoderma.2014.05.004.

Poggio, L., & Gimona, A. (2017a). 3D mapping of soil texture in Scotland. *Geoderma Regional*, *9*, 5–16. doi: 10.1016/j.geodrs.2016.11.003. Citation Key Alias: poggio3DMappingSoil2017a.

Poggio, L., & Gimona, A. (2017b). Assimilation of optical and radar remote sensing data in 3D mapping of soil properties over large areas. *Science of The Total Environment*, *579*, 1094–1110. doi: 10.1016/j.scitotenv.2016.11.078. Citation Key Alias: poggioAssimilationOpticalRadar2017a.

Poggio, L., Gimona, A., Spezia, L., & Brewer, M. J. (2016). Bayesian spatial modelling of soil properties and their uncertainty: The example of soil organic matter in Scotland using R-INLA. *Geoderma*, *277*, 69–82. doi: 10.1016/j.geoderma.2016.04.026.

Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., & Rossiter, D. (2021). SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *SOIL*, *7*, 217–240. doi: 10.5194/soil-7-217-2021. Publisher: Copernicus GmbH.

Pohjankukka, J., Pahikkala, T., Nevalainen, P., & Heikkonen, J. (2017). Estimating the prediction performance of spatial models via spatial k-fold cross validation. *International Journal of Geographical Information Science*, *31*, 2001–2019. doi: 10.1080/13658816.2017.1346255. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/13658816.2017.1346255.

Pörtner, H.-O. et al. (2021). *Scientific outcome of the IPBES-IPCC co-sponsored workshop on biodiversity and climate change*. Technical Report, Zenodo. https://zenodo.org/records/5101125. doi: 10.5281/zenodo.5101125.

QGIS Development Team (2023). QGIS: A Free and Open Source Geographic Information System.

R Core Team (2023). *R: A language and environment for statistical computing*. Technical Report, R Foundation for Statistical Computing Vienna, Austria. https://www.R-project.org/.

Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., & Thompson, J. (2018). Soil Property and Class Maps of the Conterminous United States at 100-Meter Spatial Resolution. *Soil Science Society of America Journal*, *82*, 186–201. doi: https://doi.org/10.2136/sssaj2017.04.0122. _eprint: https://acsess.onlinelibrary.wiley.com/doi/pdf/10.2136/sssaj2017.04.0122.

Reijneveld, A., van Wensem, J., & Oenema, O. (2009). Soil organic carbon contents of agricultural land in the Netherlands between 1984 and 2004. *Geoderma*, *152*, 231–238. doi: 10.1016/j.geoderma.2009.06.007.

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

Reyes Rojas, L. A., Adhikari, K., & Ventura, S. J. (2018). Projecting Soil Organic Carbon Distribution in Central Chile under Future Climate Scenarios. *Journal of Environmental Quality*, *47*, 735–745. doi: 10.2134/jeq2017.08.0329. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2134/jeq2017.08.0329.

RIVM (2020). Grootschalige Concentratie- en Depositiekaarten Nederland (GCN, GDN), Rijksinstituut voor Volksgezondheid en Milieu (RIVM).

Roberts, D. R. et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*, 913–929. doi: 10.1111/ecog.02881. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ecog.02881.

Robinson, N. J., Benke, K. K., Norng, S., Kitching, M., & Crawford, D. M. (2017). Improving the information content in soil pH maps: a case study. *European Journal of Soil Science*, *68*, 592–604. doi: 10.1111/ejss.12452. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejss.12452.

Roerink, G., & Mücher, S. (2023). Groenmonitor.

Román Dobarco, M., Arrouays, D., Lagacherie, P., Ciampalini, R., & Saby, N. P. A. (2017). Prediction of topsoil texture for Region Centre (France) applying model ensemble methods. *Geoderma*, *298*, 67–77. doi: 10.1016/j.geoderma.2017.03.015.

Roosevelt, F. D. (1937). Letter to all State Governors on a Uniform Soil Conservation Law.

Ros, G. H. (2023). Open Bodemindex (OBI): Eenvoudig en betaalbaar inzicht in bodemkwaliteit en bodemverbetering.

Ros, G. H., Haan, J. J. d., Fuchs, L. M., & Molendijk, L. (2023). *Bodembeoordeling van landbouwgronden voor diverse ecosysteemdiensten : ontwikkeling van de BLN, versie 2.0*. Rapport / Wageningen University & Research, Business unit Open Teelten WPR-OT-1030, Wageningen Plant Research Wageningen. https://doi.org/10.18174/634579.

Ros, G. H., Verweij, S. E., Janssen, S. J. C., De Haan, J., & Fujita, Y. (2022). An Open Soil Health Assessment Framework Facilitating Sustainable Soil Management. *Environmental Science & Technology*, *56*, 17375–17384. doi: 10.1021/acs.est.2c04516. Publisher: American Chemical Society.

Roudier, P., Burge, O. R., Richardson, S. J., McCarthy, J. K., Grealish, G. J., & Ausseil, A.-G. (2020). National Scale 3D Mapping of Soil pH Using a Data Augmentation Approach. *Remote Sensing*, *12*, 2872. doi: 10.3390/rs12182872. Number: 18 Publisher: Multidisciplinary Digital Publishing Institute.

Römkens, P. F. a. M., & Oenema, O. (2004). *Quick scan soils in the Netherlands; overview of the soil status with reference to the forthcoming EU soil strategy*. Alterra-rapport 948, Alterra Wageningen. https://edepot.wur.nl/120941 iSSN (Print): 1566-7197.

Sanderman, J., Hengl, T., & Fiske, G. J. (2017). Soil carbon debt of 12,000 years of human land use.

Proceedings of the National Academy of Sciences, *114*, 9575–9580. doi: 10.1073/pnas.1706103114. Publisher: Proceedings of the National Academy of Sciences.

Sanders, M. E., & Prins, A. H. (2001). Provinciaal natuurbeleid: kwaliteitsdoelen voor de Ecologische Hoofdstructuur.

Sandri, M., & Zuccolotto, P. (2008). A Bias Correction Algorithm for the Gini Variable Importance Measure in Classification Trees. *Journal of Computational and Graphical Statistics*, *17*, 611–628. doi: 10.1198/106186008X344522. Publisher: Taylor & Francis _eprint: https://doi.org/10.1198/106186008X344522.

Sandri, M., & Zuccolotto, P. (2010). Analysis and correction of bias in Total Decrease in Node Impurity measures for tree-based algorithms. *Statistics and Computing*, *20*, 393–407. doi: 10.1007/s11222-009-9132-0.

Schillaci, C., Acutis, M., Lombardo, L., Lipani, A., Fantappiè, M., Märker, M., & Saia, S. (2017). Spatio-temporal topsoil organic carbon mapping of a semi-arid Mediterranean region: The role of land use, soil texture, topographic indices and the influence of remote sensing data to modelling. *Science of The Total Environment*, *601-602*, 821–832. doi: 10.1016/j.scitotenv.2017.05.239.

Schmidinger, J., & Heuvelink, G. B. M. (2023). Validation of uncertainty predictions in digital soil mapping. *Geoderma*, *437*, 116585. doi: 10.1016/j.geoderma.2023.116585.

Schoumans, O. F. (2013). Description of the Phosphorus Sorption and Desorption Processes in Lowland Peaty Clay Soils. *Soil Science*, *178*, 291. doi: 10.1097/SS.0b013e31829ef054.

Schoumans, O. F., & Chardon, W. J. (2015). Phosphate saturation degree and accumulation of phosphate in various soil types in The Netherlands. *Geoderma*, *237-238*, 325–335. doi: 10.1016/j.geoderma.2014.08.015.

Schratz, P. (2019). Handling of Spatial Data.

Schrumpf, M., Schulze, E. D., Kaiser, K., & Schumacher, J. (2011). How accurately can soil organic carbon stocks and stock changes be quantified by soil inventories? *Biogeosciences*, *8*, 1193–1212. doi: 10.5194/bg-8-1193-2011. Publisher: Copernicus GmbH.

Schulte, R. P. O. et al. (2015). Making the Most of Our Land: Managing Soil Functions from Local to Continental Scale. *Frontiers in Environmental Science*, *3*.

Scull, P., Franklin, J., Chadwick, O. A., & McArthur, D. (2003). Predictive soil mapping: a review. *Progress in Physical Geography: Earth and Environment*, *27*, 171–197. doi: 10.1191/0309133303pp366ra.

Seidel, M., Hutengs, C., Ludwig, B., Thiele-Bruhn, S., & Vohland, M. (2019). Strategies for the efficient estimation of soil organic carbon at the field scale with vis-NIR spectroscopy: Spectral libraries and spiking vs. local calibrations. *Geoderma*, *354*, 113856. doi: 10.1016/j.geoderma.2019.07.014.

Sekulić, A., Kilibarda, M., Heuvelink, G. B. M., Nikolić, M., & Bajat, B. (2020). Random Forest Spatial Interpolation. *Remote Sensing*, *12*, 1687. doi: 10.3390/rs12101687. Number: 10 Publisher: Multidisciplinary Digital Publishing Institute.

Sharma, M., Kaushal, R., Kaushik, P., & Ramakrishna, S. (2021). Carbon Farming: Prospects and Challenges. *Sustainability*, *13*, 11122. doi: 10.3390/su131911122. Number: 19 Publisher: Multidisciplinary Digital Publishing Institute.

Shepherd, K. D., Ferguson, R., Hoover, D., van Egmond, F., Sanderman, J., & Ge, Y. (2022). A global soil spectral calibration library and estimation service. *Soil Security*, *7*, 100061. doi: 10.1016/j.soisec.2022.100061.

Slier, T., Mi-Gegotek, Y., & Lesschen, J. (2023). *Potentie voor koolstofvastlegging in minerale landbouwbodems in de Nederlandse provincies*. Slim Landgebruik, Wageningen Environmental Research Wageningen. `https://slimlandgebruik.nl/publicaties`.

Smith, P. et al. (2021). Soil-derived Nature's Contributions to People and their contribution to the UN Sustainable Development Goals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*, 20200185. doi: 10.1098/rstb.2020.0185. Publisher: Royal Society.

Smith, P. et al. (2020). How to measure, report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric greenhouse gas removal. *Global Change Biology*, *26*, 219–241. doi: 10.1111/gcb.14815. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.14815.

Soil Science Society of America (2008). *Glossary of Soil Science Terms*. Madison, WI: Soil Science Society of America. doi: 10.2136/2008.glossarysoilscienceterms.

Song, X.-D., Yang, F., Ju, B., Li, D.-C., Zhao, Y.-G., Yang, J.-L., & Zhang, G.-L. (2018). The influence of the conversion of grassland to cropland on changes in soil organic carbon and total nitrogen stocks in the Songnen Plain of Northeast China. *CATENA*, *171*, 588–601. doi: 10.1016/j.catena.2018.07.045.

Sowińska-Świerkosz, B., & García, J. (2022). What are Nature-based solutions (NBS)? Setting core ideas for concept clarification. *Nature-Based Solutions*, *2*, 100009. doi: 10.1016/j.nbsj.2022.100009.

Steinbuch, L., Brus, D. J., & Heuvelink, G. B. M. (2018). Mapping the probability of ripened subsoils using Bayesian logistic regression with informative priors. *Geoderma*, *316*, 56–69. doi: 10.1016/j.geoderma.2017.12.010.

Stettler, M., Keller, T., Weisskopf, P., Lamandé, M., Lassen, P., Schjønning, P., & others (2014). Terranimo®–a web-based tool for evaluating soil compaction. *Landtechnik*, *69*, 132–138. Publisher: Kuratorium für Technik und Bauwesen in der Landwirtschaft eV (KTBL).

Stivrins, N., Ozola, I., & Gałka, M. (2017). Drivers of peat accumulation rate in a raised bog: impact of drainage, climate, and local vegetation composition. *Mires and Peat*, (pp. 1–19). doi: 10.19189/MaP.2016.OMB.262.

Stockmann, U., Padarian, J., McBratney, A., Minasny, B., de Brogniez, D., Montanarella, L., Hong, S. Y., Rawlins, B. G., & Field, D. J. (2015). Global soil organic carbon assessment. *Global Food Security*, *6*, 9–16. doi: 10.1016/j.gfs.2015.07.001.

Stokstad, E. (2019). Nitrogen crisis threatens Dutch environment—and economy. *Science*, *366*, 1180–1181. doi: 10.1126/science.366.6470.1180. Publisher: American Association for the Advancement of Science.

Stoorvogel, J. J., Bakkenes, M., Temme, A. J. A. M., Batjes, N. H., & Brink, B. J. E. t. (2017). S-World: A Global Soil Map for Environmental Modelling. *Land Degradation & Development*, *28*, 22–33. doi: 10.1002/ldr.2656.

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*, 25. doi: 10.1186/1471-2105-8-25.

Stumpf, F., Keller, A., Schmidt, K., Mayr, A., Gubler, A., & Schaepman, M. (2018). Spatio-temporal land use dynamics and soil organic carbon in Swiss agroecosystems. *Agriculture, Ecosystems & Environment*, *258*, 129–142. doi: 10.1016/j.agee.2018.02.012.

Subburayalu, S. K., & Slater, B. K. (2013). Soil Series Mapping By Knowledge Discovery from an Ohio County Soil Map. *Soil Science Society of America Journal*, *77*, 1254–1268. doi: 10.2136/sssaj2012.0321. _eprint: https://acsess.onlinelibrary.wiley.com/doi/pdf/10.2136/sssaj2012.0321.

Sun, X.-L., Minasny, B., Wang, H.-L., Zhao, Y.-G., Zhang, G.-L., & Wu, Y.-J. (2021). Spatiotemporal modelling of soil organic matter changes in Jiangsu, China between 1980 and 2006 using INLA-SPDE. *Geoderma*, *384*, 114808. doi: 10.1016/j.geoderma.2020.114808.

Sun, X.-L., Zhao, Y.-G., Wu, Y.-J., Zhao, M.-S., Wang, H.-L., & Zhang, G.-L. (2012). Spatio-temporal change of soil organic matter content of Jiangsu Province, China, based on digital soil maps. *Soil Use and Management*, *28*, 318–328. doi: 10.1111/j.1475-2743.2012.00421.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1475-2743.2012.00421.x.

Syvitski, J. P. M., Kettner, A. J., Overeem, I., Hutton, E. W. H., Hannon, M. T., Brakenridge, G. R., Day, J., Vörösmarty, C., Saito, Y., Giosan, L., & Nicholls, R. J. (2009). Sinking deltas due to human activities. *Nature Geoscience*, *2*, 681–686. doi: 10.1038/ngeo629. Number: 10 Publisher: Nature Publishing Group.

Szatmári, G., Pirkó, B., Koós, S., Laborczi, A., Bakacsi, Z., Szabó, J., & Pásztor, L. (2019). Spatio-

temporal assessment of topsoil organic carbon stock change in Hungary. *Soil and Tillage Research*, *195*, 104410. doi: 10.1016/j.still.2019.104410.

Szatmári, G., & Pásztor, L. (2019). Comparison of various uncertainty modelling approaches based on geostatistics and machine learning algorithms. *Geoderma*, *337*, 1329–1340. doi: 10.1016/j.geoderma.2018.09.008.

Szatmári, G., Pásztor, L., & Heuvelink, G. B. M. (2021). Estimating soil organic carbon stock change at multiple scales using machine learning and multivariate geostatistics. *Geoderma*, *403*, 115356. doi: 10.1016/j.geoderma.2021.115356.

Taghizadeh-Mehrjardi, R., Schmidt, K., Eftekhari, K., Behrens, T., Jamshidi, M., Davatgar, N., Toomanian, N., & Scholten, T. (2020). Synthetic resampling strategies and machine learning for digital soil mapping in Iran. *European Journal of Soil Science*, *71*, 352–368. doi: 10.1111/ejss.12893. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejss.12893.

Taghizadeh-mehrjardi, R., Toomanian, N., Khavaninzadeh, A. R., Jafari, A., & Triantafilis, J. (2016). Predicting and mapping of soil particle-size fractions with adaptive neuro-fuzzy inference and ant colony optimization in central Iran. *European Journal of Soil Science*, *67*, 707–725. doi: 10.1111/ejss.12382. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejss.12382.

Tanneberger, F. et al. (2022). Saving soil carbon, greenhouse gas emissions, biodiversity and the economy: paludiculture as sustainable land use option in German fen peatlands. *Regional Environmental Change*, *22*, 69. doi: 10.1007/s10113-022-01900-8.

Teuling, K., Knotters, M., van Tol-Leenders, T. P., Lesschen, J. P., & Reijneveld, J. A. (2021). *Nieuwe steekproefomvang voor landelijke monitoring koolstof en bodemkwaliteit, Vervolg op rapportages CC-NL en De staat van de Nederlandse landbouwbodems in 2018*. Technical Report, Wageningen Environmental Research Wageningen.

Thomas, G. W. (1996). Soil pH and Soil Acidity. In *Methods of Soil Analysis* (pp. 475–490). John Wiley & Sons, Ltd. doi: 10.2136/sssabookser5.3.c16 section: 16 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2136/sssabookser5.3.c16.

Thunnissen, H. a. M., & Middelaar, H. J. v. (1995). *The CORINE Land Cover database of the Netherlands; final report of the CORINE Land Cover project in the Netherlands*. Technical Report 78, SC-DLO Wageningen. `https://library.wur.nl/WebQuery/wurpubs/302149`.

TNO, G. D. N. (2020). BROloket: Ondergrondgegevens.

Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, *46*, 234–240. doi: 10.2307/143141. Publisher: [Clark University, Wiley].

van Tol-Leenders, D., Knotters, M., Groot, W. d., Gerritsen, P., Reijneveld, A., van Egmond, F., Wösten, H., & Kuikman, P. (2019). *Koolstofvoorraad in de bodem van Nederland (1998-2018) : CC-NL*. Rapport 2974, Wageningen Environmental Research Wageningen. `https://doi.org/10.18174/509781`. doi: 10.18174/509781.

Törnqvist, T. E., Wallace, D. J., Storms, J. E. A., Wallinga, J., van Dam, R. L., Blaauw, M., Derksen, M. S., Klerks, C. J. W., Meijneken, C., & Snijders, E. M. A. (2008). Mississippi Delta subsidence primarily caused by compaction of Holocene strata. *Nature Geoscience*, *1*, 173–176. doi: 10.1038/ngeo129. Number: 3 Publisher: Nature Publishing Group.

United Nations (2015). *Transforming our World: The 2030 Agenda for Sustainable Development*. Technical Report, United Nations New York, NY.

Vanino, S. et al. (2023). Barriers and opportunities of soil knowledge to address soil challenges: Stakeholders' perspectives across Europe. *Journal of Environmental Management*, *325*, 116581. doi: 10.1016/j.jenvman.2022.116581.

Varón-Ramírez, V. M., Araujo-Carrillo, G. A., & Guevara Santamaría, M. A. (2022). Colombian soil texture: building a spatial ensemble model. *Earth System Science Data*, *14*, 4719–4741. doi: 10.5194/essd-14-4719-2022. Publisher: Copernicus GmbH.

Vasenev, V. I., Stoorvogel, J. J., Vasenev, I. I., & Valentini, R. (2014). How to map soil organic carbon

stocks in highly urbanized regions? *Geoderma*, *226-227*, 103–115. doi: 10.1016/j.geoderma.2014.03.007.

Vasenev, V. I., Varentsov, M., Konstantinov, P., Romzaykina, O., Kanareykina, I., Dvornikov, Y., & Manukyan, V. (2021). Projecting urban heat island effect on the spatial-temporal variation of microbial respiration in urban soils of Moscow megalopolis. *Science of the Total Environment*, *786*. doi: 10.1016/j.scitotenv.2021.147457.

Vaysse, K., & Lagacherie, P. (2017). Using quantile regression forest to estimate uncertainty of digital soil mapping products. *Geoderma*, *291*, 55–64. doi: 10.1016/j.geoderma.2016.12.017.

Viscarra Rossel, R. et al. (2016). A global spectral library to characterize the world's soil. *Earth-Science Reviews*, *155*, 198–230. doi: 10.1016/j.earscirev.2016.01.012.

Viscarra Rossel, R. A., Chen, C., Grundy, M. J., Searle, R., Clifford, D., & Campbell, P. H. (2015). The Australian three-dimensional soil grid: Australia's contribution to the GlobalSoilMap project. *Soil Research*, *53*, 845. doi: 10.1071/SR14366.

Viscarra Rossel, R. A., Jeon, Y. S., Odeh, I. O. A., & McBratney, A. B. (2008). Using a legacy soil sample to develop a mid-IR spectral library. *Soil Research*, *46*, 1. doi: 10.1071/SR07099.

Viscarra Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L. J., & Skjemstad, J. O. (2006). Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, *131*, 59–75. doi: 10.1016/j.geoderma.2005.03.007.

Visschers, R., Finke, P. A., & de Gruijter, J. J. (2007). A soil sampling program for the Netherlands. *Geoderma*, *139*, 60–72. doi: 10.1016/j.geoderma.2007.01.008.

Vos, P. (2015). *Origin of the Dutch coastal landscape: long-term landscape evolution of the Netherlands during the Holocene, described and visualized in national, regional and local palaeogeographical map series*. Groningen: Barkhuis.

Vos, P., Meulen, M. v. d., Weerts, H., & Bazelmans, J. (2020). *Atlas of the Holocene Netherlands, landscape and habitation since the last ice age*. Amsterdam: Amsterdam University Press. Last Modified: 2020-03-17T11:43 Publisher: Ministerie van Onderwijs, Cultuur en Wetenschap.

de Vries, F., & Al, E. J. (1992). *De groeiplaatsgeschiktheid voor bosdoeltypen in beeld met ALBOS*. Technical Report 234, DLO-Staring Centrum. `https://edepot.wur.nl/303208` publisher: DLO-Staring Centrum.

de Vries, F., Brouwer, F., & Walvoort, D. (2018). *Basisregistratie Ondergrond (BRO) actualisatie bodemkaart: Herkatering westelijk veengebied Waterschap Drents Overijsselse Delta*. Wageningen Environmental Research Rapport 2887, Wageningen Environmental Research Wageningen. `https://doi.org/10.18174/450341` iSSN (Print):1566-7197.

de Vries, F., Brus, D. J., Kempen, B., Brouwer, F., & Heidema, A. H. (2014). *Actualisatie bodemkaart veengebieden : deelgebied en 2 in Noord Nederland*. Alterra-rapport 2556, Alterra Wageningen. `https://edepot.wur.nl/314315` iSSN (Print: 1566-7197.

de Vries, F., de Groot, W., Hoogland, T., & Denneboom, J. (2003). *De Bodemkaart van Nederland digitaal;. Toelichting bij inhoud, actualiteit en methodiek en korte beschrijving van additionel informatie*. Alterra-rapport 811, Alterra, Research Instituut voor de Groene Ruimte Wageningen, the Netherlands.

de Vries, F., Lesschen, J. P., Akker, J. J. H. v. d., Petrescu, A. J., Huissteden, J. v., & Wyngaert, I. J. J. v. d. (2009). *Bodemgerelateerde emissie van broeikasgassen in Drenthe : de huidige situatie*. Technical Report 1859, Alterra Wageningen. `https://library.wur.nl/WebQuery/wurpubs/395179` iSSN: 1566-7197.

de Vries, F., Walvoort, D., & Brouwer, F. (2017). *Basisregistratie Ondergrond (BRO) Actualisatie bodemkaart; Herkartering van de eenheden met slappe kleilagen*. Wageningen Environmental Research Rapport 2834, Wageningen Environmental Research Wageningen. `https://doi.org/10.18174/423728`. doi: 10.18174/423728.

de Vries, W., Schulte-Uebbing, L., Kros, H., Voogd, J. C., & Louwagie, G. (2021). Spatially explicit boundaries for agricultural nitrogen inputs in the European Union to meet air and water quality

targets. *Science of The Total Environment*, *786*, 147283. doi: 10.1016/j.scitotenv.2021.147283.

Wadoux, A. M. J. C. (2019). Using deep learning for multivariate mapping of soil with quantified uncertainty. *Geoderma*, *351*, 59–70. doi: 10.1016/j.geoderma.2019.05.012.

Wadoux, A. M. J.-C., & Heuvelink, G. B. M. (2023). Uncertainty of spatial averages and totals of natural resource maps. *Methods in Ecology and Evolution*, *14*, 1320–1332. doi: 10.1111/2041-210X.14106. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.14106.

Wadoux, A. M. J. C., Heuvelink, G. B. M., de Bruin, S., & Brus, D. J. (2021a). Spatial cross-validation is not the right way to evaluate map accuracy. *Ecological Modelling*, *457*, 109692. doi: 10.1016/j.ecolmodel.2021.109692.

Wadoux, A. M. J. C., Heuvelink, G. B. M., Lark, R. M., Lagacherie, P., Bouma, J., Mulder, V. L., Libohova, Z., Yang, L., & McBratney, A. B. (2021b). Ten challenges for the future of pedometrics. *Geoderma*, *401*, 115155. doi: 10.1016/j.geoderma.2021.115155.

Wadoux, A. M. J. C., Minasny, B., & McBratney, A. B. (2020). Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews*, *210*, 103359. doi: 10.1016/j.earscirev.2020.103359.

Wadoux, A. M. J.-C., & Molnar, C. (2022). Beyond prediction: methods for interpreting complex models of soil variation. *Geoderma*, *422*, 115953. doi: 10.1016/j.geoderma.2022.115953.

Wadoux, A. M. J.-C., Padarian, J., & Minasny, B. (2019). Multi-source data integration for soil mapping using deep learning. *SOIL*, *5*, 107–119. doi: 10.5194/soil-5-107-2019. Publisher: Copernicus GmbH.

Wadoux, A. M. J.-C., Saby, N. P. A., & Martin, M. P. (2023). Shapley values reveal the drivers of soil organic carbon stock prediction. *SOIL*, *9*, 21–38. doi: 10.5194/soil-9-21-2023. Publisher: Copernicus GmbH.

Wageningen UR- Alterra (2006). Grondsoortenkaart 2006 - Simplified Soil Map of the Netherlands. doi: 10.17026/dans-xky-fsk5.

Walker, P. H., Hall, G. F., & Protz, R. (1968). Relation between Landform Parameters and Soil Properties. *Soil Science Society of America Journal*, *32*, 101–104. doi: 10.2136/sssaj1968.03615995003200010026x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2136/sssaj1968.03615995003200010026x.

Wallig, M., Microsoft, & Weston, S. (2022a). Provides Foreach Looping Construct: package 'foreach'.

Wallig, M., Microsoft Corporation, & Weston, S. (2022b). Foreach Parallel Adaptor for the 'parallel' Package: package 'doParallel'.

Walvoort, D., & Hoogland, T. (2017). Metadata for the Dutch contribution to the Global Soil Organic Carbon (GSOC) map.

Walvoort, D. J. J., & de Gruijter, J. J. (2001). Compositional Kriging: A Spatial Interpolation Method for Compositional Data. *Mathematical Geology*, *33*, 16.

Wamelink, G. W. W., Walvoort, D. J. J., Sanders, M. E., Meeuwsen, H. A. M., Wegman, R. M. A., Pouwels, R., & Knotters, M. (2019). Prediction of soil pH patterns in nature areas on a national scale. *Applied Vegetation Science*, *22*, 189–199. doi: 10.1111/avsc.12423. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/avsc.12423.

Wang, Z., & Shi, W. (2017). Mapping soil particle-size fractions: A comparison of compositional kriging and log-ratio kriging. *Journal of Hydrology*, *546*, 526–541. doi: 10.1016/j.jhydrol.2017.01.029.

Webster, R., & Oliver, M. A. (2007). *Geostatistics for Environmental Scientists*. (2nd ed.). Chichester: John Wiley & Sons, Ltd.

Weil, R., & Brady, N. (2017). *The Nature and Properties of Soils*. (15th ed.). Pearson Education.

WENR (2020). Landelijk Grondgebruik Nederland (LGN).

van der Westhuizen, S., Heuvelink, G. B. M., Hofmeyr, D. P., & Poggio, L. (2022). Measurement error-filtered machine learning in digital soil mapping. *Spatial Statistics*, *47*, 100572. doi: 10.1016/j.spasta.2021.100572.

Wetterlind, J., & Stenberg, B. (2010). Near-infrared spectroscopy for within-field soil characterization: small local calibrations compared with national libraries spiked with local samples. *European Journal*

of Soil Science, *61*, 823–843. doi: 10.1111/j.1365-2389.2010.01283.x.

Wilks, D. (2011). Chapter 8: Forecast Verification. In *Statistical Methods in the Atmospheric Sciences* (pp. 301–394). Elsevier volume 100. (3rd ed.). doi: 10.1016/B978-0-12-385022-5.00008-7.

Williams, B. J. (1976). *Nahuatl soil glyphs from the códice de Santa María Asunción*. Paris.

Williams, B. J., & Jorge, M. d. C. J. y. (2008). Aztec Arithmetic Revisited: Land-Area Algorithms and Acolhua Congruence Arithmetic. *Science*, *320*, 72–77. doi: 10.1126/science.1153976. Publisher: American Association for the Advancement of Science.

Withers, P. J. A., Sylvester-Bradley, R., Jones, D. L., Healey, J. R., & Talboys, P. J. (2014). Feed the Crop Not the Soil: Rethinking Phosphorus Management in the Food Chain. *Environmental Science & Technology*, *48*, 6523–6530. doi: 10.1021/es501670j. Publisher: American Chemical Society.

Witte, H. J. L., & Van Geel, B. (1985). Vegetational and environmental succession and net organic production between 4500 and 800 b.p. reconstructed from a peat deposit in the western dutch coastal area (assendelver polder). *Review of Palaeobotany and Palynology*, *45*, 239–300. doi: 10.1016/0034-6667(85)90004-1.

Wolf, J., Beusen, A. H. W., Groenendijk, P., Kroon, T., Rötter, R., & van Zeijts, H. (2003). The integrated modeling system STONE for calculating nutrient emissions from agriculture in the Netherlands. *Environmental Modelling & Software*, *18*, 597–617. doi: 10.1016/S1364-8152(03)00036-7.

Wood, J. (1996). *The geomorphological characterisation of Digital Elevation Models*. Ph.D. University of Leicester. Accepted: 1996.

Wood, J. (2009). Chapter 14 Geomorphometry in LandSerf. In T. Hengl, & H. I. Reuter (Eds.), *Developments in Soil Science* (pp. 333–349). Elsevier volume 33 of *Geomorphometry*. doi: 10.1016/S0166-2481(08)00014-7.

Wright, M. N., & Ziegler, A. (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, *77*. doi: 10.18637/jss.v077.i01. ArXiv: 1508.04409.

Yagüe, M., Sanz, A., Poggio, L., van Wesemael, B., Tziolas, N., Chabrillat, S., Heiden, U., Gholizadeh, A., & Ben-Dor, E. (2023). The European space agency world soils monitoring system. In *Soils, where food begins: Proceedings of the Global Symposium on soils for nutrition, 26–29 July 2022* (pp. 150–151). Rome, Italy: FAO. doi: 10.4060/cc6728en.

Yang, R.-M., Zhu, C.-M., Zhang, X., & Huang, L.-M. (2022). A preliminary assessment of the space-for-time substitution method in soil carbon change prediction. *Soil Science Society of America Journal*, *86*, 423–434. doi: 10.1002/saj2.20369. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/saj2.20369.

Yigini, Y., & Panagos, P. (2016). Assessment of soil organic carbon stocks under future climate and land cover changes in Europe. *Science of The Total Environment*, *557-558*, 838–850. doi: 10.1016/j.scitotenv.2016.03.085.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*, 338–353. doi: 10.1016/S0019-9958(65)90241-X.

van der Zee, S., van Riemsdijk, W., & de Haan, F. (1990). *Het protokol fosfaatverzadigde gronden I: toelichting, verslagen en mededelingen 1990-1A*. Landbouwuniversiteit Wageningen, Landbouwuniversiteit Wageningen Wageningen. `https://edepot.wur.nl/394261` publisher: Landbouwuniversiteit Wageningen.

Zeraatpisheh, M., Jafari, A., Bagheri Bodaghabadi, M., Ayoubi, S., Taghizadeh-Mehrjardi, R., Toomanian, N., Kerry, R., & Xu, M. (2020). Conventional and digital soil mapping in Iran: Past, present, and future. *CATENA*, *188*, 104424. doi: 10.1016/j.catena.2019.104424.

Zhang, L., Heuvelink, G. B. M., Mulder, V. L., Chen, S., Deng, X., & Yang, L. (2024). Using process-oriented model output to enhance machine learning-based soil organic carbon prediction in space and time. *Science of The Total Environment*, (p. 170778). doi: 10.1016/j.scitotenv.2024.170778.

Zhang, Y., Ji, W., Saurette, D. D., Easher, T. H., Li, H., Shi, Z., Adamchuk, V. I., & Biswas, A. (2020).

Three-dimensional digital soil mapping of multiple soil properties at a field-scale using regression kriging. *Geoderma*, *366*, 114253. doi: 10.1016/j.geoderma.2020.114253.

Zhou, Y., Hartemink, A. E., Shi, Z., Liang, Z., & Lu, Y. (2019). Land use and climate change effects on soil organic carbon in North and Northeast China. *Science of The Total Environment*, *647*, 1230–1238. doi: 10.1016/j.scitotenv.2018.08.016.

Zuur, A. J. (1958). *Bodemkunde der Nederlandse bedijkingen en droogmakerijen: Deel C: Het watergehalte, de indroging en enkele daarmede samenhangende processen*. Kampen: Ministerie van Verkeer en Waterstaat, Rijkswaterstaat, Directie van de Wieringermeer (Noordoostpolderwerken); Landbouwhogeschool, Afdeling Natte Ontginning,Ministerie van Verkeer en Waterstaat, Rijkswaterstaat, Directie van de Wieringermeer (Noordoos... Last Modified: 2023-01-17.

Summary

Humans and most life on land depends on soil for clean air, water, food, natural resources and a living environment. Yet soils vary in space and time, prompting efforts to understand and map soils ever since the development of agriculture and rise of civilizations. Even today, soil mapping is a major challenge because it is only possible to make observations or measurements at a finite number of locations. Consequently, predicting soil properties in space and time requires spatial interpolation, or mapping, using models informed by observations from visited locations. Digital soil mapping (DSM) is the process of mapping soil properties by inferring the relationship between soils and variables representative of soil-forming factors, or covariates, using mathematical and statistical models. DSM has rapidly advanced since its origins in the 1990s, largely due to scientific and technical advances and initiatives such as GlobalSoilMap (GSM), which aims to deliver information of key soil properties and their uncertainties at high-resolution for the entire world.

Even though DSM has rapidly evolved, many areas in the world still lack high-resolution soil information with quantified accuracy. One such area is the Netherlands, where information of key soil properties is crucial to support a broad range of diverse stakeholders from multiple sectors, such as agriculture, forestry, land and water management, spatial planning, infrastructure, transport and nature conservation. In addition, regardless of the study area, there are also major challenges related to assessing map accuracy, mapping in three-dimensional space and time (3D+T) and developing efficient workflows for predicting a large variety of soil properties. Therefore, the aim of this thesis is to develop a high-resolution soil modelling and mapping platform for the Netherlands in 3D+T, which I named "BIS-4D".

To address this aim, **Chapter 2** deals with developing a three-dimensional (3D) model to predict soil pH between 0 - 2 m on a national scale in the Netherlands at 25 m resolution. Soil pH is an essential soil property because it provides information on soil acidity and alkalinity, nutrient availability and affects biological activity, decomposition, metal dissolution and soil physical structure. I use quantile regression forest (QRF), not only due to its predictive performance as a machine learning algorithm, but also because it allows for probabilistic predictions, which means that (model-based) prediction uncertainty can be quantified. In this chapter, I first explore a pivotal aspect of DSM: how to evaluate map quality using external (model-free) and internal (model-based) accuracy assessment. Firstly, I build upon established sampling theory by applying design-based statistical inference using a probability sample in 3D space. Secondly, I investigate how GSM accuracy thresholds can be used as a simple rating system to visualize map uncertainty at all prediction locations. Findings from **Chapter 2** reveal large differences in map quality when comparing design-based statistical inference to other, more commonly used statistical val-

idation strategies. These findings underline the importance of properly quantifying map quality using design-based statistical inference. In doing so, the 3D soil pH maps for the Netherlands are unique among national-scale GSM products in that unbiased accuracy metrics are provided. Findings from implementing GSM accuracy thresholds reveal that less than 10 % of pixels were designated with one of the highest ratings, suggesting that further research is needed to investigate the feasibility of these ratings. Nevertheless, the ratings are promising for easily communicating map quality and comparing soil maps across scales and study areas.

In **Chapter 3**, I extend the mapping methodology from 3D to 3D+T for soil organic matter (SOM). SOM is a fundamental indicator of soil health since it enhances the availability of plant nutrients, improves moisture retention, stabilizes soil structure, increases permeability and chemical buffering and influences the biodegradability of pollutants. Moreover, among dynamic soil properties, SOM is particularly sensitive to temporal changes in the Netherlands due to large-scale excavation, drainage and cultivation of peatlands, unprecedented land reclamation projects and highly anthropogenic landscapes. In this Chapter, the main objectives are to derive covariates that are variable in two-dimensional space and time (2D+T) and 3D+T as drivers of spatio-temporal SOM dynamics and explore how 3D+T mapping can serve as a new paradigm for monitoring SOM-related soil health. I calibrate a QRF model using 869 094 SOM observations from 339 231 point locations and covariates explicit in 2D+T and 3D+T related to land use and the occurrence of peat. The 3D+T model predicts SOM and its uncertainty annually at 25 m resolution between 0 - 2 m depth from 1953 - 2022 in the Netherlands. Expressed as absolute changes in mass percentages, predictions suggest that SOM decreased by more than 25 % in many peatland areas and by 0.1 - 0.3 % in mineral soils under cropland. In contrast, SOM increased by 10 - 25 % below 80 cm depth on vast areas of reclaimed land. Supported by the literature, predicted SOM increases on reclaimed land suggest that land subsidence due to clay ripening, peat oxidation and soil compaction caused peat layers to shift upwards in terms of their depth from the surface. The analysis quantifies the substantial variations of SOM in space, depth, and time, highlighting the inadequacy of evaluating SOM-related soil health at point scale or static mapping at a single depth for policymaking.

Next, the 3D+T machine learning model is applied to predict SOM in 2050 based on a nature-inclusive land use scenario in **Chapter 4**. Nature-inclusive scenarios of the future can help address numerous societal challenges related to soil health. As nature-inclusive scenarios imply sustainable management of natural systems and resources, land use and soil health are assumed to be mutually beneficial in such scenarios. However, the interplay between nature-inclusive land use scenarios and soil health has never been modelled using DSM. By deriving dynamic covariates related to land use and the occurrence of peat for 2050, I predict SOM and its uncertainty in 2050 and assess SOM changes between 2022 and 2050 from 0 - 2 m depth at 25 m resolution. I find little changes in the majority of mineral soils. However, SOM decreases of up to 5 % are predicted in grasslands used for

animal-based production systems in 2022, which transitioned into croplands for plant-based production systems by 2050. Although increases up to 25 % SOM are predicted between 0 - 40 cm depth in rewetted peatlands, larger decreases, on reclaimed land even surpassing 25 % SOM, are predicted on non-rewetted land in peat layers below 40 cm depth. There are several limitations to the approach, mostly due to predicting future trends based on historic data. Furthermore, nuanced nature-inclusive practices, such as the adoption of agroecological farming methods, are too complex to incorporate in the model and would likely affect SOM spatial variation. Nonetheless, 3D mapping of SOM in 2050 creates new insights and raises important questions related to soil health behind nature-inclusive scenarios. Using machine learning explicit in 3D+T to predict the impact of future scenarios on soil health is a useful tool for facilitating societal discussion, aiding policy making and promoting transformative change.

Besides soil pH and SOM, **Chapter 5** develops BIS-4D into an operational, reproducible, standardised, largely automated and efficient modelling platform for mapping a broad range of soil properties on a national scale in the Netherlands. BIS-4D delivers maps of soil texture (clay, silt and sand content), bulk density, total nitrogen, oxalate-extractable phosphorus, cation exchange capacity, and updated versions of soil pH and SOM, and their uncertainties at 25 m resolution between 0 - 2 m depth. Depending on the specific soil property, the statistical model is informed by a range of 3815 - 855 950 observations and selects from a total of 366 environmental covariates. I assess the accuracy of mean and median predictions using design-based statistical inference of a probability sample and location-grouped 10-fold cross-validation. As in previous chapters, I also assess prediction uncertainty using the prediction interval coverage probability. I find that the accuracy of clay, sand and pH maps is highest, while silt, BD, SOM, N_{tot}, CEC and especially P_{ox} are more difficult to predict. One of the main limitations of BIS-4D is that prediction maps cannot be used to quantify the uncertainty of spatial aggregates. A step-by-step manual helps users decide whether BIS-4D is suitable for their intended purpose, openly available code and input data enhance reproducibility and future updates and BIS-4D prediction maps can be easily downloaded. BIS-4D fills the previous data gap of a national scale GSM product in the Netherlands and will hopefully facilitate the inclusion of soil spatial variation as a routine and integral part of decision support systems.

In the last chapter, **Chapter 6**, I discuss the main findings from **Chapters 2 - 5** and broadly discuss their implications, suggest future research and bridge science to much-needed practical implementations. BIS-4D fills the previous gap of high-resolution soil property maps on a national scale in the Netherlands and delivers state-of-the-art 3D and 3D+T maps with quantified accuracy for the GSM project. Furthermore, BIS-4D maps are being uploaded on the national online soil portal (bodemdata.nl) and are already being used for various purposes, such as estimating soil-derived greenhouse gas emissions for the Land Use, Land Use Change and Forestry sector for the United Nations Framework Convention on Climate Change.

DSM has revolutionized soil mapping at a unprecedented rate since the first soil maps were crafted more than 4000 years ago. However, I believe a new paradigm shift lies ahead. Digital soil assessment is underway towards scalable soil estimation services that combine the advantages of proximal and remote soil sensing to monitor soil health on-demand for a tailored target area. In changing times ahead, it is paramount that attention and resources are not only directed towards research, but also towards practical solutions and successful implementation of scientific ideas. Only then will soils also benefit society as a whole and contribute to human well-being.

Acknowledgements

In the scorching 40°C heat, the gentle breeze offered little respite on the rooftop. Below, Red-whiskered Bulbuls flitted through the treetops, while a proud peacock perched on pink walls nearby, surrounded by a garden ablaze with lush greenery and vibrant blooms. In the distance, a dog's bark mingled with the rhythmic breaths of a cow by the building's entrance, chewing dry hay. Here I sat in Tiruvannamalai, Tamil Nadu, India, moments away from an online PhD interview for Wageningen University and Research. Despite the sweltering conditions and shaky internet, my two-month journey through Karnataka, Kerala, and Tamil Nadu had offered me distance from my daily life, presenting a promising path ahead. The freedom to choose one's trajectory often hinges on the people encountered along the way, and for that, I am grateful to those who have journeyed with me over the years.

First and foremost, I extend my gratitude to my PhD supervisors: Titia, Gerard, Mirjam and Joop. Titia, your expertise and guidance as my daily supervisor and co-promoter has been invaluable. During moments of doubt, you provided unwavering support. You challenged my perspectives with thought-provoking questions, fostering both personal and professional growth. Your mentorship has empowered me to navigate my own path, while underscoring the collaborative and team-work essence of science and research. Gerard, having you as a promotor makes it seem easy to do a PhD. Your knack for simplifying complex concepts, refining my writing, and nurturing objectivity has shaped me into a more adept and well-rounded scientist. Mirjam, I am deeply grateful for my seamless integration into the Bodem, Water en Landgebruik (BWL) team and the invaluable connections you've facilitated. Your encouragement has been a constant source of resilience during challenging times. I'll always cherish the memory of our unexpected midnight encounter on one of my first dates with Maria Giovanna, during which you casually sat down next to us and joined our conversation! Joop, without you and Titia this PhD would not have happened. Even from my very first day at WUR, I always felt tremendous support from you. The privilege of speaking at your retirement ceremony remains a cherished highlight, and even post-WUR, your insights shared through "Bodem en Kunst" articles and our ongoing correspondence continue to inspire me.

I am immensely thankful for the enriching experiences shared with my colleagues in the Soil Geography and Landscape (SGL) group, spanning coffee breaks, lunches, leisurely walks, team outings, and insightful discussions. Among these cherished moments, my interactions with fellow (and former) PhDs stand out prominently. Jasper and Marijn, your assistance with integration and moving apartments eased transitions and fostered a sense of belonging. Selçuk, your guidance and shared ideas during moments of struggle has been invaluable. Cynthia and Doina, whether augering marine clay soils after a drought, getting stuck on dirt roads in southern Limburg, collecting earthworms, or letting cows,

horses and sheep inspect our soil profile descriptions, every shared moment during our field work was a delight. Doina, I thoroughly enjoyed our conversations about vegetable farming. Stephan, our engaging discussions about random forest and machine learning were enlightening. Eric, it was fun to play football together and I extend my best wishes to you and your family. Luc, your humility and the invaluable contributions you make behind the scenes are truly admirable. There would be more sustainability, peace and social harmony with more individuals like you. As SGL has flourished in size and diversity, I will fondly remember the vibrant new dynamic and collaborative spirit that emerged post-pandemic. Jakob, your leadership style characterized by humility, self-reflection and encouragement to prioritize well-being, including ample holiday time, has left a lasting impression. I will undoubtedly miss the luxury of eight weeks of holidays!

My journey at WUR would have been incomplete without my colleagues in the BWL team. Despite my physical presence at BWL being limited to two days a week, your collective influence on my PhD, the project, and my personal growth during my "Dutch experience" was profound. I vividly recall my first "Basisregistratie Ondergrond borrel," where after one or two "Wageningse blond", I found myself pleasantly surprised to comprehend and converse in Dutch. Fokke, your patience and assistance with my endless inquiries about soil surveying, BIS data, and your companionship during field campaigns alongside Cynthia, Doina, and Willy, are deeply appreciated. Dennis, our discussions veering into programming or statistics during our "walk and talks" were consistently enlightening. Martin K., I will never forget when you brought me to the archives and showed me evidence of the intricate and multi-faceted processes involved in making soil maps during the days of the "Stichting voor Bodemkartering" (StiBoKa). Your creativity to combine music with soils, advise on which bike route to take for my holiday plans and discussions about science in a philosophical way ("On Bullshit" and "Hoezo plausibel?") expanded my horizons in unexpected ways. Fenny, your tireless dedication to standardization, connecting science with society, and fostering synergies across various projects exemplifies what I think we need more of. Dorothee, thank you for helping realize our shared vision of providing BIS-4D maps on bodemdata.nl, a dream turned into reality. I am grateful to Kees, Willy, Simone, Amanda, Erik, Tom, Ab, Joost(s), Stijn and Hessel for the conversations and moments we shared. I wish to express my gratitude to all soil surveyors from the former StiBoKa, the Staring center, Alterra and today, Wageningen Environmental Research for the arduous efforts in collecting, measuring, and harmonizing the data pivotal to my PhD. Their meticulous work in the field and laboratory underscores the indispensable role of observations in science, a fact that model developers should always remember.

I probably would never have made it to the interview on the roof terrace in Tiruvannamalai if it had not been for my supervisors during my Master thesis: Philipp, Madlene, Raphael and Johan. Philipp, thank you for being patient and investing so much time to teach me about R programming and soil spectroscopy. Madlene, your expertise in digital soil

mapping and statistics were invaluable and I know that I can always count on you for your advise. In addition, I would like to thank Prof. Christine Alewell; her excellent pedology lectures and meticulous soil profile descriptions introduced me to the world of soil science during my Bachelor.

I would like to express my gratitude to my family for their love and guidance. To my parents, for prioritizing my education and giving me the opportunity and financial support to study and chase my dreams. To my mother, for teaching me to be courageous, unconventional and (literally) making me stand on the table so that I would believe in myself as a teenager. To my father, for his constant optimism, even during my wildest moments, and for his impeccable proof-reading of countless essays and reports in school. Julian, you are not only my older brother, but my constant companion and closest friend. We formed alliances during playground feuds in Bern, we fought fiercely during pillow fights in the attic in Urbana, you exploded with excitement when I found a juvenile goshawk tail feather in the forest, we howled to coyotes in Big Bend and we shared a bottle of wine on a "jurassische Wytweide" as we watched the sunset on one of our many camping trips. Furthermore, you were with me on that day in Tiruvannamalai, and during that trip in India, when I decided to start a PhD in Wageningen. Anna, you have grown close to my heart, I cherish your opinions and I will miss being able to see you and Julian so frequently and spontaneously.

Maria Giovanna, what I could not have known at the time in Tiruvannamalai is that the best thing about coming to Wageningen is that it is where I met you. You make me feel like I know you from a previous life, yet you never cease to amaze me. Thanks for sharing your wisdom, enriching my life in countless ways and for your constant support and love. Ti amo!

Special thanks to all my friends that I met in Wageningen: Mark, Anne, Susana, Francisco, Carlo, Francoise, Doina, Pieter, Jan, Maarten, Giustina, Sebas, Marce, Fer, David, Daniel, Mercy, Mickael, Corentin, Ilaria and Dhanush. Mark, I cherish our every encounter, our philosophical and hilarious conversations, sailing in Friesland, and dinners, bathing and ice-skating in the "uiterwaarden". Susana, you were with me through good and tough times; thanks for helping me understand other people's perspectives and adding spice to our many dinner conversations. Franciscito, you are like my "bonus" older brother; I have learned so much because of you and I deeply admire you. Carletto, my intellectual Sardinian shepherd, you are a constant source of inspiration. I also want to thank all my dear friends abroad in Switzerland or elsewhere. I always enjoyed your visits or (video) calls, which were a constant support especially during Corona. Thanks to Claudia and Huib for the Sundays in Nijmegen and passing on your tandem. Thanks to Elsa and GJ from Sheba Coffee, Grinda and Delio from Sole d'Italia, Chris and Pauline from Tuinderij Het Lichtveen, Ivo's Bread and Pizza, the Stadsbrouwerij Wageningen and all Streekwaar members; it is largely thanks to the local food community that we felt at

home in Wageningen.

The people I met on my journey are like the characters in a story or the data in a model. Listen to them. What do they say?

About the author

Anatol Helfenstein was born on July 17th, 1993 in Bern, Switzerland. At the age of seven, he moved with his family to the United Sates, where he eventually completed his high school with a bilingual International Baccalaureate (IB) in Houston, Texas. He went on to pursue a Bachelor degree at the University of Basel, Switzerland, with a major in Environmental Geosciences and a minor in Geology. During his studies, he became increasingly enthusiastic about soil science. He investigated the use of stable isotopes to trace carbon in drained organic soils during his Bachelor thesis with Prof. Christine Alewell and Dr Jens Leifeld and participated in a 3-week soil science excursion in Russia. After completing his BSc, Anatol conducted civil service in the Forest Soils and Biogeochemistry group at the Swiss Federal Institute for Forest, Snow and Landscape Research (WSL). He continued to pursue his passion for environmental and soil science at the Swiss Federal Institute of Technology (ETH) in Zürich, from which he obtained a MSc in Environmental Sciences, with a major in Biogeochemistry and Pollutant Dynamics and a minor in Agricultural Plant Production. In parallel, he also pursued a Teaching Certificate in Environmental Sciences. As part of his studies, he completed an internship at the Swiss Soil Monitoring Network (NABO) and became aware of the challenge of obtaining spatial soil information. Motivated by this challenge, he learned about statistical modelling techniques by using soil spectroscopy and digital soil mapping during his Master thesis with Dr Philipp Baumann, Dr Madlene Nussbaum and Prof. Johan Six. This included a 6-week research exchange at the Commonwealth Scientific and Industrial Research Organisation (CSIRO) in Canberra, Australia, where he worked with Prof. Raphael Viscarra Rossel. Due to his growing interest in pedometrics, he started a PhD at Wageningen University and Research (WUR) in the Netherlands with Dr Titia Mulder, Prof. Gerard Heuvelink, Dr Joop Okx and Dr Mirjam Hack-ten Broeke. Anatol fondly reminisces about his time as a PhD candidate at WUR and the meaningful personal and professional connections he forged, culminating not only in invaluable personal growth but also in the completion of this dissertation. Currently, he works at the Swiss Federal Office for the Environment (FOEN) as a project manager to coordinate the national soil mapping in Switzerland from a methodological perspective.

Peer-reviewed journal publications

Helfenstein, A., Baumann, P., Viscarra Rossel, R., Gubler, A., Oechslin, S., Six, J. (2021). Quantifying soil carbon in temperate peatlands using a mid-IR soil spectral library. *SOIL* 7, 193–215. `https://doi.org/10.5194/soil-7-193-2021`

Baumann, P., **Helfenstein, A.**, Gubler, A., Keller, A., Meuli, R.G., Wächter, D., Lee, J., Viscarra Rossel, R., Six, J. (2021). Developing the Swiss mid-infrared soil spectral li-

brary for local estimation and monitoring. *SOIL* 7, 525–546. https://doi.org/10.5194/soil-7-525-2021

Helfenstein, A., Mulder, V.L., Heuvelink, G.B.M., Okx, J.P. (2022). Tier 4 maps of soil pH at 25 m resolution for the Netherlands. *Geoderma* 410, 115659. https://doi.org/10.1016/j.geoderma.2021.115659

Helfenstein, A., Mulder, V.L., Heuvelink, G.B.M., Hack-ten Broeke, M.J.D. (2024). Three-dimensional space and time mapping reveals soil organic matter decreases across anthropogenic landscapes in the Netherlands. *Communications Earth & Environment* 5, 1–16. https://doi.org/10.1038/s43247-024-01293-y

van Doorn, M., **Helfenstein, A.**, Ros, G.H., Heuvelink, G.B.M., van Rotterdam-Los, D.A.M.D., Verweij, S.E., de Vries, W. (2024). High-resolution digital soil mapping of amorphous iron- and aluminium-(hydr)oxides to guide sustainable phosphorus and carbon management. *Geoderma* 443, 116838. https://doi.org/10.1016/j.geoderma.2024.116838

Helfenstein, A., Mulder, V.L., Hack-ten Broeke, M.J.D., van Doorn, M., Teuling, K., Walvoort, D.J.J., and Heuvelink, G.B.M. BIS-4D: Mapping soil properties and their uncertainties at 25 m resolution in the Netherlands. *Earth System Science Data Discussions* 1–43 [preprint], https://doi.org/10.5194/essd-2024-26, accepted, 2024.

Helfenstein, A., Mulder, V.L., Hack-ten Broeke, M.J.D., Breman, B. (2024). A nature-inclusive future with healthy soils? Mapping soil organic matter in 2050 in the Netherlands. *European Journal of Soil Science*, in review.

Peer-reviewed dataset publications

Helfenstein, A., Mulder, V.L., Heuvelink, G.B.M., Okx, J.P. (2021). Tier 4 maps of soil pH at 25 m resolution for the Netherlands. https://doi.org/10.4121/16451739

Helfenstein, A., Teuling, K., Walvoort, D.J.J., Hack-ten Broeke, M.J.D., Mulder, V.L., van Doorn, M., Heuvelink, G.B.M. (2024). Georeferenced point data of soil properties in the Netherlands. https://doi.org/10.4121/C90215B3-BDC6-4633-B721-4C4A0259D6DC

Helfenstein, A., Mulder, V.L., Hack-ten Broeke, M.J.D., van Doorn, M., Teuling, K., Walvoort, D.J.J., Heuvelink, G.B.M. (2024). Spatially explicit environmental variables at 25m resolution for spatial modelling in the Netherlands. https://doi.org/10.4121/6AF610ED-9006-4AC5-B399-4795C2AC01EC

Helfenstein, A., Mulder, V.L., Hack-ten Broeke, M.J.D., van Doorn, M., Teuling, K., Walvoort, D.J.J., Heuvelink, G.B.M. (2024). BIS-4D: Maps of soil properties and their uncertainties at 25 m resolution in the Netherlands. https://doi.org/10.4121/0C934AC6-2E95-4422-8360-D3A802766C71

Conference contributions

Helfenstein, A., Baumann, P., Viscarra Rossel, R., Gubler, A., Oechslin, S., Six, J. (2019). Predicting soil carbon by building a national soil spectral library and using digital soil mapping: a pilot study for the Rhine valley, Switzerland. Digital Soil Mapping Workshop for the German and Swiss National Soil Associations (DBG & BGS), April 4-5, Tübingen, Germany.

Helfenstein, A., Baumann, P., Viscarra Rossel, R., Gubler, A., Oechslin, S., Six, J. (2019). Predicting soil carbon by building a Swiss soil spectral library and using digital soil mapping. Post-Graduate Research Conference at the University of Agricultural and Horticultural Sciences, May 20, Shivamogga, India.

Helfenstein, A., Mulder, V.L., Heuvelink, G.B.M., Okx, J. (2021). BIS-3D: high resolution 3D soil maps for the Netherlands using accuracy thresholds. European Geosciences Union General Assembly, April 19-30, online. `https://doi.org/10.5194/egusphere-egu21-7836`

Helfenstein, A., Mulder, V.L., Heuvelink, G.B.M., Hack-ten Broeke, M.J.D. (2022). Machine learning in four dimensions for mapping soil organic matter changes between 1953-2018 at 25 m resolution in the Netherlands. European Geosciences Union General Assembly, May 23-27, online. `https://doi.org/10.5194/egusphere-egu22-5624`

Helfenstein, A., Sekulić, A., Mulder, V.L., Heuvelink, G.B.M. (2022). Random forest spatial interpolation for 3D digital soil mapping. 22nd World Congress of Soil Science, August 1-4, Glasgow, Scotland.

Helfenstein, A., Mulder, V.L., Heuvelink, G.B.M., Hack-ten Broeke, M.J.D. (2022). A four-dimensional soil modelling and mapping platform at 25 m resolution for the Netherlands. Netherlands Earth Sciences Conference, September 5-6, Utrecht, the Netherlands.

Helfenstein, A., Mulder, V.L., Heuvelink, G.B.M., Hack-ten Broeke, M.J.D. (2023). Machine learning for mapping soil organic matter changes over 70 years in 3D at 25 m resolution in the Netherlands. Soil Mapping for a Sustainable Future: 2nd joint Workshop of the IUSS Working Groups Digital Soil Mapping and Global Soil Map, February 7-9, Orleans, France.

Helfenstein, A., Mulder, V.L., Heuvelink, G.B.M., Hack-ten Broeke, M.J.D. (2023). 3D+T mapping reveals soil organic matter changes between 1953 and 2022 at 25 m resolution in the Netherlands. Wageningen Soil Conference, 28 August – 1 September, Wageningen, the Netherlands.

Helfenstein, A., Baumann, P., Viscarra Rossel, R., Gubler, A., Oechslin, S., Six, J. (2023). Operationalizing soil spectral libraries: a case study for soil carbon in peat soils of Switzerland. Wageningen Soil Conference, 28 August – 1 September, Wageningen, the Netherlands.

Helfenstein, A., Mulder, V.L., Heuvelink, G.B.M., Hack-ten Broeke, M.J.D. (2023). 3D+T mapping in the Netherlands reveals soil organic matter changes between 1953 and 2022 at 25 m resolution. European Healthy Soils Conference Series, 1. Edition: Soil Fertility, September 13-15, Muttenz, Switzerland.

Other contributions

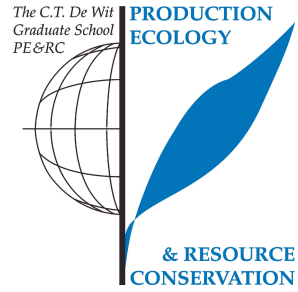
Helfenstein, A. (2016). C_4 carbon accumulation potential of *Miscanthus* cropping on a drained organic soil (Bachelor thesis). University of Basel, Switzerland.

Helfenstein, A. (2019). Predicting soil carbon by building a national soil spectral library and using digital digital soil mapping: a pilot study for the Rhine valley, Switzerland (Master thesis). Swiss Federal Institute of Technology (ETH) Zürich, Switzerland.

Helfenstein, A. (2021). Four dimensional information about the skin of the Earth (video). Wageningen University and Research, the Netherlands. `https://www.youtube.com/watch?v=ENCYUnqc-wo`

PE&RC Training and Education Statement

With the training and education activities listed below the PhD candidate has complied with the requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)



Review/project proposal (4.5 ECTS)

- Developing a high-resolution 3D soil modelling and mapping platform for the Netherlands

Post-graduate courses (7.1 ECTS)

- R and big data; PE&RC (2019)
- Linux basic course; HPC centre WUR (2019)
- HPC basic course; HPC centre WUR (2019)
- Geostatistics; PE&RC (2019)
- Uncertainty propagation in spatial environmental modelling; PE&RC (2019)
- Summer school; OpenGeoHub (2020)
- Soil, geomorphology and landuse of Galloway, Scotland in microcosm; World Congress of Soil Science (2022)

Laboratory training and working visits (1.2 ECTS)

- Digital soil mapping of soil organic matter in 3D space and time; GlobalSoilPartnership, Food & Agriculture Organization of the United Nations (2023)

Invited review of journal manuscripts (5 ECTS)

- Soil Security: mid-infrared soil spectroscopy modelling (2021)
- Geoderma Regional: digital soil mapping (2021)
- Spatial Statistics: digital soil mapping (2021)
- Land Degradation & Development: digital soil mapping (2022)
- Geoderma Regional: digital soil mapping (2022)

Competence, skills and career-oriented activities (2.25 ECTS)

- Project and time management; WGS (2020)

- Reviewing a scientific manuscript; WGS (2021)
- PhD workshop carousel: essentials of scientific writing; designing an attractive & effective poster; impostor syndrome; WGS (2022)
- PhD workshop carousel: research data & scientific publishing; essentials of scientific writing; entrepreneurship as a skillset for career development; visual thinking; WGS (2023)

Scientific integrity/ethics in science activity (0.3 ECTS)

- Scientific integrity; WGS (2022)

PE&RC Annual meetings, seminars and PE&RC weekend/retreat (2.1 ECTS)

- PE&RC first years weekend (2019)
- PE&RC day (2019, 2022)
- PE&RC Last year's retreat (2023)

Discussion groups/local seminars or scientific meetings (7.6 ECTS)

- R user meeting (2019-2022)
- Modelling and simulation (2019-2022)
- Presentation for OpenGeoHub seminar (2020)
- Presentation for ISRIC & WENR group meeting (2021)
- Presentation at farewell retirement symposium of Dr. Joop Okx (2021)
- Presentation of SOM 3D+T maps at WENR group meeting (2023)
- Presentation of SOM 3D+T maps at FAO for GlobalSoilPartnership in Rome, Italy (2023)
- Organized and taught masterclass at Wageningen Soil Conference (2023)
- Workshop at Wageningen model & data day (2023)

International symposia, workshops and conferences (17.6 ECTS)

- Digital soil mapping workshop for the German and Swiss national soil associations; oral presentation; Tübingen, Germany (2019)
- Post-graduate research conference at the University of Agricultural and Horticultural Sciences; poster presentation; Shivamogga, India (2019)
- Pedometrics webinar; oral presentation; online (2021)
- Pico presentation at EGU 2021; online (2021, 2022)
- World Congress of Soil Science; poster presentation; Glasgow, Scotland (2022)
- NAC; oral presentation; Utrecht, the Netherlands (2022)
- Soil Mapping for a Sustainable Future; oral presentation; Orleans, France (2023)
- Wageningen Soil Conference; oral & poster presentation; Wageningen, the Netherlands (2023)
- Healthy Soils Conference; oral presentation; Basel/Muttenz, Switzerland (2023)

Societally relevant exposure (1.5 ECTS)

- WIMEK and soil cluster video pitch (2021)
- Presentation stand and oral pitch at WUR digital innovation expo (2023)
- Podcast/audio recording of research article and PhD thesis (2023)

Lecturing/supervision of practicals/tutorials (11.4 ECTS)

- Spatial and temporal analysis for earth and environment (2020-2023)
- Geology, soils and landscapes of the Northwest European lowlands (2021)
- Introduction to soil geography (2021-2023)

MSc thesis supervision (3 ECTS)

- Cristina Cavero Panez: The impact of temporal land use variability on soil organic carbon changes: a case study at point level in the Netherlands from 1998-2018 (2021)

The research described in this thesis was financially supported by the Dutch Ministry of Agriculture, Nature and Food Quality (WOT-04-013-010 and KB-36-001-014).

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design by Anatol Helfenstein

Printed by ProefschriftMaken on FSC-certified paper