2021

Predictive Digital Soil Mapping

White Paper prepared for LWS 548

Lindsay Nelson

Masters of Land and Water Systems – University of British Columbia

August 17, 2021

Table of Contents

Executive Summary

Soil is important to human and environmental health and ecological functioning. There is a demand for high-resolution quantitative soil data in many applications including agriculture, forestry, mining, land reclamation, and environmental science. Soil scientists are trying to meet this demand using new mapping methods.

Soil scientists combine soil observations with models of soil-landscape relationships to interpret how soil varies over space. Soil mapping is how this information is compilated and shared with wider users. Conventional mapping depends on mental models and produces sharply bounded map units with qualitative estimates of accuracy and uncertainty. The process of conventional mapping is time-consuming and resource-intensive, and the end results are not updatable or completely reproducible.

Recent technological advancement (GIS, remote sensing, computer processing) has enabled soil mapping to become more quantitative. Predictive digital mapping harnesses statistical techniques to generate soil maps with gridded continuous data and quantitative uncertainty estimates. These maps can be updated dynamically as new data becomes available. The inputs for predictive mapping are soil values used to calibrate and validate models and environmental covariates. Maps are commonly produced using existing or legacy data, often due to budget constraints, but the collection of new data using statistical sampling designs and an independent dataset for model validation have the potential to generate the greatest accuracy.

Predictive digital soil mapping has progressed from academic to operational over the last decade. Soil survey agencies in many countries now incorporate these techniques with good results. Some soil properties and classifications can be predicted with suitable levels of accuracy, but others are more difficult to estimate, providing opportunities for future developments. With this progress, the development of minimum quality standards has improved, but a universal system is yet to be enforced. Communication with end-users is critical to the success of PDSM, to ensure that mapping products are valuable and being applied judiciously.

List of Definitions and Acronyms

DEM	Digital elevation model
Covariates	Observed continuous variables that vary with the unknown target variable
GIS	Geographic information systems
LiDAR	Light Detection and Ranging. Remote sensing technique that uses laser pulses to measure the distance to Earth's surface. Common data source of DEMs.
Map scale	The relationship (or ratio) between distance on a map and distance on the ground. So-called small-scale maps (e.g. 1:500,000 - meaning one unit on the map equals 500,000 units on the ground) represent large areas but have less detail, while large-scale maps (e.g. 1:10,000) have a larger ratio and therefore represent smaller areas in greater detail.
PDSM	Predictive digital soil mapping. Also referred to as digital soil mapping (DSM) and predictive soil mapping (PSM) in the literature. Predictive digital soil mapping will be used in this work, as it is the most descriptive name, and the nomenclature used by the Canadian Society of Soil Science (CSSS).
Pedology	Science of soil formation, including the description and classification of soils.
Pedometrics	The application of mathematical and statistical methods in studying the distribution and genesis of soils. Includes modelling of space-time soil variation and soil-landscape processes, linking proximal or remote sensing data to soil properties, analyzing uncertainty propagation within soil models, and sampling design optimization. http://pedometrics.org/
Raster	A raster consists of cells, or pixels, organized into a grid where each cell contains data values. Raster data are one of the main formats used in GIS, examples included aerial photographs and satellite imagery. (Esri)
Remote Sensing	Collection of landscape information using non-direct or distanced methods such as satellites and aerial surveys. Remote sensing is an efficient technique for collecting large amounts of data with more complete coverage than on-the-ground methods.
Soil attribute	Individual characteristic or property of a soil, such as colour, pH, abundance of certain element, texture (sand, silt, clay content), etc.
Soil classification	Organization of soils into categories based on their characteristics. Includes soil order, suborder, great group, subgroup, family, and series. (Brady & Weil, 2008)
Soil profile	Vertical cross-section through the soil, extending from surface to parent material at depth. Displays soil horizons. (Brady & Weil, 2008)
Spatial autocorrelation	Principle that locations nearer to each other will be more related (positively or negatively) than points farther away.
Vector	In GIS, a coordinate-based data model that represents geographic features as points, lines, and polygons. Attributes are connected to each feature. (Esri)

1. Introduction

Soils, being at the nexus of landscape, climate, and biogeochemical cycling, are essential to understanding environment and ecosystem health. In addition to providing ecosystem services such as crop production, temperature regulation, water filtration, and carbon storage (Bouma, 2014; Nolan et al., 2021), soils also hold intrinsic and cultural value (Minami, 2009). If we can characterize how soil varies across landscapes and through time, we are more capable of making land-management decisions that support and enhance soil functioning rather than degrading the soil and damaging ecosystems (Doran, 2002; Gregorich et al., 1997; Lal, 2007).

The compilation and interpretation of soil information is communicated through soil mapping. Conventional (or traditional) soil mappers generally rely on field data combined with their expert knowledge of unformalized rules or assumptions about soil-forming factors to delineate sharp boundaries between soil units (Hudson, 1992). Predictive digital soil mapping (PDSM) is a mapping technique enabled by advances in computing and data science that uses numerical or statistical models to combine information about soil-forming factors obtained from spatial data layers (environmental covariates) with point soil data in order to estimate soil features for a map area (MacMillan & Hengl, 2019; McBratney et al., 2003a; Scull et al., 2003). PDSM outputs are generally gridded data that treat soil as a continuum rather than the sharp map units of traditional mapping. This grid representation has the potential to be a more accurate depiction of soil variability over space. Gridded quantitative soil data is also more desirable in interdisciplinary applications such as environmental modelling (Grunwald, 2006a). However, the quality of the data inputs and assumptions made during the modeling strongly influence the accuracy of the results.

Predictive mapping offers further advantages over traditional soil mapping. By explicitly describing the variables and models used, the mapping process becomes transparent and repeatable, and errors or assumptions are more easily interrogated. Additionally, due to the quantitative nature of these techniques, corresponding maps of the uncertainty of predictions can be generated (Macmillan, 2012). Areas of the map with poor model fit can be prioritized for sample collection over well-characterized zones. These features allow soil formation models to be extended over large regions with greater confidence, without necessarily needing the same time or resources as a traditional survey would to complete a map (Drozdowski et al., 2019; McSweeney et al., 1994; Minasny & McBratney, 2016). The iterative, updatable nature of PDSM systems allows for maps to be dynamic, creating new possibilities for monitoring soil properties over time.

Work in PDSM has been gradually moving from the academic research phase into application (McBratney et al., 2003a; Minasny & McBratney, 2016). The development of PDSM methods and specific systems are guided, in principle, by end-user requirements (Drozdowski et al., 2019; Florinsky, 2016). However, most reviews and descriptions of PDSM are oriented towards technical experts completing research in this field. Potential applications of PDSM projects can get lost in complicated details about how the mapping was done and not why, and reporting of accuracy and uncertainty can be difficult to understand (Arrouays, McBratney, et al., 2020). For these techniques to be embraced by end-users and decision-makers, such as government agencies, environmental consultants, agrologists, foresters, or mining industry members, the current status and future potential of PDSM must be communicated clearly and directly.

The objectives of this paper are to first explain what PDSM is, why it is important, and describe the fundamental processes and techniques used in this mapping in a way that is comprehensible to non-experts; and secondly, determine whether predictive mapping results are at a stage where they are compatible with the needs of government, industry, and other end-users of soil data.

2. Methodology

Research for this paper was conducted from April-July 2021. A literature review was completed with a focus on established predictive mapping projects, often with some degree of government involvement, rather than smaller-scale research using cutting-edge techniques. This distinction was made due to the objective of examining the current application of PDSM techniques outside of purely research contexts. Scientific papers and government reports were the main sources of information.

Consultation with subject experts was essential in the development of this project. Semi-structured interviews with predictive mapping practitioners were integral to gaining a better understanding of the mapping process and discussion of issues facing predictive mapping. The "Progress in predictive digital mapping and proximal soil sensing" session of the Canadian Society of Soil Science (CSSS) 2021 conference was another valuable source of information.

3. Context

3.1. Soil Formation Principles

Soil is a mixture of organic and inorganic particles, gases, and liquid distinguished from other unconsolidated materials on Earth's surface by its ability to support life. The complex physical, chemical, and biological features and processes in soil change not only laterally across landscapes, but also with depth, creating distinctive layers called soil horizons.

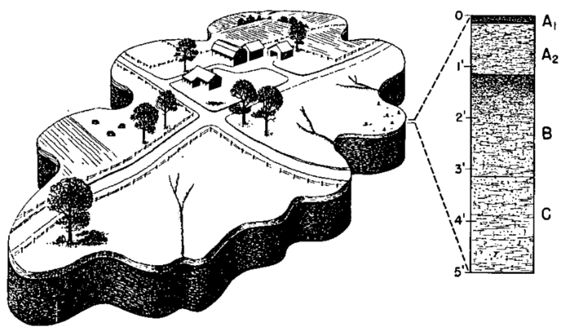

Figure 1. Sketch of a soil unit with a representative soil profile. Depth markers are on the left of the profile, and horizon labels (A1, A2, B, C) are on the right (Simonson, 1959).

Soil scientists conceptualize the development of soils and their properties as a function of multiple environmental factors. The selection of these factors is the result of field observation and soil research. Therefore, if functions connecting soil to those factors can be defined and quantified, it should be possible to predict soil properties for a set of environmental conditions. This hypothesis was first published by Vasily Dokuchaev in 1886 (Florinsky, 2012) and popularized by Hans Jenny through his seminal book on the subject (Jenny, 1941).

This original function is as follows:

$$S = f(cl, o, r, p, t, \dots) \qquad (1)$$

Where S is soil, and cl, o, r, p, t are the soil-forming factors of climate, organisms, relief (topography), parent material, and time, respectively. Each of these factors encompasses many potential variables. For example, topography includes elevation, slope, aspect (e.g., north-facing), curvature, and many other features that impact soil development. The ellipsis in the equation indicates the possibility of adding other factors as our understanding of soil progresses.

Other influential pedologic models use a different approach than Dokuchaev and Jenny (Brevik et al., 2016). Roy Simonson's "Theory of Soil Genesis" changed the focus from external factors to processes occurring within a soil unit that impact the development of soil horizons, such as additions, removals, translocations, or transformations (Simonson, 1959). Geomorphology considers models of erosion and deposition in open or closed basins and the influence of slope curvature on water dynamics and resulting soil properties (Ruhe, 1975). Both these models are useful in soil classification and mapping as they promote the understanding of soil-landscape relationships.

More recent advances in statistical soil models led to the development of the scorpan-SSPFe (soil spatial prediction function with spatially autocorrelated errors) method (McBratney et al., 2003a):

$$S = f(s, c, o, r, p, a, n) + \varepsilon \qquad (2)$$

Where S = soil class or individual soil attribute at a point, can be predicted by:

s: soil, other properties of the soil at that point;

c: climate;

o: organisms, vegetation or fauna or human activity;

r: topography, landscape attributes;

p: parent material, lithology;

a: age, the time factor;

n: space, spatial position.

With an estimate of error or uncertainty, ε.

The *scorpan* model builds on the soil formation-factor model first introduced by adding soil and space as variables. The inclusion of information about the soil and spatial position changes the model to account for spatial autocorrelation, or the concept that points closer to each other will be more related to each other than points that are far away (Webster & Oliver, 2007). The development of the *scorpan* model is the result of efforts to create a generic model and framework to use in predictive soil mapping (McBratney et al., 2003a). This approach to modelling has largely been accepted and forms the foundation of most modern PDSM.

3.2. Conventional Soil Mapping

To represent soils on a conventional map, the extreme and continuous variation of soil must be simplified. Soil mappers apply the principles of soil formation-factors (*clorpt*) in a qualitative way, using their own mental models to draw boundaries between soil sample sites with different properties (Hudson, 1992). Soil point data is correlated with geological maps, topography, vegetation, and air-photo interpretation to delineate soil units based on the mapper's understanding of these features' control over soil formation. The logic used to decide which landscape features most control soil distribution, and any formalized rules applied during mapping, may not be available or well documented.

Sample design in a conventional soil survey is also commonly driven by the mapper's mental model of soil-landscape relationships. Survey sites are selected to gain the best coverage of certain features and variability in the landscape. This is a way to optimize the resources of a given survey but introduces bias to the data collection, and the sample distribution may not be suitable for statistical analysis (Carré et al., 2007).

Modern conventional soil mapping is also digital and completed using GIS programs, the end products of which are multi-layer vector maps with polygons representing different soil units and associated attribute tables listing soil classifications and attributes for each shape. The layers of landscape information guiding mappers come from up-to-date remote sensing methods such as LiDAR and spectral imaging. However, in principle and use, these maps are still more like the hand-drafted soil maps of the past when compared to predictive digital mapping outputs.

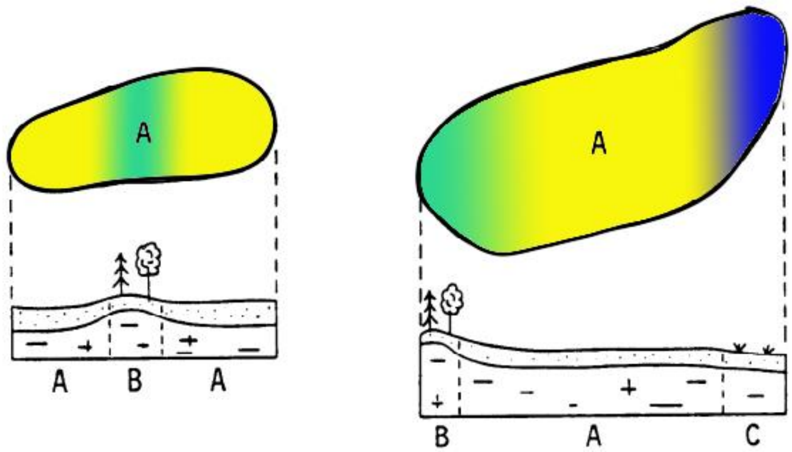

Figure 2. Image from a conventional mapping guide showing landscape transects with representation of how a simple soil unit would be represented as a polygon on a map (oval shapes). The predominant soil type must make up 75% or more of the unit, and minor components are aggregated. Modified from (Mapping System Working Group, 1981).

Conventional soil mapping is a well-established discipline that has served to improve our understanding of soil systems for many decades (Brevik et al., 2016). However, these traditional techniques have some inherent limitations. The necessity for sharp boundaries and generalization of detail required in mapping for a particular scale results in a simplified representation of the continuous variation of soil properties, and one or more soil types may be present in a mapped polygon (Grunwald, 2006b; Liu et al., 2016). It is difficult to extract quantitative data from conventional soil maps, where single values are assigned to entire sharply-bounded map units (Grunwald, 2006a). Measuring and reporting uncertainty or error is difficult to accomplish in conventional maps. Protocols have been established to record the amount of aggregation in a map and provide general estimates of uncertainty, but this cannot be quantified (Brevik et al., 2016). Conventional maps cannot be updated dynamically when new data becomes available. The area of interest would need to be re-mapped, requiring significant time and person-hours (Bouma, 1989). Soil properties such as soil organic carbon (SOC) change in a matter of years to decades (Mermut & Eswaran, 2001; Prescott, 2010) and are crucial to understanding global systems of carbon balance and biogeochemical cycling. It is not feasible to re-map on this timescale using conventional methods. Additionally, because the mapping process depends on the individual judgment of a mapper, a mapping process may not be completely reproducible.

4. Predictive Digital Soil Mapping

The exponential growth of computer processing, satellite data collection, software capability, and sharing of digital information has greatly impacted all earth science disciplines. Increasingly, scientists are challenged with becoming literate in coding, statistics, and data science as well as their chosen discipline, to be able to understand how to best apply the enormous computer power now at our disposal. The digital revolution has also made interdisciplinary and transdisciplinary work more possible with the potential for data sharing and integration worldwide if the hurdles of infrastructure, access, and standardization can be overcome. The challenges and benefits of this progress have been applied in the development of predictive digital soil mapping (PDSM).

PDSM uses numerical or statistical models that combine information about soil-forming factors (*scorpan*) obtained from spatial data layers with point soil data to estimate soil features for a map area. This technique can be used to generate maps of discrete features, like soil classifications, or continuous features like salinity, pH, and soil organic carbon. The outputs differ from conventional mapping, as they are gridded data represented by pixels in raster formats. This product is more suited to representing the spatial variability of soil and avoids the generalization required in conventional mapping (Hartemin & McBratney, 2008; McBratney et al., 2003b).

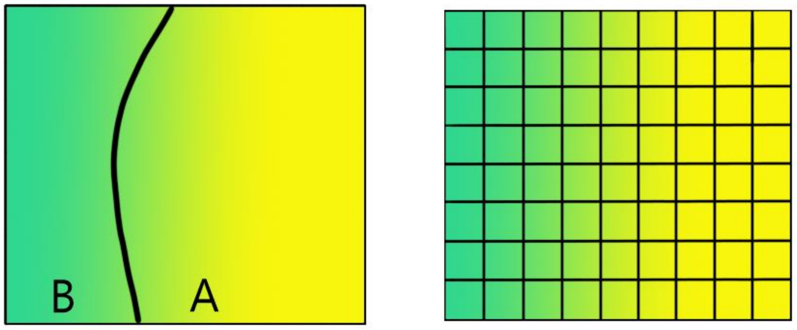

Figure 3. Representation of a conventional vector map with sharp boundaries between units (left) and a raster map with pixels that each have a unique value (right). It is simple to aggregate the raster map into larger units or classifications but much more complicated to disaggregate a vector map. After (Soil Science Division Staff (USDA-NRCS), n.d.).

The methodology used in PDSM can vary significantly, and academic reviews describe the particulars of different techniques with greater depth than will be covered here (Arrouays et al., 2017a; Florinsky, 2016; T. Hengl et al., 2019; Ma et al., 2019; McBratney et al., 2003a; Nussbaum et al., 2018; Scull et al., 2003; Wadoux et al., 2020). Despite differences in methods, the general process is the same (as illustrated in Figure 4):

- The objectives and parameters of the mapping are determined (target attributes, classifications, map extent, and map scale.)
- Data is collected (real soil observations and environmental covariates).
- A spatial prediction method is applied to the data to generate a model.
- The output of the model (i.e., the soil map) is assessed using validation procedures and critical evaluation by experts.
- If the outputs do not meet the necessary standards or do not achieve the goals of the project, further iterations of the mapping process will be completed.

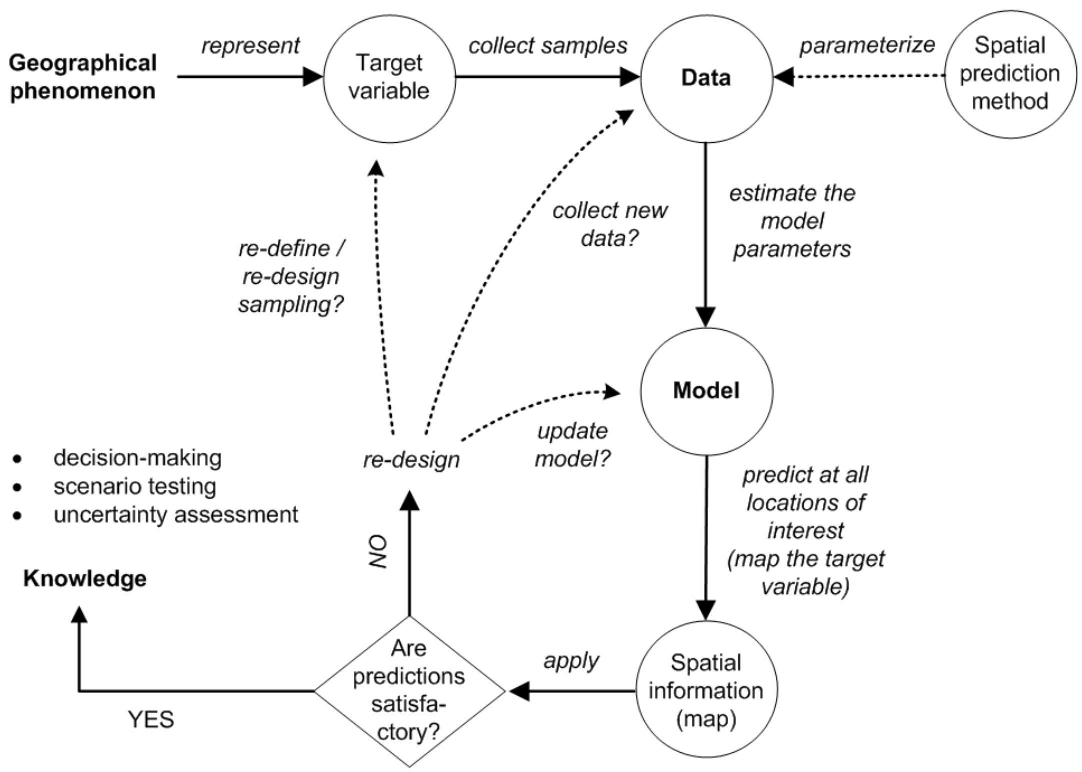

Figure 4. Generalized spatial prediction schematic. (T. Hengl et al., 2019)

The fundamental basis of the complex mathematical models driving soil predictions is the universal model of soil variation (T. Hengl et al., 2019):

$$Z(s) = m(s) + \varepsilon'(s) + \varepsilon''(s) \quad (3)$$

Each term represents a source of variation. The first term is the structural component or trend, determined by the soil formation factors and relationships between them (pedologic model). The second term describes the spatial correlation that produces variability; this term is treated as stochastic, meaning that it involves probability and is often solved for using regression models. The third term is to account for the variation from noise. Natural processes all have a degree of randomness that contributes to this term, along with errors from measurement and other sources. Understanding these three basic components and the different ways they introduce error and uncertainty is important to the conceptualization of what more complex and obscure predictive mapping methods are trying to achieve.

The types of prediction models used can be divided into three general categories: unsupervised covariate classification techniques; supervised methods such as taxonomic distance techniques, multinomial logistic regression techniques, and geostatistical approaches; and lastly, models based on qualitative information and expert knowledge (T. Hengl et al., 2019; Ma et al., 2019). The selection of a data-driven or knowledge-based modelling approach largely depends on the amount of available data and desired outcomes. Sophisticated data-driven statistical methods using recent developments in machine learning like artificial neural networks can produce accurate results (Brungard et al., 2015) and provide new insight into soil-landscape relationships by identifying relationships that were not recognized or anticipated (Ma et al., 2019). However, these "big data" or "black box" techniques must be used with caution due to potential issues with overfitting models to sparse soil data or selecting the simplest models instead of the most suitable from a pedological standpoint (Arrouays, McBratney, et al., 2020).

4.1. Data Sources for Predictive Digital Soil Mapping

4.1.1. Soil Information

Observations of soil conditions are necessary for all soil mapping. In PDSM, soil information from sample sites is used to train the predictive models – connecting real soil properties to the environmental factors. A second set of samples can also be used to check, or validate, the model results. The soil observations can come from new field collection or legacy surveys.

4.1.1.1. New soil data

In PDSM, the selection of soil sample locations is based on statistical methods, which are necessary to produce the most accurate results possible. These approaches include random samples, uniform grids, or more complex techniques like conditioned Latin hypercube – which can optimize the sample distribution across various units (Brus et al., 2011; Minasny & McBratney, 2006; Zhang et al., 2016). The precision and coverage required by these different sampling schemes can increase the cost of sampling programs in remote or inaccessible areas (Kidd et al., 2015). However, mappers have had success generating predictive maps with convenience-based sampling schemes that prioritize accessibility and increase the potential for establishing soil monitoring networks (Chartin et al., 2017).

New technology has significantly changed how soil information can be collected through improvements in remote sensing and proximal sensing tools. Portable devices such as spectrometers and magnetic susceptibility meters allow for characterization and estimation of soil properties in the field, saving time and lab analysis costs (Adamchuk et al., 2017; Horta et al., 2015; Silva et al., 2021). These advances complement PDSM as they allow for the collection of more data points, increasing the data available for modelling.

4.1.1.2. Legacy soil data

Legacy soil data refers to past soil survey maps, profiles, and other related soil analyses (Hengl et al., 2019). Digitizing this information has been a major focus of many government survey organizations to make it accessible as GIS layers in web-based data repositories. The georeferencing of this data, especially determining the precise location of point soil data (soil profiles), is critical to the successful use of legacy data in modern PDSM projects (Hartemin & McBratney, 2008; MacMillan & Hengl, 2019). Legacy data may also need to be harmonized if multiple methods of description and lab analysis have been used. Subjective or qualitative observations will have higher levels of uncertainty than quantitative field measurements. A mismatch between the age of different data sources (covariates, training vs. validation soil data) is another potential source of error, especially when attempting to model dynamic soil properties such as biological attributes (Brevik et al., 2016).

4.1.2. Environmental Covariates

Covariates, or environmental covariates, are spatial data layers related to the factors of soil formation. The mathematical relationship between soil characteristics and covariate data is what predictive models are calculating. Relief and other topography-related values are the foundation of a covariate dataset and are combined with many other features such as climate, geology, and biological factors (Table 1). The significance of certain covariates can depend on the location and extent of mapping and the quality of covariate data available. Geology, both bedrock and surficial, was the most important covariate in a new national map of Nepal (Lamichhane et al., 2021), while a review of over 40 studies in Iran determined that quantitative information such as terrain attributes and remote sensing data are more commonly used as covariates in that region, over qualitative layers such as geology and land use (Zeraatpisheh et al., 2020).

Table 1. Examples of Environmental Covariates (after Hengl et al., 2019)

Formation-factor	**Covariates**
Soil	Soil classification map, soil attributes map
Climate	Temperature, precipitation, snow cover, potential evapotranspiration
Organisms	Vegetation indices, land cover maps
Relief (topography)	DEM, slope, curvatures, flow models, landform classes, hydrological accumulation/deposition indices
Parent material	Bedrock and surficial geology maps, classifications and properties, volcanic activity

Formation-factor	Covariates
Age	Geological ages, ages of recent disturbances
Spatial position and context	Position, distance to landscape features (e.g. ocean, river, hills)
Human influences	Land use maps, probability/intensity of different land uses, soil disturbance or sealing, fertilization maps, infrastructure

Covariates must be selected and prepared carefully, depending on the map scale, extent, resolution, and purpose of mapping. The covariates used in a model can be selected using expert knowledge, especially in areas where the contributing factors to soil development are well understood. Pedological knowledge can also be useful in assessing whether the preparation of covariate layers is appropriate (e.g., extent of smoothing for a DEM) (Soil Science Division Staff (USDA-NRCS), 2017). Data-driven approaches can also be used to narrow down a covariate set from all available inputs and can lead to discoveries of new pedological relationships (Ma et al., 2019; Nussbaum et al., 2018).

4.2. Accuracy and Validation

"…defining and applying a common accuracy assessment framework is probably the greatest challenge of Digital Soil Mapping on its way toward practical applications." (Lagacherie, 2008)

Mapping accuracy refers to the difference between a predicted value and a "true" value. Accuracy assessments can be subjective, based on what is reasonable considering expert knowledge or calculated and assessed using objective approaches (Hengl et al., 2019).

Common pitfalls of PDSM that can impact map accuracy were identified by (McBratney et al., 2003b) and continue to be relevant in contemporary projects:

- Lack of sufficient soil observations that are appropriate for statistical modelling methods.
- Poor data quality, and inadequate or inappropriate data preparation.
- Overfitting models to data, either due to lack of data points or modelling methods. This is increasingly important to consider as data-driven methodologies using machine learning are the cutting edge of the research, but the field collection of soil samples

necessary to support these techniques has not been prioritized (MacMillan & Hengl, 2019).

- Circularity, as in covariates being derived from one DEM then used to derive more covariates and then all be combined in predicting the target soil variables.
- Discrepancies between databases and data mining techniques adopted by different organizations.

Validation of model results is key to evaluating and communicating map accuracy. The highest standard of validation is to have an independent set of soil observations collected using a statistical sampling method to compare to predicted values (Brus et al., 2011; Carré et al., 2007). However, funding restrictions mean that there is generally not a budget for this external sampling. A review of PDSM mapping of soil organic carbon assessing studies from 2013-2019 found that though all studies used objective validation techniques, only 9% performed external validation (Lamichhane et al., 2019). When new or external data is not available, validation is performed through data-splitting and cross-validation techniques where the original sample set used to generate the model is divided into calibration and validation sets, and then the analysis is repeated (T. Hengl et al., 2019). Cross-validation is cost-effective but may not be entirely suitable, especially in situations where the data was not collected using an unbiased sampling design (Lamichhane et al., 2019).

The recent proliferation of PDSM has led to the greater codification of analysis and validation methods through collaborative initiatives like the GlobalSoilMap (Arrouays, Poggio, et al., 2020) and widespread application of predictive techniques in different regions (Arrouays et al., 2017b; J L Boettinger et al., 2010; Chaney et al., 2016; Searle et al., 2021; van Zijl, 2019). However, contemporary critiques are still calling for the development of accuracy standards for PDSM products, a need that is more urgent as the software and data necessary to conduct this work has become more available to non-academic users. Without enforced minimum standards, there is a greater risk of the generation of poor quality map products in an applied setting like the private sector (Arrouays, McBratney, et al., 2020).

4.3. Uncertainty

"End users and decision makers frequently feel uneasy with uncertainty… most of them are unsure of how to use and communicate [it]" (Arrouays, McBratney, et al., 2020)

Uncertainty is a critical feature of all maps, conventional and digital. No map perfectly captures nature. They are a representation of our best estimate given the resources and understanding at a given moment in time. Sources of error such as sampling bias, location accuracy, measurement accuracy and error, and numerous other human errors are all sources of uncertainty in addition to random natural variation. The quantification and graphical representation of uncertainty is a core advantage of PDSM over conventional mapping, and the potential of this feature of predictive maps draws attention from end-users of soil data. Prediction uncertainty can be used to design optimized future sampling plans (Zhang et al., 2016), and reporting uncertainty is important to guiding land management decisions and completing risk assessments. However, understanding how best to interpret and apply quantitative uncertainty information to these situations is not entirely clear. When surveyed, soil map end-users expressed a strong interest in having uncertainty indicators reported but also suggested that they would need significant technical support to apply these findings (Richer-de-Forges et al., 2019).

Determination and representation of uncertainty in PDSM can be achieved with several methods, often depending on the modeling technique used. Validation through the collection and statistical analysis of a new set of soil samples is one way of determining the uncertainty of a soil map (Brus et al., 2011), but it is expensive and not always performed. Uncertainty reporting through direct geostatistical modelling of soil properties is more common but relies on the quality of the assumptions used in making the model (T. Hengl et al., 2019). Prediction intervals (PI) are a common technique. A 90% PI reports the upper and lower values within which the true value is expected to occur 9 times out of 10 (GlobalSoilMap Science Committee, 2015). If a PI is narrow (e.g. range of 5-6 units) the uncertainty is lower than if the PI is wide (e.g. range of 1-10 units). In addition to PI, maps of probability distributions or density functions, equiprobable simulations, and prediction error maps can be used (Hengl et al., 2019). The complexity of these determinations of uncertainty depends on the type of soil data being investigated (e.g., continuous values, categorical data, narrative data). (Heuvelink & Brown, 2006) proposed a framework for handling uncertainty in soil information databases that accounts for multiple types of data.

Understanding uncertainty in soil mapping is not trivial. Even the best possible predictive maps made by experts using the best available data may not be able to explain a significant amount of the variation. This is especially true for smaller survey areas where strong regional trends in topography, geology, and climate are not present to explain the soil variability. In such cases, explaining only 50% of the variation of the target variable modeled may be considered a good result (Cook et al., 2008; Kempen et al., 2011)

5. Application of Predictive Digital Soil Mapping

Predictive mapping has been applied at all scales, from global and continental maps to field-scale evaluations in support of precision agriculture (J L Boettinger et al., 2010). The key examples selected focus on how predictive mapping techniques have been integrated with established soil surveys or agencies operating on a regional or national scale.

5.1. Predictive Soil Mapping in Australia

In the 1990s, the recognition of soil degradation in agricultural areas of Western Australia renewed soil survey efforts to improve soil information coverage at smaller map scales (1:250,000 or finer). Most of this mapping was undertaken using conventional methods (Cook et al., 2008). However, during this reinvigoration of soil survey in the country, predictive digital mapping projects were initiated, and the country has since become a leader in the operationalization of predictive mapping and the development of these techniques (Kidd et al., 2020; Lamichhane et al., 2019; Searle et al., 2021). Key figures in the development of modern PDSM, such as Dr. Alexander McBratney, are also based in Australia.

The re-mapping of the Murray-Darling Basin from 1995-1998 using spatial modelling techniques to fill in gaps in the conventional survey is one of the earliest examples of applied PDSM and has become a seminal study in the field (Bui & Moran, 2003). This project used the expert knowledge from existing soil maps to generate classification rules for the rest of the basin. During modeling, training data pixels were matched with unmapped areas based on the similarities in environmental conditions with the assumption that the same factors would be controlling soil properties in each location. Despite the presence of complicated soil-landscape relationships, the replication of mental models from historic maps was generally successful. Accuracy assessments of the final predictions had more mixed results (Figure 5), and were restricted by the lack of new sampling, but the overall project was considered a success.

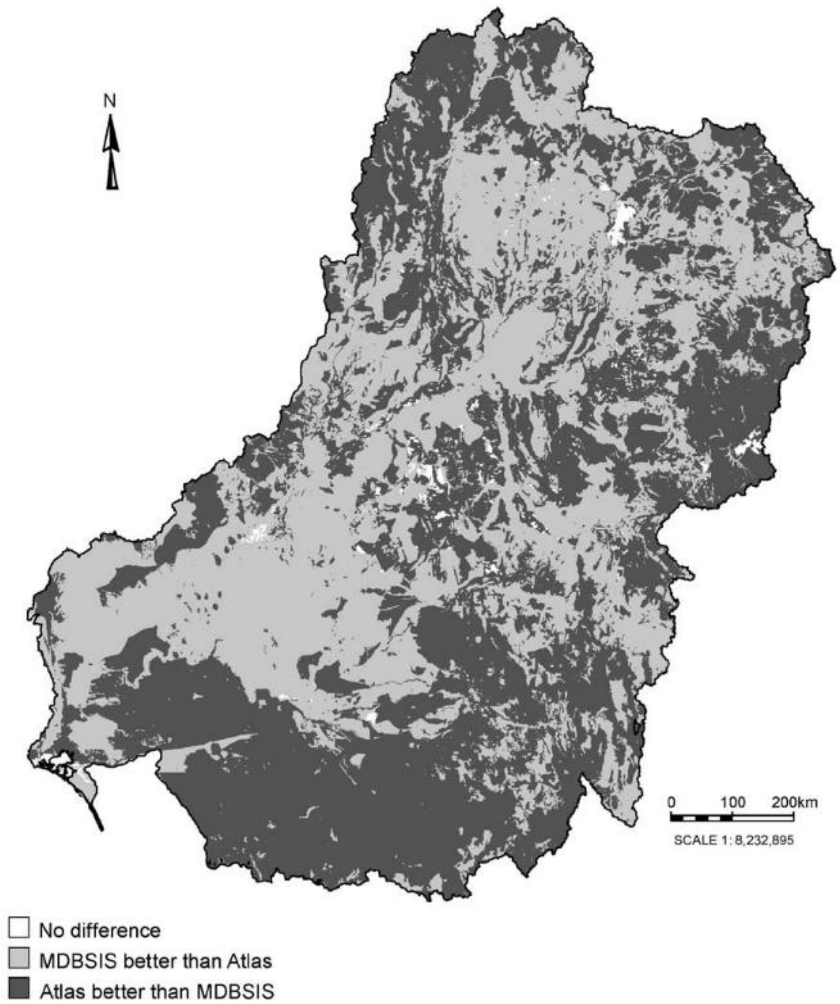

Figure 5. Comparison of the Digital Atlas of Australian soils (former conventional map) and the new Murray-Darling Basin Soil Information System (MDBSIS) (Bui & Moran, 2003).

Though there was some integration with surveys, PDSM in Australia was still largely academic until approximately ten years ago (Kidd et al., 2020). Since then, more widespread application of PDSM has occurred. The Soil and Landscape Grid of Australia (SLGA) (Grundy et al., 2015) is a publicly funded nationwide map of soil attributes at a 90-m resolution. The data used to generate the map are historic soil data, with new measurements using spectroscopic sensors and environmental covariates. The map includes estimates for six standard depths with a 90% prediction interval and complies with the GlobalSoilMap standards (Kidd et al., 2020). Through the online data portal, national and regional maps of soil and landscape attributes are freely available for download.

5.2. Predictive Soil Mapping in the United States

Soil surveys in the United States began in the 1890s, and the "modern" soil survey started in the 1950s when the United States Department of Agriculture Natural Resources Conservation Service (USDA-NRCS, originally called the Soil Conservation Service) was established, and maps were consistently produced using air photos as a base. These surveys covered 85% of the United States by the 1990s (Indorante et al., 1996). However, at that time, advancements in understanding of soil and environmental systems, land use changes, and technological improvements caused the traditional soil survey reports to be considered outdated. Increasingly, the modernization of soil survey included calls for accessible systems with common standards that can readily incorporate new data (Brevik et al., 2016; Grunwald, 2006a; Indorante et al., 1996). PDSM has been accepted as the solution to these issues.

Early work in the United States consisted of local or regional PDSM projects and test sites. The USDA-NRCS experimented with developing digital soil-landscape modelling in California to serve as an aid for conventional soil mappers to more efficiently plan their field work (Howell et al., 2008). Topography and satellite imagery (ASTER) were the primary covariates used, and several soil parameters were modelled. Not all modelling was successful, but the modelled depth to secondary carbonates and estimates for particle size classes were sufficiently accurate for over 70% of the test points (Howell et al., 2008). Embedding some PDSM evaluation within an operating conventional survey is mutually beneficial; field mappers gain useful insight and then are also able to independently validate the predictions through their subsequent data collection and mapping.

In the years since the California study, the advancement of PDSM and the establishment of more consistent standards have led to the greater acceptance of digital mapping as a stand-alone product. A number of ambitious PDSM projects have been completed or are currently in development.

- POLARIS is a map of soil series probabilities at a 30-m spacing that has complete coverage of the contiguous United States (Chaney et al., 2016). The map was produced using a machine learning algorithm (DSMART-HPC) to combine high-resolution environmental covariates with conventional soil map data from the Soil Survey Geographic (SSURGO) database. The intent of this mapping was to fill in gaps in the SSURGO coverage, remove the existing discontinuities at political boundaries, and disaggregate large polygons. The results were validated using the NRCS National Soil Information System (NASIS), a database of soil observations. Most of these observations were used to produce the SSURGO map units, so it is not an independent dataset, but no other appropriate set of observations exist. The uncertainty of the predicted classification is fairly high for large areas of the map (see Fig. 5 in Chaney et al., 2016), which has been attributed to having insufficient covariates. The initial accuracy assessment showed that the accuracy was highly variable throughout the map. The POLARIS map has since been refined to improve the accuracy of soil series predictions and also generate soil property maps (Chaney et al., 2019). This new map is able to explain 41% of the variation of soil properties, on average, with pH, organic matter, and fractions of sand, silt, and clay being the most accurately predicted.
- (Ramcharan et al., 2018) used a different approach to mapping soil classifications and properties of the conterminous United States at a 100-m resolution. Instead of disaggregating polygons like POLARIS, point data from three national datasets was used, along with an extensive set of environmental covariates. Cross-validation of soil property results indicated similar to slightly better results than POLARIS. Cross-validation estimated that prediction accuracy was 60% for soil great groups (GG) and 66% for modified particle size classes (mPSCs). Independent validation was also completed and found that accuracy was between 24-58% for GG and 24-93% of mPSCs.

- The USDA-NRCS has recently announced that they will be completing a new inventory of soils for the entire United States, using PDSM technology to be completed by 2026 (Thompson et al., 2020). Soils2026 will generate 30-m resolution products predicting 12 key soil properties at six depth intervals, following the standards set by the GlobalSoilMap project.

5.3. Predictive Ecosystem Mapping in British Columbia, Canada

Predictive techniques have also been operationalized in fields outside of soil mapping. Predictive ecosystem mapping (PEM) uses the same principles of PDSM, with environmental covariates being linked to ecosystem rather than soil classifications. The predictive ecosystem mapping project in British Columbia, Canada, is one of the earliest examples of predictive modeling standards being established for a government-backed mapping initiative (Resources Inventory Committee (Canada). Terrestrial Ecosystem Mapping Alternatives Task Force., 1999). By early 2005, 53% of the province had been mapped with predictive ecosystem mapping, compared to only 15% with conventional methods. The results of this mapping are available in a public database. Applications of this ecosystem data include site productivity assessment, wildlife habitat and biodiversity interpretations, rare and special ecosystem inventories, riparian management area interpretations, archaeological overview assessments, and support responsible forest management (Cortex Consultants Inc. and JMJ Holdings Inc., 2005; Forest Service of British Columbia, n.d.). An 8.2 million ha 1:20,000 scale map in the former Caribou Forest Region was completed as part of this initiative. This study generated a predictive map using only existing knowledge and data. Independently classified field data was then collected to for validation of the model results. The final PEM maps were found to have an average accuracy of 69% across the study area, which was deemed an acceptable result (MacMillan et al., 2010). In addition to this technical success, it was reported that end-users of the data enthusiastically accepted the results once the utility of predictive mapping methods was communicated.

6. Future Opportunities and Barriers to Adoption

PDSM has made major strides since its inception, with exponential increases in map extent and resolution and significant improvements in accuracy (Minasny & McBratney, 2016). The potential for the dynamic re-mapping of areas as new data is collected makes PDSM applicable to adaptive management techniques that are becoming more widespread (Allen et al., 2011; Cook et al., 2008). However, certain soil features are not yet being modelled by PDSM or are difficult to map accurately (Richer-de-Forges et al., 2019). Highly dynamic soil properties related to hydrology like permeability and infiltration capacity are in high demand by end-users but are not being produced in PDSM (Figure 6). Additionally, current PDSM techniques are not modelling system dynamics or able to predict changes in soil properties (Ma et al., 2019). This is an area of potential expansion that has also been identified as desirable by end-users in research, agriculture, and other natural resource industries (Richer-de-Forges et al., 2019).

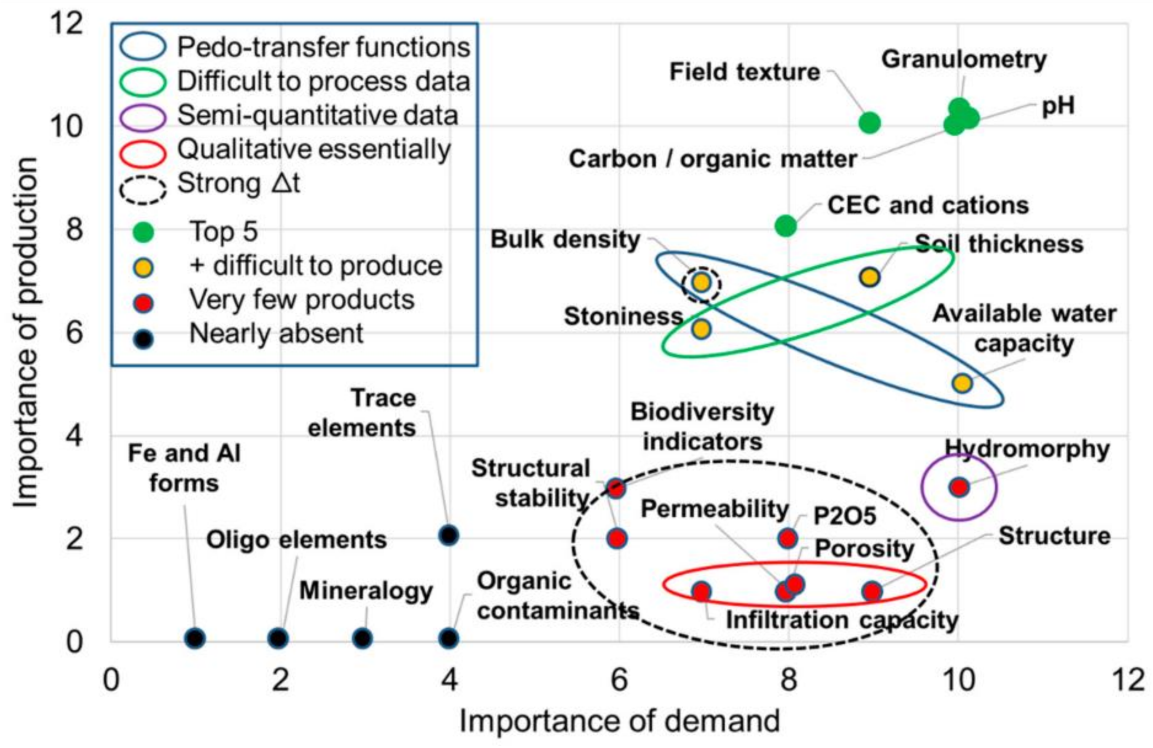

Figure 6. Correlation between soil attributes prioritized by end-users (x-axis) and attributes being produced by soil mappers (y-axis) (Richer-de-Forges et al., 2019)

There are numerous challenges facing the progression of PDSM into applied use; some of the most critical are outlined here.

Access to data and need for new data collection:

PDSM requires large amounts of high-quality soil data for models to be as accurate as possible and to complete independent model validations. When legacy soil data is used with new covariate data layers, predictions may not be accurate, and the uncertainty is compounded with original errors of data collection (Hartemin & McBratney, 2008). Large-scale mapping programs can benefit from new in-situ measurement technologies to collect soil data more rapidly and cheaply. For local or regional maps, new data could be obtained from private industry as companies are commonly required to collect baseline soil data and routine monitoring for environmental impact assessments, but it is difficult to convince stakeholders to make this data available (Drozdowski et al., 2019).

The use of new data is also essential for developing and progressing soil knowledge and our understanding of soil systems. There is a close relationship between soil mapping, classifications, and pedologic models. Advances in one area will necessarily change the others (Brevik et al., 2016).

Map quality standards:

Challenges in creating and adopting unified international standards of soil classification and analysis persist, and the World Reference Base (WRB) for soil classification only started being accepted as the default system, over previous national systems, towards the end of the 20th century (Mermut & Eswaran, 2001). PDSM faces similar challenges. Having a robust and consistent framework of minimum quality standards for PDSM outputs is essential. Users should be able to easily check and understand the validity and credibility of the data, then move on to the important work of figuring out what to do with it.

Though many thorough descriptions of PDSM methodology and workflow are available, there is yet to be an accepted and enforced minimum standard of practice. This is not a simple undertaking due to the many permutations of data sources, modelling method, map scale and purpose, and beyond. Standards have been developed for significant large-scale PDSM endeavours (WorldSoilMap) that are now being applied to more regional projects. This is a good starting point.

Using unfamiliar types of mapping products:

The representation of the variability of soil across landscapes using raster formats, rather than units with clearly delineated boundaries as in conventional mapping, produces a mismatch between how many soil map end-users are used to conceptualizing soil data. Complicated, gradational information is not intuitive and may not be what land managers or administrators need. Maps of both basic soil properties, as well as higher-level interpretations such as agricultural suitability, should be generated by map makers instead of leaving all the interpretation to users that may be less familiar with pedology or the limitations of a specific data set. (Arrouays, McBratney, et al., 2020)

Similarly, quantifications of uncertainty are lauded as a great advantage of digital over conventional mapping. However, communicating what these values mean in a practical way to more general audiences is not adequately covered by the current discourse. Quantified uncertainty or probability can also have legal implications for decisions made using this data – this is not yet well understood (Tomislav Hengl & MacMillan, 2019).

7. Recommendations

Based on the discussion above, some key recommendations can be made:

- The collection of new soil data for both primary mapping and map validation is critical to producing the best and most accurate products. Recent data is especially important for understanding dynamic soil properties that are currently difficult to model.

- Further development and enforcement of minimum quality standards is necessary as PDSM becomes more applied. WorldSoilMap is a good starting point for large mapping projects, but standards for more local or regional projects are less clear.

- Clear and accessible communication of both the potential and limitations of PDSM programs should be prioritized. This includes audiences within the soil science community – improving training and knowledge-building in new methods – as well as end users of soil data and stakeholders such as government bodies and private industry. Feedback from end-users about the ease of use and applicability of PDSM outputs should direct future work.

8. Conclusion

Technological advancement (GIS, remote sensing, computer processing) has enabled soil mapping to become more quantitative. There is a demand for high-resolution quantitative soil data in many disciplines from end-users in academia, private industry, and government. Predictive digital soil mapping can provide gridded maps with continuous values for soil properties and predict discrete soil classifications, with quantitative estimates of uncertainty for the results.

Techniques of predictive digital soil mapping are sufficiently developed to be applied outside of research contexts, and these methods have been increasingly adopted over the last decade. Most contemporary soil mapping projects are already using predictive mapping techniques in some capacity. However, mapping projects are often limited by the amount and quality of the soil data available. Therefore, instead of asking if we should attempt predictive mapping, the more significant question is, are we willing to invest in what is necessary to produce the highest quality and most accurate maps using these techniques? This could mean new sampling campaigns in support of updating outdated or sparse soil surveys or accessing recent soil information collected by natural resource or agriculture projects.

9. References

Adamchuk, V., Allred, B., Doolittle, J., Grote, K., & Viscarra Rossel, R. (2017). Tools for Proximal Soil Sensing. In *Soil Survey Manual* (USDA Handb, pp. 355–394). Government Printing Office, Washington, D.C.

Allen, C. R., Fontaine, J. J., Pope, K. L., & Garmestani, A. S. (2011). Adaptive management for a turbulent future. *Journal of Environmental Management*, *92*(5), 1339–1345. https://doi.org/10.1016/j.jenvman.2010.11.019

Arrouays, D., Lagacherie, P., & Hartemink, A. E. (2017a). Digital soil mapping across the globe. *Geoderma Regional*, *9*, 1–4. https://doi.org/10.1016/j.geodrs.2017.03.002

Arrouays, D., Lagacherie, P., & Hartemink, A. E. (2017b). Digital soil mapping across the globe. In *Geoderma Regional* (Vol. 9, pp. 1–4). Elsevier B.V. https://doi.org/10.1016/j.geodrs.2017.03.002

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*, *20*. https://doi.org/10.1016/j.geodrs.2020.e00255

Arrouays, D., Poggio, L., Salazar Guerrero, O. A., & Mulder, V. L. (2020). Digital soil mapping and GlobalSoilMap. Main advances and ways forward. *Geoderma Regional*, *21*, e00265. https://doi.org/10.1016/j.geodrs.2020.e00265

Boettinger, J L, Howell, D., Moore, A. C., Hartemink, A. E., & Kienast-Brown, S. (2010). Digital Soil Mapping: Bridging Research, Environmental Application, and Operation. In *Progress in Soil Science*. Springer. https://doi.org/10.1007/978-90-481-8863-5_19

Bouma, J. (1989). *Using Soil Survey Data for Quantitative Land Evaluation BT - Advances in Soil Science: Volume 9* (B. A. Stewart (ed.); pp. 177–213). Springer US. https://doi.org/10.1007/978-1-4612-3532-3_4

Bouma, J. (2014). Soil science contributions towards Sustainable Development Goals and their implementation: Linking soil functions with ecosystem services. *Journal of Plant Nutrition and Soil Science*, *177*(2), 111–120. https://doi.org/10.1002/jpln.201300646

Brady, N. C., & Weil, R. R. (2008). *The nature and properties of soils* (Rev. 14th). Pearson Prentice Hall.

Brevik, E. C., Calzolari, C., Miller, B. A., Pereira, P., Kabala, C., Baumgarten, A., & Jordán, A. (2016). Soil mapping, classification, and pedologic modeling: History and future directions. *Geoderma*, *264*, 256–274. https://doi.org/10.1016/j.geoderma.2015.05.017

Brungard, C., Boettinger, J., Duniway, M., Wills, S., & Jr, T. (2015). Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma*, *s 239–240*, 68–83. https://doi.org/10.1016/j.geoderma.2014.09.019

Brus, D. J., Kempen, B., & Heuvelink, G. B. M. (2011). Sampling for validation of digital soil maps. *European Journal of Soil Science*, *62*(3), 394–407. https://doi.org/10.1111/j.1365-2389.2011.01364.x

Bui, E. N., & Moran, C. J. (2003). A strategy to fill gaps in soil survey over large spatial extents: An example from the Murray-Darling Basin of Australia. *Geoderma*, *111*(1–2), 21–44. https://doi.org/10.1016/S0016-7061(02)00238-0

Carré, F., McBratney, A. B., & Minasny, B. (2007). Estimation and potential improvement of the quality of legacy soil samples for digital soil mapping. *Geoderma*, *141*(1–2), 1–14. https://doi.org/10.1016/j.geoderma.2007.01.018

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*(4), 2916–2938. https://doi.org/10.1029/2018WR022797

Chaney, N. W., Wood, E. F., McBratney, A. B., Hempel, J. W., Nauman, T. W., Brungard, C. W., & Odgers, N. P. (2016). POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. *Geoderma*, *274*, 54–67. https://doi.org/10.1016/j.geoderma.2016.03.025

Chartin, C., Stevens, A., Goidts, E., Krüger, I., Carnol, M., & van Wesemael, B. (2017). Mapping Soil Organic Carbon stocks and estimating uncertainties at the regional scale following a legacy sampling strategy (Southern Belgium, Wallonia). *Geoderma Regional*, *9*, 73–86. https://doi.org/10.1016/j.geodrs.2016.12.006

Cook, S. E., Jarvis, A., & Gonzalez, J. P. (2008). *A New Global Demand for Digital Soil Information BT - Digital Soil Mapping with Limited Data* (A. E. Hartemink, A. McBratney, & M. de L. Mendonça-Santos (eds.); pp. 31–41). Springer Netherlands. https://doi.org/10.1007/978-1-4020-8592-5_3

Cortex Consultants Inc. and JMJ Holdings Inc. (2005). *Predictive Ecosystem Mapping User's Guide*.

Doran, J. W. (2002). Soil health and global sustainability: translating science into practice. *Agriculture, Ecosystems & Environment*, *88*(2), 119–127. https://doi.org/https://doi.org/10.1016/S0167-8809(01)00246-8

Drozdowski, B., Aumann, C., & Powter, C. (2019). *Predictive Soil Mapping Seminar: Summary Report*.

Florinsky, I. V. (2012). The Dokuchaev hypothesis as a basis for predictive digital soil mapping (on the 125th anniversary of its publication). *Eurasian Soil Science*, *45*(4), 445–451. https://doi.org/10.1134/S1064229312040047

Florinsky, I. V. (2016). Predictive Soil Mapping. *Digital Terrain Analysis in Soil Science and Geology*, 287–311. https://doi.org/10.1016/b978-0-12-804632-6.00011-0

Forest Service of British Columbia. (n.d.). *Biogeoclimactic Ecosystem Classification Program*. Retrieved April 20, 2021, from https://www.for.gov.bc.ca/hre/becweb/system/how/index.html

GlobalSoilMap Science Committee. (2015). Specifications Tiered GlobalSoilMap products. *GlobalSoilMap Project*, *2.4*, 52. http://www.globalsoilmap.net/

Gregorich, E. G., Carter, M. R., & Books, E. A. A. (1997). *Soil quality for crop production and ecosystem health* (Vol. 25). Elsevier.

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*(8), 835–844. https://doi.org/10.1071/SR15191

Grunwald, S. (2006a). Future of Soil Science. In A. E. Hartemink (Ed.), *The Future of Soil Science* (pp. 51–53). IUSS International Union of Soil Sciences.

Grunwald, S. (2006b). What do we really know about the space-time continuum of soil-landscapes? In *Environmental Soil-Landscape Modeling: Geographic Information Technologies and Pedometrics* (pp.

3–36). CRC Press. https://doi.org/10.1201/9781420028188.sec1

Hartemin, & McBratney. (2008). *Digital Soil Mapping With Limited Data*.

Hengl, T., Heuvelink, G. B. ., & MacMillan, R. A. (2019). Statistical theory for predictive soil mapping. In *Predictive Soil Mapping With R*. https://soilmapper.org/statistical-theory.html

Hengl, Tomislav, & MacMillan, R. A. (2019). *Predictive Soil Mapping with R*. OpenGeoHub foundation. www.soilmapper.org

Heuvelink, G. B. M., & Brown, J. D. (2006). Chapter 8 Towards a Soil Information System for Uncertain Soil Data. *Developments in Soil Science*, *31*(C), 97–106. https://doi.org/10.1016/S0166-2481(06)31008-2

Horta, A., Malone, B., Stockmann, U., Minasny, B., Bishop, T. F. A., McBratney, A. B., Pallasser, R., & Pozza, L. (2015). Potential of integrated field spectroscopy and spatial analysis for enhanced assessment of soil contamination: A prospective review. *Geoderma*, *241–242*, 180–209. https://doi.org/10.1016/j.geoderma.2014.11.024

Howell, D., Kim, Y. G., & Haydu-Houdeshell, C. A. (2008). *Development and Application of Digital Soil Mapping Within Traditional Soil Survey: What will it Grow Into? BT - Digital Soil Mapping with Limited Data* (A. E. Hartemink, A. McBratney, & M. de L. Mendonça-Santos (eds.); pp. 43–51). Springer Netherlands. https://doi.org/10.1007/978-1-4020-8592-5_4

Hudson, B. D. (1992). The Soil Survey as Paradigm-based Science. *Soil Science Society of America Journal*, *56*(3), 836–841. https://doi.org/10.2136/sssaj1992.03615995005600030027x

Indorante, S. J., McLeese, R. L., Hammer, R. D., Thompson, B. W., & Alexander, D. L. (1996). Positioning soil survey for the 21st century. *Journal of Soil and Water Conservation*, *51*(1), 21–28.

Jenny, H. (1941). *Factors of soil formation; a system of quantitative pedology,*. McGraw-Hill.

Kempen, B., Brus, D. J., & Stoorvogel, J. J. (2011). Three-dimensional mapping of soil organic matter content using soil type-specific depth functions. *Geoderma*, *162*(1–2), 107–123. https://doi.org/10.1016/j.geoderma.2011.01.010

Kidd, D., Malone, B., McBratney, A., Minasny, B., & Webb, M. (2015). Operational sampling challenges to digital soil mapping in Tasmania, Australia. *Geoderma Regional*, *4*, 1–10. https://doi.org/https://doi.org/10.1016/j.geodrs.2014.11.002

Kidd, D., Searle, R., Grundy, M., McBratney, A., Robinson, N., O'Brien, L., Zund, P., Arrouays, D., Thomas, M., Padarian, J., Jones, E., Bennett, J. M. L., Minasny, B., Holmes, K., Malone, B. P., Liddicoat, C., Meier, E. A., Stockmann, U., Wilson, P., … Triantafilis, J. (2020). Operationalising digital soil mapping – Lessons from Australia. *Geoderma Regional*, *23*. https://doi.org/10.1016/j.geodrs.2020.e00335

Lagacherie, P. (2008). *Digital Soil Mapping: A State of the Art BT - Digital Soil Mapping with Limited Data* (A. E. Hartemink, A. McBratney, & M. de L. Mendonça-Santos (eds.); pp. 3–14). Springer Netherlands. https://doi.org/10.1007/978-1-4020-8592-5_1

Lal, R. (2007). Soil Science and the Carbon Civilization. *Soil Science Society of America Journal*, *71*(5), 1425–1437. https://doi.org/10.2136/sssaj2007.0001

Lamichhane, S., Kumar, L., & Adhikari, K. (2021). Updating the national soil map of Nepal through digital soil mapping. *Geoderma*, *394*, 115041. https://doi.org/https://doi.org/10.1016/j.geoderma.2021.115041

Lamichhane, S., Kumar, L., & Wilson, B. (2019). Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. *Geoderma*, *352*(June), 395–413. https://doi.org/10.1016/j.geoderma.2019.05.031

Liu, F., Geng, X., Zhu, A.-X., Fraser, W., Song, X.-D., & Ganlin, Z. (2016). Soil polygon disaggregation through similarity-based prediction with legacy pedons. *Journal of Arid Land*, *8*. https://doi.org/10.1007/s40333-016-0087-7

Ma, Y., Minasny, B., Malone, B. P., & Mcbratney, A. B. (2019). Pedology and digital soil mapping (DSM). *European Journal of Soil Science*, *70*(2), 216–235. https://doi.org/10.1111/ejss.12790

Macmillan, R. A. (2012). History and Evolution of Digital (Predictive) Soil Mapping. *Canadian Land Resource Network (CLRN) Annual Meeting*.

MacMillan, R. A., & Hengl, T. (2019). The future of predictive soil mapping. In *Predictive Soil Mapping With R*. https://soilmapper.org/the-future-of-predictive-soil-mapping.html

MacMillan, R. A., Moon, D. E., Coupé, R. A., & Phillips, N. (2010). *Predictive Ecosystem Mapping (PEM) for 8.2 Million ha of Forestland, British Columbia, Canada BT - Digital Soil Mapping: Bridging Research, Environmental Application, and Operation* (Janis L Boettinger, D. W. Howell, A. C. Moore, A. E. Hartemink, & S. Kienast-Brown (eds.); pp. 337–356). Springer Netherlands. https://doi.org/10.1007/978-90-481-8863-5_27

Mapping System Working Group. (1981). *A soil mapping system for Canada: Revised* (Issue 142).

McBratney, A. B., Mendonça Santos, M. L., & Minasny, B. (2003a). On digital soil mapping. In *Geoderma* (Vol. 117, Issues 1–2). https://doi.org/10.1016/S0016-7061(03)00223-4

McBratney, A. B., Mendonça Santos, M. L., & Minasny, B. (2003b). On digital soil mapping. *Geoderma*, *117*(1–2), 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4

McSweeney, K., Slater, B. K., Hammer, R. D., Bell, J. C., Gessler, P. E., & Peterson, G. W. (1994). Towards a New Framework for Modeling the Soil-Landscape Continuum. In *Factors of Soil Formation: A Fiftieth Anniversary Retrospective* (SSSA Speci). Soil Science Society of America.

Mermut, A. R., & Eswaran, H. (2001). Some major developments in soil science since the mid-1960s. *Geoderma*, *100*(3–4), 403–426. https://doi.org/10.1016/S0016-7061(01)00030-1

Minami, K. (2009). Soil and humanity: Culture, civilization, livelihood and health. *Soil Science and Plant Nutrition*, *55*(5), 603–615. https://doi.org/10.1111/j.1747-0765.2009.00401.x

Minasny, B., & McBratney, A. B. (2006). A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers and Geosciences*, *32*(9), 1378–1388. https://doi.org/10.1016/j.cageo.2005.12.009

Minasny, B., & McBratney, A. B. (2016). Digital soil mapping: A brief history and some lessons. *Geoderma*, *264*, 301–311. https://doi.org/10.1016/j.geoderma.2015.07.017

Nolan, M., Stanton, K. J., Evans, K., Pym, L., Kaufman, B., & Duley, E. (2021). From the ground up: prioritizing soil at the forefront of ecological restoration. *Restoration Ecology*, 1–5. https://doi.org/10.1111/rec.13453

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), 1–22. https://doi.org/10.5194/soil-4-1-2018

Prescott, C. E. (2010). Litter decomposition: What controls it and how can we alter it to sequester more carbon in forest soils? *Biogeochemistry*, *101*(1), 133–149. https://doi.org/10.1007/s10533-010-9439-0

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*(1), 186–201. https://doi.org/10.2136/sssaj2017.04.0122

Resources Inventory Committee (Canada). Terrestrial Ecosystem Mapping Alternatives Task Force. (1999). *Standards for predictive ecosystem mapping : inventory standard*. Resources Inventory Committee.

Richer-de-Forges, A. C., Arrouays, D., Bardy, M., Bispo, A., Lagacherie, P., Laroche, B., Lemercier, B., Sauter, J., & Voltz, M. (2019). Mapping of soils and land-related environmental attributes in France: Analysis of End-Users' needs. *Sustainability (Switzerland)*, *11*(10), 1–15. https://doi.org/10.3390/su11102940

Ruhe, R. V. (1975). *Geomorphology: Geomorphic processes and surficial geology*. Houghton Mifflin.

Scull, P., Franklin, J., Chadwick, O. A., & McArthur, D. (2003). Predictive soil mapping: A review. *Progress in Physical Geography*, *27*(2), 171–197. https://doi.org/10.1191/0309133303pp366ra

Searle, R., McBratney, A., Grundy, M., Kidd, D., Malone, B., Arrouays, D., Stockman, U., Zund, P., Wilson, P., Wilford, J., Van Gool, D., Triantafilis, J., Thomas, M., Stower, L., Slater, B., Robinson, N., Ringrose-Voase, A., Padarian, J., Payne, J., … Andrews, K. (2021). Digital soil mapping and assessment for Australia and beyond: A propitious future. *Geoderma Regional*, *24*. https://doi.org/10.1016/j.geodrs.2021.e00359

Silva, S. H. G., Weindorf, D. C., Faria, W. M., Pinto, L. C., Menezes, M. D., Guilherme, L. R. G., & Curi, N. (2021). Proximal sensor-enhanced soil mapping in complex soil-landscape areas of Brazil. *Pedosphere*, *31*(4), 615–626. https://doi.org/10.1016/S1002-0160(21)60007-3

Simonson, R. W. (1959). Outline of a Generalized Theory of Soil Genesis. *Soil Science Society of America Journal*, *23*(2), 152–156. https://doi.org/10.2136/sssaj1959.03615995002300020021x

Soil Science Division Staff (USDA-NRCS). (n.d.). *Digital Soil Mapping (DSM)*. https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/geo/?cid=stelprdb1254424

Soil Science Division Staff (USDA-NRCS). (2017). *Soil Survey Manual, USDA Handbook 18*. *18*, 639.

Thompson, J. A., Kienast-Brown, S., D'Avello, T., Philippe, J., & Brungard, C. (2020). Soils2026 and digital soil mapping – A foundation for the future of soils information in the United States. *Geoderma Regional*, *22*, e00294. https://doi.org/10.1016/j.geodrs.2020.e00294

van Zijl, G. (2019). Digital soil mapping approaches to address real world problems in southern Africa. *Geoderma*, *337*(November 2017), 1301–1308. https://doi.org/10.1016/j.geoderma.2018.07.052

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*(January), 103359. https://doi.org/10.1016/j.earscirev.2020.103359

Webster, R., & Oliver, M. A. (2007). *Geostatistics for Environmental Scientists* (2nd ed.). John Wiley & Sons, Inc.

Zhang, S. J., Zhu, A. X., Liu, J., Yang, L., Qin, C. Z., & An, Y. M. (2016). An heuristic uncertainty directed field sampling design for digital soil mapping. *Geoderma*, *267*, 123–136. https://doi.org/10.1016/j.geoderma.2015.12.009