

# Perspectives on validation in digital soil mapping of continuous attributes—A review

Kristin Piikki<sup>1,2</sup> | Johanna Wetterlind<sup>1</sup> | Mats Söderström<sup>1</sup> | Bo Stenberg<sup>1</sup>

<sup>1</sup>Swedish University of Agricultural Sciences (SLU), Skara, Sweden

<sup>2</sup>International Center for Tropical Agriculture (CIAT), Nairobi, Kenya

## Correspondence

Kristin Piikki, Swedish University of Agricultural Sciences (SLU), Skara, Sweden

Email: kristin.piikki@slu.se

## Abstract

We performed a systematic mapping of validation methods used in digital soil mapping (DSM), in order to gain an overview of current practices and make recommendations for future publications on DSM studies. A systematic search and screening procedure, largely following the *RepOrting standards for Systematic Evidence Syntheses* (ROSES) protocol, was carried out. It yielded a database of 188 peer-reviewed DSM studies from the past two decades, all written in English and all presenting a raster map of a continuous soil property. Review of the full-texts showed that most publications (97%) included some type of map validation, while just over one-third (35%) estimated map uncertainty. Most commonly, a combination of multiple (existing) soil sampe datasets was used and the resulting maps were validated by single data-splitting or cross-validation. It was common for essential information to be lacking in method descriptions. This is unfortunate, as lack of information on sampling design (missing in 25% of 188 studies) and sample support (missing in 45% of 188 studies) makes it difficult to interpret what derived validation metrics represent, compromising their usefulness. Therefore, we present a list of method details that should be provided in DSM studies. We also provide a detailed summary of the 28 validation metrics used in published DSM studies, how to interpret the values obtained and whether the metrics can be compared between datasets or soil attributes.

## KEYWORDS

accuracy, digital soil mapping, systematic map, uncertainty, validation

## 1 | INTRODUCTION

A *systematic map* is a summary of a research area based on systematic searches, screening and coding of scientific literature (James et al., 2016). It aims to give an overview of a subject area and identify trends, knowledge gaps and knowledge clusters in scientific reporting. In contrast to a *systematic review*, the aim is not to answer specific close-framed

questions, but rather to describe the state of knowledge in the subject, in order to guide future research. Periodic reviewing of research areas can be a healthy exercise in order to discover trends in practices—good or bad—and guide future research, so that good trends can be continued, bad habits can be broken, and practice gaps can be filled. It also enables discussion on the implications of current practices for stakeholders such as policymakers and practitioners.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Soil Use and Management* published by John Wiley & Sons Ltd on behalf of British Society of Soil Science.

Spatial soil information is essential for research and decision-making in several disciplines and at various scales, for example, as support for individual farmers' decisions on how to vary the rate of lime across fields, as input for biophysical modelling of the global carbon cycle and as support for high-level policy decisions. A digital soil map can also be the starting point for formulating hypotheses in soil science (Wadoux & McBratney, 2021). By the end of the 20<sup>th</sup> century, methods were becoming available to create digital maps of soil properties by empirical modelling using spatial covariate data. Early examples of digital soil property maps were created by McKenzie & Austin (1993) and Moore et al. (1993). Later, Scull et al. (2003) defined the term *predictive soil mapping*<sup>1</sup> and McBratney et al. (2003) defined the concept *digital soil mapping*<sup>2</sup> (DSM) and presented a systematic approach to the task. Since then, many DSM studies have been published. Typically, stacks of environmental raster data are combined with soil property data from laboratory analyses of samples. Empirical models are used to translate the raster datasets into a digital soil map of the target soil property and, if applicable, prediction residuals may be spatially interpolated and used to correct the prediction map. DSM has become a rapidly growing and evolving research field at the intersection of soil science and mathematics (Arrouays et al., 2017; Minasny & McBratney, 2016). Several technical handbooks on production of digital soil property maps have been published (Malone et al., 2017; Hengl et al., 2019; Yigini et al., 2017).

A digital soil property map is a spatial representation of the actual variation in soil properties but, like other representations in the form of maps or models, it is always a generalization of reality. Map validation is therefore necessary to determine whether a digital soil property map is good enough for a certain practical application, or to enable comparison of information accuracy between maps. Validation is generally done by comparing map values (e.g. soil properties) against observed values at known positions. To summarize various aspects of map accuracy, different evaluation measures (hereafter validation metrics) are computed. Some are sensitive to random, unpredictable, errors, some are sensitive to systematic errors, like offsets and scale-shifts, and some are sensitive to both. Some are normalized to enable comparisons between map attributes or between map areas (i.e., datasets with different ranges of variation). The sampling design used

<sup>1</sup>Definition by Scull et al. (2003): "Predictive soil mapping (PSM) can be defined as the development of a numerical or statistical model of the relationship among environmental variables and soil properties, which is then applied to a geographic data base to create a predictive map".

<sup>2</sup>Definition by McBratney et al. (2003): "Digital soil mapping is defined as the creation of geographically referenced soil databases based on quantitative relationships between spatially explicit environmental data and measurements made in the field and laboratory".

### Highlights

- A systematic map of validation practices in digital soil mapping (DSM) was performed.
- Most studies (97%) validated maps using soil observations, and 35% estimated map uncertainty.
- It was most common to use one or more existing sample datasets and validate by cross-validation.
- Essential method information such as sampling design and sample support was frequently missing.
- Based on current practices, we present recommendations that could improve future DSM studies.

for collecting validation data against which to compare the map values is important, as it has implications for the conclusions that can be drawn. For example, if mean absolute error (MAE) is computed based on comparisons with data from a probability sampling,<sup>3</sup> the MAE is (or can be made) an unbiased measure of how accurate the map is across its area (Brus et al., 2011, 2019). Spatial support for observed values is also fundamental for interpretation of validation metrics. If, for example, the validation data consist of laboratory analyses of soil samples with point support, the validation measure will show how well the map represents reality at point locations. If the validation samples represent one-hectare averages, the map accuracy for one-hectare areas is evaluated.

There have been a number of general overviews of the scientific field of DSM, by, for example, Arrouays et al. (2020), Grunwald (2009), Minasny & McBratney (2016), and Zhang et al. (2017). In other studies, various specific aspects of DSM have been reviewed or discussed in detail. These include environmental covariates (Dewitte et al., 2012; Mulder et al., 2011), algorithms (Lamichhane et al., 2019; Padarian et al., 2020), sampling designs (Biswas & Zhang, 2018; Brus et al., 2011), the scale concept (Malone et al., 2013a, 2017), DSM of specific attributes (e.g. Minasny et al., 2013) and DSM in specific geographical regions (e.g. Paterson et al., 2015; Zeraatpisheh et al., 2020). The reviews by Lamichhane et al. (2019) and Zeraatpisheh et al. (2020) are systematic

<sup>3</sup>Definition by OECD (2007): "Any method of selection of a sample based on the theory of probability; at any stage of the operation of selection the probability of any set of units being selected must be known. It is the only general method known which can provide a measure of precision of the estimate. Sometimes the term random sampling is used in the sense of probability sampling".

mappings, the study by Grunwald et al. (2009) is a systematic review, and the study by Padarian et al. (2020) is a systematic review based on machine learning. The other reviews and overviews are narrative or semi-systematic, that is, systematic but not following a predefined search and screening protocol. Brus et al. (2011) and Biswas & Zhang (2018) assessed soil sampling for digital soil mapping, but none of the publications cited above is a systematic map of validation procedures in digital soil mapping literature.

The overall aim of the present study was to conduct a systematic mapping of how validation in DSM is performed and reported in peer-reviewed scientific studies, following the procedure of *RepOrting standards for Systematic Evidence Syntheses* (ROSES) (Haddaway et al., 2018). Specific objectives were to:

1. Make a detailed summary of the validation strategies and validation measures used in studies producing continuous soil property maps.
2. Identify trends, practice gaps and practice clusters in validation procedures.
3. Formulate recommendations for future publications in the subject area regarding good practices for validation and its reporting; identify hitherto often neglected aspects that may be important to consider; and establish a minimum set of method descriptors to be used in DSM studies.

## 2 | MATERIALS AND METHODS

The ROSES framework for systematic reviews and systematic mapping was designed for the field of conservation and environmental management and consists of a *pro forma* and a flow diagram (Haddaway et al., 2018). The *pro forma* is a form for metadata recording, while the flow diagram is a log of the numbers of articles found and discarded in the consecutive search and screening steps. In the present study, we based our search and screening workflow on the ROSES flow diagram, in order to ensure a transparent and systematic procedure.

### 2.1 | Literature search

We searched the following databases: ProQuest (Natural Science Collection), Scopus and Web of Science (Core collection). The search date was 15 October 2018, and the exact search strings used are those presented in Table 1. Essentially, we searched for "Digital soil map" OR "Digital soil maps" OR "Digital soil mapping" in the title or abstract and limited the results to original studies written in English. We set no limitation on year of publication. The search results were imported to the open reference management software Zotero (version 5.0.85).

### 2.2 | Screening criteria and procedure

First, duplicate hits were removed. The remaining studies were then randomly split between four senior researchers in pedometrics, who conducted screening and coding of their share of studies. Studies were excluded if they met any of the following exclusion criteria: not a DSM study (to qualify as a DSM study, at least one covariable was required; that is, digital soil maps created by spatial interpolation of point location soil property data were not included); full text not in English; no continuous maps produced (i.e., all soil classification studies resulting in polygon maps or raster map with attributes expressed on an ordinal or nominal value scale were excluded); not peer-reviewed; and not an original research article (i.e., review studies, book chapters and publications of other types were excluded). The screening was conducted in three phases:

1. **Title screening:** Studies were removed if it was obvious from the *title* that any one of the exclusion criteria was fulfilled.
2. **Abstract screening:** Studies were removed if it was obvious from the *abstract* that any one of the exclusion criteria was fulfilled.
3. **Full-text screening:** Studies were removed if it was obvious from the *full text* that any one of the exclusion criteria was fulfilled. In this step, an additional exclusion criterion was added: the mapped soil property is not a laboratory-analysed soil property, such as clay content, pH, or soil organic carbon content (i.e., DSM studies focusing on, e.g., soil depth or crop suitability were excluded).

The exclusion criteria for map data model (raster or feature), value scales and soil properties were applied to delimit a homogeneous set of literature that allowed a more detailed summary.

### 2.3 | Coding

A Microsoft Excel Table was populated with method details, which included: Purpose of the study and of the map, general information on the mapping area and soil properties, extensive information on sampling and sampling strategy, general information on mapping method, extensive information on validation strategy, and uncertainty estimation including validation and uncertainty measures. Details (columns and values in the coding sheet) are presented in Appendix S1 and definitions of terms in Appendix S2. Since the studies were split among different researchers, five studies were coded jointly in order to co-calibrate judgements. In addition, the four researchers met regularly over the coding period to discuss judgment calls. All coding sheets were then merged and carefully checked for consistency. If any values seemed incorrect, the information was checked in the full-text article.

**TABLE 1** Number of publications added or removed in the search and screening steps

Step	Action	Number of publications		Details
		Change	Remaining after	
1	Database searches			
	<i>Ia: ProQuest</i>	+701	701	<b>Search:</b> noft("Digital soil map") OR noft("Digital soil maps") OR noft("Digital soil mapping"). Additional limits - Source type: Scholarly Journals; Document type: Article; Language: English
	<i>Ib: Scopus</i>	+708	1,409	<b>Search:</b> TITLE-ABS-KEY ("Digital soil map" OR "Digital soil maps" OR "Digital soil mapping") AND DOCTYPE "ar"
	<i>Ic: Web of Science</i>	+807	2,216	<b>Search:</b> TOPIC: ("Digital soil map" OR "Digital soil maps" OR "Digital soil mapping") OR TITLE: ("Digital soil map" OR "Digital soil maps" OR "Digital soil mapping") Refined by: DOCUMENT TYPES: (ARTICLE) AND RESEARCH DOMAINS: (SCIENCE TECHNOLOGY) Timespan: All years. Databases: WOS, BCI, CABI, CCC, DRCI, DIIDW, KJD, MEDLINE, RSCI, SCIELO, ZOOREC. Search language = Auto
2	Duplicate removal	-1,338	878	Duplicates were removed
3	Title screening	-353	525	Title revealed that the study did not produce a digital soil map or a method for digital soil mapping.
4	Abstract screening	-204	321	Abstract revealed that the study did not produce a digital soil map or a method for digital soil mapping.
5	Full-text retrieval	-0	321	All full texts were available
6	Full-text screening	-133	188	Full text reveals that this study did not produce a digital soil map or a method for digital soil mapping.
7	Full texts from other sources	+0	188	No texts from other sources were added
Summary			188	<b>Remaining studies</b>

### 3 | RESULTS AND DISCUSSION

#### 3.1 | Search and screening

The numbers of publications found (+) or discarded (-) in the consecutive steps of the search and screening process are presented in Table 1. In total, 188 publications remained for coding after the three screening steps.

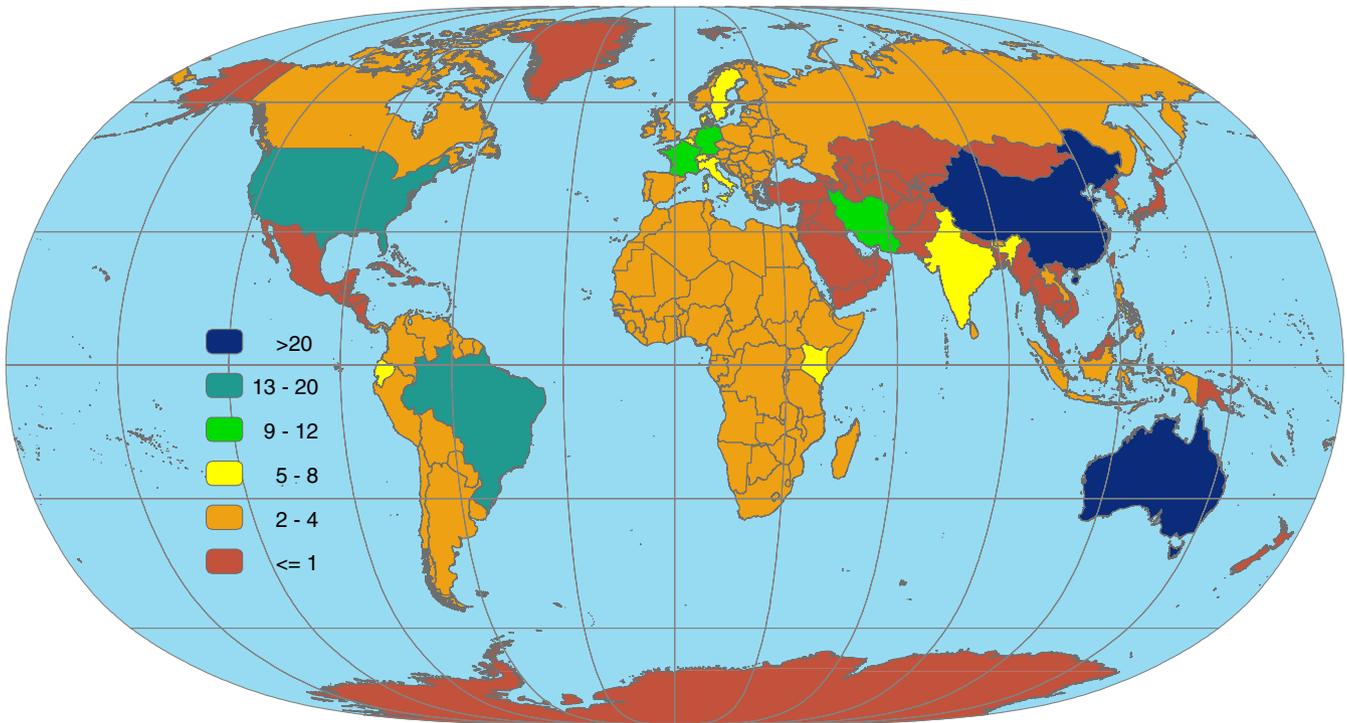
##### 3.1.1 | Spatial distribution of coded studies

The number of coded studies per country is presented in Figure 1. As can be seen, some countries dominated this field of research, assuming that countries that are frequently mapped also host active research groups. Studies mapping areas in Australia and China were most common, followed by Brazil, the United States, France, Germany and Iran. However, when interpreting the map, it should be borne in mind that only studies with the full text in English were included. It can be noted that most of the world is now covered by some kind

of continuous digital soil map, either a local map product or a continental or global map product (Figure 1).

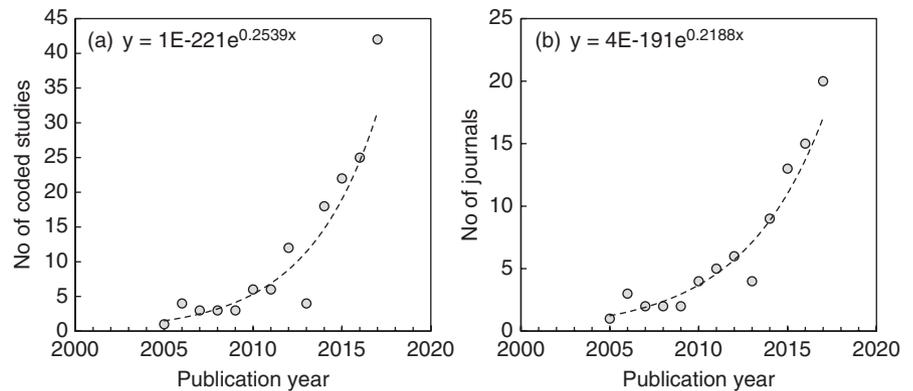
##### 3.1.2 | Number of studies per journal and year

As observed also by, for example, Lamichhane et al. (2019), the number of peer-reviewed studies on DSM has increased rapidly in recent decades (Figure 2). The first published article coded in our systematic mapping was not published until 2005. The reason for the first article located being relatively late was probably the strict screening process applied, with high specificity (i.e., high likelihood that search hits will be retained) but rather low sensitivity (i.e., a risk of missing publications that should be included). This may be because 'digital soil map' was used as a search term or because the analysis was restricted to maps of continuous soil properties and excluded soil class maps. However, this was not considered a problem, as the aim was not to include all DSM studies, but to describe the field of science based on a systematic sample of the existing DSM literature.



**FIGURE 1** Number of coded studies per country. Administrative boundaries from GADM version 3.6. Retrieved from [www.gadm.org](http://www.gadm.org) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

**FIGURE 2** Time series of (a) number of coded studies and (b) number of journals with coded studies



The DSM studies located were published in a large number of journals (Figure 2b; Table 2). The journal *Geoderma* dominated, with 32% of coded studies, followed by seven other journals with at least five coded studies: *Geoderma Regional*; *Science of the Total Environment*; *Catena*, *Soil Research*; *Ecological Indicators*; *European Journal of Soil Science*; *Remote Sensing*; and *PlosONE*. In addition, 44 other journals with fewer than five DSM studies each were retrieved in the systematic mapping. It can be noted that the number of journals increased exponentially, in the same manner as the number of coded studies (Figure 2).

### 3.2 | Study aim and map use

The aims and uses of maps produced in the coded studies are summarized in Table 3. More than 60% of the 188 studies

focused on method only, and 52% provided no information on intended map use. This indicates that the methodology for deriving digital soil maps is a science in itself, and not just a means to produce maps for other purposes. There were no clear trends in study aim or map use over time (data not shown).

### 3.3 | Map cell size, horizontal extent and soil depths

Summaries of map cell size, map extent and map soil depth layers revealed that relatively high resolution ( $\leq 100$  m) was common up to subcontinent level and that the most common type of area mapped was watersheds (Tables 4 and 5). Almost one-third of the studies included one or more subsoil depth layers, which was possibly an effect

**TABLE 2** Number of coded studies per journal, expressed as a percentage of the 188 coded studies included in the analysis

Journal	Prevalence
Geoderma	32%
Geoderma Regional	6%
Science of the Total Environment	5%
Catena	5%
Soil Research	5%
Ecological Indicators	4%
European Journal of Soil Science	3%
Remote Sensing	3%
PLOS ONE	3%
Forty-four different journals with <5 coded studies each	34%
Total	100%

**TABLE 3** Study aims and map uses, expressed as a percentage of the 188 coded studies included in the analysis

Map use	Study aim			Total
	Map	Method	Map and method	
Soil assessment and monitoring	9%	10%	3%	22%
Decision support	5%	10%	2%	17%
Model input	3%	2%	1%	6%
Research	1%	1%	1%	3%
No info	9%	36%	7%	52%
Total	27%	59%	14%	100%

*Note:* Method: The aim was to propose, demonstrate, test and/or compare methods. Map: the aim was to produce, interpret and/or use the map. Soil assessment and monitoring: the map was used or intended for spatially explicit assessment and/or monitoring of one or more soil properties. Decision support: the map was used or intended to guide legislation, policy and/or management. Model: the map was used or intended as input data to a specific model. Research: the map was used or intended to answer a specific research question.

of GlobalSoilMap specifications with six standard depths (GlobalSoilMap, 2015). Mapping soil properties in six depth layers across the globe is a very ambitious task, and the resources required for developing a high-quality map should be considered. Depending on data availability (covariates and reference soil samples), in some cases a less ambitious map may be a better choice. There were no clear trends in map extent, map cell size or maximum soil depth over time (data not shown). However, 23% of the studies provided no information on the cell size in the final map. These included studies where the covariates had different cell sizes and there was no information on how the differently specified raster datasets were resampled and fused. In cases where all covariate data were resampled to a common grid, this was assumed to be the

cell size of the final map. Thus, these studies are not included in the 23% providing no information, even if the cell size was not explicitly stated. Malone, McBratney, et al. (2013) point out that there is often a mismatch between the scale (extent, resolution and support) required for practical applications of digital soil maps and the scale of digital soil maps available. This is illustrated by, for example, Li et al. (2019) and Söderström et al. (2017), who demonstrated that local adaptation of large-scale maps might be needed before practical application. As the DSM method itself, rather than the application of the map, was the focus of many studies, it is perhaps understandable that covariate resolution and/or computational considerations is the main guide for choosing the map resolution suitable for practical applications.

### 3.4 | Sampling design

The sampling designs used in the coded studies are listed in Table 6. In 13% of the studies, various types of probability sampling were used (grid,<sup>4</sup> random and random stratified). This means that unbiased estimations of map accuracy and precision metrics can be made if evaluation is done by random data-splitting, which preserves the properties of the probability sampling. It may be noted that preservation of the properties of probability sampling may prove difficult in case of random splitting of a grid sample (i.e., the design-based statistical inference that goes which such sample may not be easy to derive). If non-probability sampling is used, the validation subset will not be a probability sample (Brus et al., 2011).

Brus et al. (2011) found use of existing soil sample data to be very common, a finding confirmed in the present study. The most common practice was to use a mix of different sampling campaigns (Table 6). The samples were often collected for different purposes and with different sampling designs. A mixed sampling design made it difficult to interpret the validation metrics reported. Around 25% of the studies provided no information on sampling design or only made reference to another publication describing the design (not always easily accessible or in English). This is surprising, because with an unknown sampling design, evaluation measures cannot be interpreted and the value of presenting them may be questionable. We highly recommend that the soil sampling design is always described, at least briefly, even if the samples were collected in a previous study or survey, and even in cases where references to other publications are given.

In addition to the spatial distribution of soil samples, *sample support* is important for interpretation of validation

<sup>4</sup>Grid sampling is a type of probability sampling provided that the starting point of the grid is chosen randomly.

**TABLE 4** Raster cell sizes for different map extents, expressed as percentage of the 188 coded studies included in the analysis

Extent Type/IQR (km <sup>2</sup> )	Cell size						Multiple	No. info	Total
	≤10 m	≤30 m	≤100 m	≤500 m	≤5,000 m				
Subcontinent [10,000,000, 30,000,000]	---	---	1%	1%	1%	---	1%	4%	
Country [72,000, 920,000]	---	1%	4%	2%	2%	1%	---	10%	
Region/province [820, 120,000]	---	7%	12%	1%	4%	---	8%	32%	
Watershed <i>etc.</i> [31, 1,200]	8%	12%	7%	1%	1%	2%	8%	39%	
Field/farm [0.1, 1]	6%	2%	---	---	---	1%	6%	15%	
Total	14%	22%	24%	5%	8%	4%	23%	100%	

Note: Multiple: several maps with different cell sizes were produced. Subcontinent: subcontinent or larger area (including global maps).

---: 0%.

Abbreviation: IQR, Interquartile range.

**TABLE 5** Number of mapped soil depth layers or horizons for different map extents, expressed as a percentage of the 188 coded studies included in the analysis

Extent Type/ IQR (km <sup>2</sup> )	Number or depth layers or horizons				Total
	1	2	3 to 10	No info	
Subcontinent [10,000,000, 30,000,000]	2%	---	---	1%	3%
Country [72,000, 920,000]	5%	1%	4%	1%	11%
Region/province [820, 120,000]	18%	2%	10%	2%	32%
Watershed <i>etc.</i> [31, 1,200]	30%	3%	6%	2%	41%
Field/farm [0.1, 1]	8%	1%	2%	2%	13%
Total	63%	7%	22%	8%	100%

Note: --- 0%. Subcontinent: subcontinent or larger area (including global maps).

Abbreviation: IQR, Interquartile range.

metrics. In 46% of the coded studies, information on sample support was lacking (Table 7). Of the 87% of studies that reported sample support, relatively small-scale spatial sample support (≤100 m<sup>2</sup>) was provided. This included studies with point support (i.e., sample collected by a single auguring) and samples taken in dug pits. When deciding on sample support, the intended map use should be borne in mind, for example, small spatial sample support may not be relevant for maps used to guide large-scale decision-making. When describing a digital soil map, providing information on

**TABLE 6** Prevalence of sampling designs, expressed as a percentage of the 188 coded studies included in the analysis

Sampling design	Prevalence
Grid	8%
Random	3%
Random stratified (geometric strata)	3%
Random stratified (targeted strata)	13%
Targeted	13%
Convenience	2%
Mixed	33%
No information	25%
Total	100%

Note: Definitions of the sampling designs are given in Appendix S2.

sample support is important, as it has implications for: (i) the distribution of observed property data, for example, larger support means that local extreme conditions are smoothed (De Grujter et al., 2006); (ii) the coupling between the environmental raster datasets used as covariates and the soil property data, as different procedures to link spatial support data can lead, for example to different inferences (Young & Gotway, 2007); and (iii) what the raster values in the created map or the validation metrics represent. The soil samples used for model calibration (if no smoothing is done) determine the support for the raster cell values in the digital soil map produced. Additionally, the support for the validation statistics is the same as that for the soil samples used for validation. Malone et al. (2013) provide a comprehensive review of spatial scaling operations for digital soil maps in raster format, including changes in extent, cell size and

**TABLE 7** Sample support (area represented and number of subsamples) given as a percentage of the 188 coded studies included in the analysis

No. of subsamples	Represented area						No info	Total
	Pit	≤1 m <sup>2</sup>	≤100 m <sup>2</sup>	≤1 ha	>1 ha	Multiple		
1	8%	3%	---	---	---	---	---	11%
2–10	---	5%	8%	4%	---	1%	---	18%
>10	---	---	---	1%	---	---	---	1%
Multiple	---	---	1%	1%	---	---	---	2%
No information	19%	1%	2%	1%	---	---	45%	68%
Total	27%	9%	11%	7%	---	1%	45%	100%

Note: ---: 0%. Multiple: different numbers of subsamples taken for different samples (in different areas).

**TABLE 8** Prevalence of mapped attributes, expressed as a percentage of the 188 coded studies included in the analysis

Map attribute	Prevalence
Soil organic carbon/organic matter	68%
Soil inorganic carbon	2%
Total carbon	1%
Soil texture	40%
pH	19%
Bulk density	10%
CEC/sum of base cations/base saturation	9%
Plant macronutrients (N, P, K, Ca, Mg, S)	15%
Trace elements (B, Ni, Cu, Zn, Fe, Cr, As, Pb)	4%
Total	100%

CEC: cation exchange capacity.

support. Bishop et al. (2015) compared the accuracy of digital soil maps with different validation support and found the highest accuracy for maps with the largest support. In a later study, Piikki & Söderström (2019) compared map accuracy for different validation supports and map extents. Validation at point support, a common practice, may underestimate map accuracy at the support of relevance for intended map use.

### 3.5 | Mapped attributes

It is very common for DSM studies to focus on more than one soil attribute. A total of 314 attributes were mapped in the 188 coded studies, which is on average 1.7 attributes per study. Soil organic carbon (SOC) dominated, with 68% of the coded studies producing maps of SOC concentration or SOC stock, followed by soil texture (40%), soil pH (19%) and plant macronutrient concentrations in the soil (15%) (Table 8). Plant micronutrient concentrations in the soil were not commonly mapped in the 188 studies, despite their importance for

**TABLE 9** Prevalence of different validation strategies, expressed as a percentage of the 188 coded studies included in the analysis

Validation strategy	Prevalence
Independent	7%
Cross-validation	43%
Data-splitting	31%
Calibration samples only	3%
Multiple	13%
No validation	3%
Total	100%

crop growth and human nutrition (e.g. Kihara et al., 2020). Only 4% of the studies mapped at least one micronutrient or other trace element. One possible explanation for the low frequency of trace element mapping could be that it may be difficult to map certain element concentrations from available covariates.

### 3.6 | Validation strategies

Most of the 188 coded studies included validation of the maps produced. Only five studies (~3%) did not present any validation and another six studies (also ~3%) presented validation statistics only for the calibration dataset (Table 9). Different types of data-splitting and cross-validation were the most common procedures and, of these, leave-one-out cross-validation was the most common (Table 10). It was not unusual for more than one type of validation to be carried out, for example, with an independent sample and by cross-validation.

As already summarized by Biswas & Zhang (2018), tests and discussions by Brus et al. (2011), Mueller et al. (2004) and Schmidt et al. (2014) have shown that:

1. independent probability sampling is preferable for validation of digital soil maps.

**TABLE 10** Prevalence of types of cross-validation, expressed as a percentage of the 95 studies using cross-validation (cross-validation + part of the multiple category in Table 8)

Cross-validation type	Prevalence
Leave-one-out	43%
<i>k</i> -fold	33%
Repeated <i>k</i> -fold	13%
Bootstrapping	5%
Multiple	4%
Other	2%
Total	100%

Note: *k*: number of folds. Multiple: more than one type of cross-validation used. The different types of cross-validation are described in Appendix S2.

- if time and money do not permit this, leave-one-out cross-validation or bootstrapping is the next best choice.
- one-time data-splitting is less good, if it is not random, as the two subsets may become biased and
- using calibration samples for validation seriously overestimates map accuracy.

Data-splitting is also a problem in studies with relatively few samples, because models created by a smaller number of observations can be less accurate, and validation in that case can underestimate the accuracy of the mapping (when all data are used). Cross-validation produces much more stable results because it uses all data for validation and should be preferred over data-splitting.

In the cross-validation studies, number of folds (*k*) ranged from two to 10, with 10 being the most common, and the number of repetitions in *repeated k*-fold cross-validation ranged from five to 5,000, with 100 being the most common (data not shown). If samples included in the final map-making are left out from the evaluation, it is not the final map *per se* that is evaluated but rather the digital soil mapping framework (the combination of algorithms, geometric specifications, reference data and covariate data). How well this estimated accuracy represents the accuracy of the final map depends on the impact of the left-out sample(s) on the final map. The final map is validated only when an independent sampling is carried out, and possibly also when the data are split once.

In the data-splitting studies, random selection of validation samples was most common (Table 11). This is good because, provided that the original sampling design is a probability sampling, the subsets can be used for unbiased estimation accuracy and precision. The fraction of samples used for calibration ranged from 33% to 95%. Using a relatively large fraction for calibration of samples may be preferred because there will be more representative observations in the calibration set.

**TABLE 11** Prevalence of types of data-splitting, expressed as a percentage of the 74 studies using data-splitting

Data-split	Prevalence
Systematic	1%
Random	74%
Random stratified or targeted in covariate space	11%
Random stratified in geography	4%
No information	10%
Total	100%

### 3.6.1 | Prevalence of less appropriate methods

Unfortunately, problematic validation procedures were relatively common. In 28% of the coded studies, there were possible flaws with the validation (Table 12). In an additional 9% of the studies, not enough information was presented to judge whether any of these problems occurred. We interpret this as an indication that there is some degree of un-reflected routine in DSM validation. Validation should be designed with thought and care, bearing in mind that methods are not justified simply by the fact that they are commonly used in practice.

When spatially clustered soil sample datasets are split, without keeping the clusters together, the validation dataset may not be independent from the calibration dataset. The validation metrics are then not representative of the entire area. It has previously been demonstrated that this can be a problem when validation is used for identification of robust models (Piikki et al., 2016). A robust model can be identified through the manner in which the model was validated. The validation design shall be challenging, so that it mimics a real-world application, and not just test how the model performs within the calibration dataset. In some circumstances, leave-one-out cross-validation and *k*-fold cross-validation with random splitting of data into the *k* folds tend to overestimate map accuracy, as demonstrated by e.g. Piikki et al. (2016) and Meyer et al. (2018). Whether this is the case or not depends on the sampling design. In a grid sampling or random sampling, overestimation should not be a problem, but if samples are spatially clustered or if the dataset contains multiple samples from the same soil profile, this is a real risk. In these cases, leave-one-out cross-validation is not recommended, instead a *k*-fold cross-validation with suitable data-splitting method (e.g. leaving one cluster or one profile out at a time) would be recommended. In Table 11, summary statistics of the prevalence of different methods to split data into *k* folds is reported.

Targeted sampling designs or semi-targeted sampling designs (i.e., sampling designs that are stratified in covariate space) were commonly used in the coded studies. This is suitable for collecting data for calibration of models, but when data from a targeted or semi-targeted sampling are split

**TABLE 12** Prevalence of observed problems in validation design, expressed as a percentage of the 188 coded studies included in the analysis

Possible flaw	Prevalence
Calibration and validation data are the same	3%
Splitting of targeted or semi-targeted sample sets	22%
Splitting of spatially clustered data	20%
Splitting of profiles	1%
Poor areal coverage of the validation set	1%
No validation	3%
Other	1%

between a calibration and a validation dataset, the following happens: (i) the calibration data no longer capture all the variation they were designed to capture, and (ii) the validation statistics can be misleading.

### 3.7 | Validation measures

Almost 30 different validation metrics were used in the coded studies (Table 13; Appendix S3), with on average 2–3 metrics per study. However, several reported validation metrics provide almost the same information about model performance/map quality. For example, MAE and root-mean-square error (RMSE) are both absolute measures of the average magnitude of error, although RMSE is more sensitive to outliers than MAE (Janssen & Heuberger, 1995). Yet 15% of the 188 coded studies reported both MAE and RMSE. We recommend choosing a faceted set of validation metrics to characterize multiple aspects of the validation performance, rather than using several similar metrics.

There were some sources of confusion in the reporting of validation metrics. The abbreviation  $R^2$  was used both for coefficient of determination and for amount of variance explained. To prevent confusion among readers, it is therefore good practice to be very specific when describing validation metrics. Several of the evaluation measures were also reported under different names; for example, RMSE was also denominated RMSD, RMSEP and RMS (Appendix S3).

The possibility to compare model performance and map accuracy between attributes or datasets depends on whether the error (or accuracy) is expressed in the units of the attribute or in relation to the level or spread of observed values. When for example comparing prediction results, absolute errors given in the units of the attribute in question (such as the RMSE and MAE) can be compared between datasets of different sizes (e.g. representing different map areas), while measures of covariation between observed and predicted values (such as the coefficient of determination ( $R^2$ ), the adjusted  $R^2$ , or amount of variance explained (e.g. Nash-Sutcliffe

**TABLE 13** Prevalence of validation measures, expressed as percentage of the 188 coded studies included in the analysis

Validation measure used	Prevalence
Total error (proportion of variance explained)	
LCCC	22%
Nash-Sutcliffe E	3%
Other measure used in < 5 studies	1%
SUM	26%
Total error (absolute)	
RMSE	79%
MAE	20%
MSE	7%
MAD	3%
Other measure used in < 5 studies	4%
SUM	113%
Total error (relative to average* in observed data)	
NRMSE	10%
Other measure used in < 5 studies	3%
SUM	12%
Total error (relative to spread in observed data)	
RPD	7%
RPIQ	3%
Other measure used in < 5 studies	2%
SUM	10%
Total error (relative to spread in predicted data)	
MSDR	4%
Random error only**	
$R^2$	56%
$R^2$ -adj	4%
r	3%
Other measure used in < 5 studies*	2%
SUM	65%
Systematic error only	
ME	38%
Other measure used in < 5 studies	3%
SUM	41%

Note: For abbreviations, see Appendix S3. For definition of terms, see Appendix S2.

\*Average is the mean or the median.

\*\*Covariation of predicted and observed values (or lack thereof caused by random error).

modelling efficiency (E), mean absolute percentage error (MAPE) and the ratio of performance to deviation (RPD)) cannot be directly compared between datasets with different spreads in observed values (see Appendix S3 for a summary of possible comparisons). However, when comparing models for different attributes, or models developed and validated on data with different ranges in the attribute variable, measures

**TABLE 14** Prevalence of uncertainty estimation strategies, expressed as a percentage of the 188 coded studies included in the analysis

Uncertainty estimation strategy	Prevalence
Multiple realizations (variation induced by internal random sampling of calibration data in the algorithm model fitting procedure)	10%
Multiple realizations (variation induced by repeated random sampling of calibration data)	8%
Multiple realizations (variation induced using cross-validation subsets of calibration data)	5%
Geostatistics (e.g. kriging standard deviation)	6%
Other strategy	5%
Unclear method description*	1%
No uncertainty estimation	65%
Total	100%

\*Unclear method: uncertainty estimation not described well enough to be coded.

taking variation into account will help in the comparison. A good strategy is therefore to include at least one evaluation measure from each category.

### 3.8 | Uncertainty estimation strategy

The term accuracy is often used for quantitative (or qualitative) measures of how close a predicted value is to the true (or in practice the observed) value. To compute accuracy means to compute an error. The term uncertainty is, in digital soil mapping, often defined as the expected, or observed, variation in predictions or prediction means for a given target variable at each prediction location. Uncertainty is often quantified by statistical parameters of a distribution, in this case a distribution of soil property predictions for a specific location. The distribution of predictions is not always uniform. It is commonly a normal distribution, but the shape of the distribution can vary depending on the data. Accuracy metrics are computed, while uncertainty metrics are estimated. The accuracy of a digital soil map can be determined at soil sample locations, while the uncertainty of the same map commonly is estimated for every raster cell. In this context, it can be noted that variability between multiple model predictions will in almost all cases produce a gross underestimation of how accurate the digital soil map is (in terms of lack of error of any kind).

In digital soil mapping literature we found that, to assess prediction uncertainty of a digital soil map, prediction distributions are derived in different ways, often by repeated reparameterization of models, using different subsets of data for calibration (e.g. by bootstrapping that is built into the algorithm or by repeated parameterization of new models from random subsets of the calibration data). In 35% of the 188 coded studies, prediction uncertainty was estimated (Table 14), and in 85% of these, the spatial variation in prediction uncertainty was presented in the form of a map.

**TABLE 15** Prevalence of uncertainty measures, expressed as a percentage of the 65 studies in which uncertainty was assessed

Uncertainty measure	Prevalence
Measures of prediction spread*	52%
Prediction intervals**	28%
Confidence intervals	17%
Other	3%
Total	100%

\*Standard deviation, variance and different interpercentile ranges.

\*\*In eight of these 18 studies, the quality of the prediction interval estimation was quantified by prediction interval coverage probability (PICP).

The most common uncertainty measures were different measures of spread in predictions, such as standard deviation, variance and different interpercentile ranges, but also prediction intervals (PI) and confidence intervals (CI) were used (Table 15). An  $l\%$  PI gives information on the range within which  $l\%$  of future predictions can be expected, while an  $l\%$  CI gives information on the range within which the true mean of prediction can be expected with a confidence of  $l\%$ . The  $l\%$  PI is always wider than the  $l\%$  CI and, when the number of calibration samples increases, the CI becomes narrower while the PI remains unaffected. It is as important to assess the quality of the uncertainty assessment as it is to assess the quality of the predictions (Malone et al., 2011). For this, the prediction interval coverage probability (PICP) (Solomatine & Shrestha, 2009) can be used; in eight of the 18 studies where the PI of predictions was presented, the quality of the PI estimation was quantified by PICP (Table 15).

### 3.9 | Reported method information

The prevalence of missing important method information has already been mentioned at various stages in this paper.

**TABLE 16** Prevalence of missing information, expressed as a percentage of the 188 coded studies included in the analysis

Missing information	Prevalence
General	
Intended map use	52%
Sampling	
Sampling design	25%
No. of sample locations	2%
Soil depth(s)	18%
No of subsamples	68%
Sample support	45%
Validation	
Data-splitting method	10%
Uncertainty estimation	
Uncertainty estimation method (includes unclear descriptions)	1%
Final map	
Resolution	23%
Soil depth(s)	8%

Table 16 provides a summary to give a better overview and a more complete picture. We suggest that the method details listed in Table 16 and Box 2 be taken as a minimum information requirement in scientific studies on DSM. These pieces of information are essential for the reader to understand and assess the presented research.

In a previous review by Biswas and Zhang (2018), 15% of the 95 DSM studies assessed provided no information on sampling design. This is a high incidence considering the importance of this information. Our review confirmed that this is common and revealed an even higher incidence: in 25% of the 188 coded studies, information on sampling design was missing.

## 4 | SUMMARY

Digital soil mapping has evolved into a well-established method framework and a mature scientific subject, and validation of models and maps is a central part of the work. The primary purpose of the present systematic mapping was to obtain an overview of validation practices used in the DSM research, based on a systematic selection of literature. Digital soil property maps of areas in Australia and China were most common, followed by maps of areas in Brazil, the United States, France, Germany, and Iran. The number of peer-reviewed DSM studies has increased exponentially over in recent decades and they are published in a large number of different journals, but particularly in *Geoderma*.

### BOX 1 Recommended practices and good examples

**Design!** Evaluate what you want to know. Beware of un-reflected routine in map validation. Consider what you want to know about the map and evaluate accordingly; do not routinely present average level of absolute errors at point support if that does not suit your specific application. Larger support or the prevalence of error under a certain limit may be more interesting.

- Malone et al. (2011) evaluated map area that is accurate enough for practical use.
- Bishop et al. (2015) compared map validations for multiple support.
- Angelini et al., (2017) and many others used multiple well-chosen and well-defined validation measures to assess different aspects of map accuracy.

**Inform!** Provide the required information and give a summary. To allow readers to understand and assess the results presented, it is important to provide some fundamental details of methods used (Box 2). In addition to providing the details, an overview is helpful. DSM studies often have complex workflows and many data management steps. Our overall experience from the coding process was that graphical abstracts and schematic study overviews were very useful.

- A good comprehensible method overview is provided by e.g. Wang et al. (2018).

**Interpret!** Compare with a reference. In a study evaluating a map for practical use, it is important to put the accuracy into perspective and to compare the evaluation measures with that required for the practical use in question. It may also be important to know whether the map is better than the mean of the observational data used to derive it. This also applies when comparing several DSM methods. All maps may be good enough and the difference compared with a reference may be negligible, or the error may be close to the laboratory error of the reference soil samples and further improvements may be difficult to assess.

The methodology for deriving digital soil property maps is a science in itself. It is common for DSM studies to focus solely on method development, and not even mention intended or possible map use. Even when the focus is method development, it is important to put the method into context.

**BOX 2 Recommended method information**

Values of validation metrics should always be accompanied by a description of what the metrics represent (validation strategy, sampling design and sample spatial support). If possible, this should also be stated in an abstract or summary. For example: “The MAE in leave-one-out cross-validation of random sampling with point support was 7% clay” is much more informative than: “The MAE was 7% clay”. If the latter is the only information given, the reader cannot know if this is a biased or unbiased estimation of the MAE and the support used to evaluate the map. The MAE may be very different if the same map is evaluated using one-hectare averages or even one-square-kilometre averages.

- **Intended map use.**
- **Sampling design, geometry and method including:**
  - method used to determine sample locations (including specification of any covariate data used)
  - number of sample locations
  - sampling depths
  - sample support
  - number of subsamples
  - Validation strategy
- **Validation metrics**
- **Uncertainty estimation strategy (if any)**
- **Uncertainty estimation metrics (if any)**
- **Raster map specifications**
  - map area
  - raster resolution
  - represented soil depth(s)

As regards the geometry and attributes of the maps produced, watersheds were the most common type of study area and relatively high resolution ( $\leq 100$  m) was common, even for large-extent maps. The spatial support was often small, most often  $\leq 100$  m<sup>2</sup>. Most studies focused on topsoil only, but almost one-third of the coded studies mapped one or more subsoil depth layers or horizons. Soil organic carbon was by far the most common map attribute in the 188 studies reviewed, followed by soil texture.

The cost, both in time and money, for complete, optimized soil sampling is often high, and the use of existing soil information from earlier sampling campaigns is common. In one-third of the coded studies, data from multiple samplings were combined. Only 13% of the studies used some type of probability sampling, allowing for unbiased estimations of map

accuracy and precision. Single data-splitting and different types of cross-validation were the most common validation strategies. When designing a DSM study, it is important to be aware of possible limitations in interpretation of the validation results.

A large number of validation measures were used, with RMSE and R<sup>2</sup> being most common. Several similar metrics had different names, and some metric names referred to several different metrics, which may be a source of confusion. Therefore, it is important to be specific in reporting the metrics used. However, a much larger source of confusion was that it was often impossible to know what the metrics represented, because of mixed or targeted sampling designs or lack of information on sampling design. Overall, information crucial for the reader to understand and assess the research conducted and maps produced was frequently missing.

In this systematic mapping of validation strategies and validation measures used in published studies producing continuous soil property maps, we identified trends, practice gaps and practice clusters in DSM validation. We used these to formulate recommendations for future publications in the subject area. We hope that this summary is useful as guidance for coming DSM studies.

**ACKNOWLEDGEMENTS**

Salaries were partly funded by Formas/SIDA (contract: 220-1975-2013) and Västra Götaland Region/SLU (contract: RUN 2018-00141).

**REFERENCES**

- Angelini, M. E., Heuvelink, G. B. M. & Kempen, B. (2017). Multivariate mapping of soil with structural equation modelling. *European Journal of Soil Science*, 68(5), 575–591. <https://doi.org/10.1111/ejss.12446>
- Arrouays, D., Lagacherie, P. & Hartemink, A. E. (2017). Digital soil mapping across the globe. *Geoderma Regional*, 9, 1–4. <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. & Mulder, V. L. (2020). Impressions of digital soil maps: The good, the not so good, and making them ever better. *Geoderma Regional*, 20, e00255. <https://doi.org/10.1016/j.geodrs.2020.e00255>
- Bishop, T. F. A., Horta, A. & Karunaratne, S. B. (2015). Validation of digital soil maps at different spatial supports. *Geoderma*, 241, 238–249. <https://doi.org/10.1016/j.geoderma.2014.11.026>
- Biswas, A. & Zhang, Y. (2018). Sampling designs for validating digital soil maps: A review. *Pedosphere*, 28(1), 1–15. [https://doi.org/10.1016/S1002-0160\(18\)60001-3](https://doi.org/10.1016/S1002-0160(18)60001-3)
- Brus, D. J. (2019). Sampling for digital soil mapping: A tutorial supported by R scripts. *Geoderma*, 338, 464–480. <https://doi.org/10.1016/j.geoderma.2018.07.036>
- 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>
- De Gruijter, J., Brus, D. J., Bierkens, M. F. & Knotters, M. (2006). *Sampling for natural resource monitoring*, Berlin Heidelberg,

- Germany: Springer-Verlag. <https://doi.org/10.1007/3-540-33161-1>
- Dewitte, O., Jones, A., Elbelrhiti, H., Horion, S. & Montanarella, L. (2012). Satellite remote sensing for soil mapping in Africa: An overview. *Progress in Physical Geography*, *36*(4), 514–538. <https://doi.org/10.1177/0309133312446981>
- GlobalSoilMap. (2015). *Specifications Tiered GlobalSoilMap products*. Release 2.4. 52 pages. Retrieved from: [https://www.isric.org/sites/default/files/GlobalSoilMap\\_specifications\\_december\\_2015\\_2.pdf](https://www.isric.org/sites/default/files/GlobalSoilMap_specifications_december_2015_2.pdf) pp. (1–52). Wageningen, the Netherlands: ISRIC - World Soil Information.
- Grunwald, S. (2009). Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma*, *152*(3–4), 195–207. <https://doi.org/10.1016/j.geoderma.2009.06.003>
- Haddaway, N. R., Macura, B., Whaley, P. & Pullin, A. S. (2018). ROSES RepORting standards for Systematic Evidence Syntheses: Pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. *Environmental Evidence*, *7*(1), 7. <https://doi.org/10.1186/s13750-018-0121-7>
- Hengl, T., MacMillan, R. A. (2019). *Predictive Soil Mapping with R*, (pp. 1–370). Wageningen, the Netherlands: OpenGeoHub foundation. Retrieved from <https://soilmapper.org/>
- James, K. L., Randall, N. P. & Haddaway, N. R. (2016). A methodology for systematic mapping in environmental sciences. *Environmental Evidence*, *5*(1), 7. <https://doi.org/10.1186/s13750-016-0059-6>
- Janssen, P. H. M. & Heuberger, P. S. C. (1995). Calibration of process-oriented models. *Ecological Modelling*, *83*(1–2), 55–66. [https://doi.org/10.1016/0304-3800\(95\)00084-9](https://doi.org/10.1016/0304-3800(95)00084-9)
- Kihara, J., Bolo, P., Kinyua, M., Rurinda, J. & Piikki, K. (2020). Micronutrient deficiencies in African soils and the human nutritional nexus: Opportunities with staple crops. *Environmental Geochemistry and Health*, *42*(9), 3015–3033. <https://doi.org/10.1007/s10653-019-00499-w>
- Lamichhane, S., Kumar, L. & Wilson, B. (2019). Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. *Geoderma*, *352*, 395–413. <https://doi.org/10.1016/j.geoderma.2019.05.031>
- Li, N., Zhao, X., Wang, J., Sefton, M. & Triantafyllis, J. (2019). Digital soil mapping based site-specific nutrient management in a sugarcane field in Burdekin. *Geoderma*, *340*, 38–48. <https://doi.org/10.1016/j.geoderma.2018.12.033>
- 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*(3), 1032–1043. <https://doi.org/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*(3), 890–902. <https://doi.org/10.2136/sssaj2012.0419>
- Malone, B. P., Minasny, B. & McBratney, A. B. (2017). *Using R for Digital Soil Mapping*, Cham, Switzerland: Springer International Publishing. <https://doi.org/10.1007/978-3-319-44327-0>
- McBratney, A. B., Santos, M. M. & Minasny, B. (2003). On digital soil mapping. *Geoderma*, *117*(1–2), 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- McKenzie, N. J. & Austin, M. P. (1993). A quantitative Australian approach to medium and small scale surveys based on soil stratigraphy and environmental correlation. *Geoderma*, *57*(4), 329–355. [https://doi.org/10.1016/0016-7061\(93\)90049-Q](https://doi.org/10.1016/0016-7061(93)90049-Q)
- 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. <https://doi.org/10.1016/j.envsoft.2017.12.001>
- 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>
- Minasny, B., McBratney, A. B., Malone, B. P. & Wheeler, I. (2013). Digital Mapping of Soil Carbon. *Advances in Agronomy*, Volume *118*, (pp. 1–47). Amsterdam, Netherlands: Elsevier. <https://doi.org/10.1016/B978-0-12-405942-9.00001-3>
- Moore, I. D., Gessler, P. E., Nielsen, G. A. E. & Peterson, G. A. (1993). Soil attribute prediction using terrain analysis. *Soil Science Society of America Journal*, *57*(2), 443–452. <https://doi.org/10.2136/sssaj1993.03615995005700020026x>
- Mueller, T. G., Pusuluri, N. B., Mathias, K. K., Cornelius, P. L., Barnhisel, R. I. & Shearer, S. A. (2004). Map quality for ordinary kriging and inverse distance weighted interpolation. *Soil Science Society of America Journal*, *68*(6), 2042–2047. <https://doi.org/10.2136/sssaj2004.2042>
- Mulder, V. L., De Bruin, S., Schaepman, M. E. & Mayr, T. R. (2011). The use of remote sensing in soil and terrain mapping—A review. *Geoderma*, *162*(1–2), 1–19. <https://doi.org/10.1016/j.geoderma.2010.12.018>
- OECD. (2007). *Glossary of statistical terms*. Organisation for Economic Co-operation and Development (OECD), Paris, France, 863 pp. Retrieved from: <https://stats.oecd.org/glossary/index.htm> Paris, France: OECD.
- Padarian, J., Minasny, B. & McBratney, A. B. (2020). Machine learning and soil sciences: A review aided by machine learning tools. *Soil*, *6*(1), 35–52. <https://doi.org/10.5194/soil-6-35-2020>
- Paterson, G., Turner, D., Wiese, L., Van Zijl, G., Clarke, C. & Van Tol, J. (2015). Spatial soil information in South Africa: Situational analysis, limitations and challenges. *South African Journal of Science*, *111*(5–6), 1–7.
- Piikki, K., & Söderström, M. (2019). Digital soil mapping of arable land in Sweden—Validation of performance at multiple scales. *Geoderma*, *352*, 342–352. <https://doi.org/10.1016/j.geoderma.2017.10.049>
- Piikki, K., Söderström, M., Eriksson, J., Muturi John, J., Ileri Muthee, P., Wetterlind, J. & Lund, E. (2016). Performance evaluation of proximal sensors for soil assessment in smallholder farms in Embu County, Kenya. *Sensors*, *16*(11), 1950.
- Schmidt, K., Behrens, T., Daumann, J., Ramirez-Lopez, L., Werban, U., Dietrich, P. & Scholten, T. (2014). A comparison of calibration sampling schemes at the field scale. *Geoderma*, *232–234*, 243–256.
- Scully, P., Franklin, J., Chadwick, O. A. & McArthur, D. (2003). Predictive soil mapping: A review. *Progress in Physical Geography*, *27*(2), 171–197.
- Söderström, M., Piikki, K. & Cordingley, J. (2017). Improved usefulness of continental soil databases for agricultural management through local adaptation. *South African Journal of Plant and Soil*, *34*(1), 35–45.
- Solomatine, D. P. & Shrestha, D. L. (2009). A novel method to estimate model uncertainty using machine learning techniques. *Water Resources Research*, *45*(12), 1–16. <https://doi.org/10.1029/2008WR006839>

- Wadoux, A. & McBratney, A. (2021). Hypotheses, machine learning and soil mapping. *Geoderma*, 383, 1–2. <https://doi.org/10.1016/j.geoderma.2020.114725>
- Wang, B., Waters, C., Orgill, S., Gray, J., Cowie, A., Clark, A. & Li Liu, D. (2018). High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern Australia. *Science of the Total Environment*, 630, 367–378. <https://doi.org/10.1016/j.scitotenv.2018.02.204>
- Yigini, Y., Olmedo, G. F., Reiter, S., Baritz, R., Viatkin, K., Vargas, R. (eds). (2017). *Soil Organic Carbon Mapping Cookbook*. 1–180). Rome, Italy: nd edn. : FAO. 1 pp. Retrieved from: <http://www.fao.org/global-soil-partnership/pillars-action/4-information-and-data/global-soil-organic-carbon-gsoc-map/en/> <http://www.fao.org/3/a-bs901e.pdf>.
- Young, L. J. & Gotway, C. A. (2007). Linking spatial data from different sources: The effects of change of support. *Stochastic Environmental Research and Risk Assessment*, 21(5), 589–600.
- Zeraatpisheh, M., Jafari, A., Bodaghabadi, M. B., Ayoubi, S., Taghizadeh-Mehrjardi, R., Toomanian, N., ... Xu, M. (2020). Conventional and digital soil mapping in Iran: Past, present, and future. *Catena*, 188, 104424.
- Zhang, G. L., Feng, L. I. U. & Song, X. D. (2017). Recent progress and future prospect of digital soil mapping: A review. *Journal of Integrative Agriculture*, 16(12), 2871–2885. [https://doi.org/10.1016/S2095-3119\(17\)61762-3](https://doi.org/10.1016/S2095-3119(17)61762-3)

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Piikki K, Wetterlind J, Söderström M, Stenberg B. Perspectives on validation in digital soil mapping of continuous attributes—A review. *Soil Use Manage.* 2021;37:7–21. <https://doi.org/10.1111/sum.12694>