On the accuracy of soil survey in Queensland, Australia

Andrew J. W. Biggs A,B,C, Mark Crawford A, Kaitlyn Andrews A, Mark Sugars A, Dan Smith A, and Warrick Brown A

ADepartment of Natural Resources, Mines and Energy, Qld, Australia.
BUniversity of Queensland, School of Agriculture and Food Sciences, St Lucia, Qld, Australia.
CCorresponding author. Email: andrew.biggs@dnrme.qld.gov.au

Abstract. The accuracy of soil survey is not well described in a consistent manner for either conventional or digital soil mapping in Queensland or more generally in Australia. Concepts of accuracy are often poorly understood and the rise of digital soil mapping has led to further terminology confusion for clients. Despite long-standing recommendations for derivation of accuracy statistics of soil surveys via statistically-based external validation, accuracy assessment by this method has been limited. Concepts for accuracy description (overall, producers and users accuracy) from the remote sensing discipline are applicable to soil survey and their use should be encouraged. An analysis of 12 published 1:50 000 and 1:100 000 soil surveys in Queensland revealed a 73% to 97% match between mapped polygonal and site data. This, in conjunction with accuracy standards for similar mapping disciplines and published soil survey accuracy assessments, leads us to recommend that a benchmark of 80% accuracy is realistic for all types of soil surveys. The adoption of a benchmark is however dependent upon further development and evaluation of accuracy assessment methods and standards, particularly in relation to minimum sample size and acceptance criteria. These outcomes will only be achieved if all surveys include accuracy assessment within the survey design.

Keywords: standards, validation, soil mapping, soil survey, accuracy, remote sensing, digital soil mapping, assessment.

Received 19 May 2020, accepted 15 January 2021, published online 4 March 2021

Introduction

The accuracy of soil survey in Queensland or Australia in general has not been evaluated or described well, despite the existence of the profession for more than 70 years. This has led to variable and at times ill-informed views of both the quality and limits to use of soil survey data. To non-practitioners (clients), the question is simple – how good is the data? This may seem an easy question, but it is one that can lead to considerable debate and confusion, as accuracy is a term that can have differing meaning to different people e.g. reliability, confidence, precision, uncertainty. The use of geostatistical terminology in digital soil mapping (DSM) for reporting various model uncertainty statistics, has created further confusion for many users of soil survey data. Associated with this has been a growing paradigm that it is up to the consumer of the data to decide if it is accurate enough for their purposes. Achieving a consistent and meaningful understanding of the description and measurement of accuracy is crucial for the discipline of soil survey, both in terms of internal standards and external communication.

These thoughts are not new – the Guidelines for Surveying Soil and Land Resources1 (McKenzie et al. 2008a) make recommendations regarding the need for accuracy assessment and benchmarks for soil survey in Australia. Earlier, McBratney et al. (2003) identified the need to research methods for the quality assessment of evolving DSM methods. Arrouays et al. (2020a) flagged the need to develop standards for map products and that the production of poor quality (DSM) map products is counter-productive. Despite the recommendations of these and other authors, surprisingly little progress has been made in terms of evaluating the accuracy of soil surveys in a consistent manner within Australia (and specifically in Queensland) and the development of associated standards.

In this paper we provide a brief review of the key issues associated with describing and measuring accuracy of soil survey, using examples from Queensland (due to the wide range of surveys available). We follow with revision of published survey accuracy statistics for soil surveys in Queensland and analysis of some historical mapping data. We conclude with some recommendations regarding principles for accuracy assessment, the use of related terminology and priority matters for attention. Throughout the text, our focus is primarily on the matter of accuracy rather than uncertainty. We acknowledge at the start that we do not canvas all possible options and activities, but our intent is to provoke further thought and action regarding these matters.

1Referred to hereon as ‘The Guidelines’.
Key issues concerning accuracy

There are many texts that describe aspects of accuracy (and uncertainty) in soil survey and related disciplines such as remote sensing e.g. Story and Congalton (1986), Congalton and Green (1999), Carré et al. (2007), Minasny and Bishop (2008), Biswas and Zhang (2018). It is not our intention here to repeat the content of those texts, rather to highlight some of the key issues that require attention in the current environment.

We discuss the issues under the headings of:

- Accuracy terms
- The general approach
- Site selection
- Sufficient validation
- Validation criteria (the response design)
- Benchmarks

Accuracy terms

In the case of conventional soil survey, the general usage of the term accuracy often encompasses (rightly or wrongly) related concepts and terms such as confidence, scale, resolution and site density. Some of those terms, for example scale, can have multiple meanings in the context of spatial data (Goodchild 2011). On the whole though, accuracy is not assessed or reported very often in conventional survey. In DSM, other terms such as error, precision, bias, uncertainty, reliability and confidence interval can be relevant in relation to uncertainty and accuracy (Minasny and Bishop 2008).

Statistics such as confusion index (Burrough et al. 1997), \( r^2 \), root mean square error and kappa (Cohen 1960) are often reported. Many of these and other terms have been defined in The Guidelines and elsewhere, but as noted by Minasny and Bishop (2008), terms are often used loosely, incorrectly, inconsistently, or not at all to describe the quality of soil surveys. Minasny and Bishop (2008) described accuracy as ‘a measure of how close the prediction is to being correct’. While this definition is useful, it is apparent that many terms can, or possibly must be used to describe the accuracy of a soil survey, as no single term encompasses all desired aspects or can be applied in all circumstances.

It is easy for non-geostatistical readers to become confused between terms that apply to uncertainty, those that apply to accuracy and the difference between these concepts. While the two have quite specific and different meanings, this difference is often lost on the novice reader. This illustrates one of the major issues with the use of statistics to report accuracy – the choice of language should be dictated by the need to communicate at the ‘lowest intellectual level’, not the highest.

Terms used to describe soil survey spatial data should preferably be consistent with the AS/NZS ISO 19115.1:2015 (Geographic metadata) standard, as this is the standard by which spatial data are described in Australia. Internationally there are related standards such as Digital Catalog Vocabulary - Version 2 (DCAT2, https://www.w3.org/TR/vocab-dcat-2/) and Data Quality Vocabulary (DQV, https://www.w3.org/TR/vocab-dqv/) that contain related and relevant standardised terms for the description of data quality and resolution of spatial attributes.

Within the Australian Spatial Information Council (ANZLIC) standard for example, scope exists to describe terms such as positional accuracy, thematic accuracy and attribute accuracy. In soil survey, the main concern has been thematic accuracy i.e. how accurately are the spatial entities represented on the map. Insufficient attention has been given to describing positional accuracy, even though it is highly relevant to the creation and use of soil data in both conventional surveys and DSM.

The general approach

Hewitt et al. (2008) in The Guidelines, and others such as Brus et al. (2011), have indicated that the preferred method for determining the accuracy of a soil survey is external, independent validation. The term ‘independent’ appears to have three possible meanings in the context of validation – the first being that the sites are chosen in an independent (statistically unbiased) manner. The second is that the sites are independent of the mapping process and there is no personal bias. We use the terms ‘statistically-based’, ‘independent’ and ‘external’ in relation to these three concepts.

Hewitt et al. (2008) provided more detail when stating the key requirements for validation as ‘…collection of a statistically based sample of the complete survey area leading to derivation of statistics of success reported in a manner that can be understood by the user’. Brus et al. (2011) provided a thorough description of sampling strategies for validation of DSM outputs, drawing upon lessons from the remote sensing discipline, where concepts of the determination and communication of accuracy have been explored for decades. Brus et al. (2011) noted the value of error matrices (Rosenfield 1986; Rosenfield and Fitzpatrick-Lins 1986; Congalton and Green 1999; Stehman and Foody 2019) to soil survey map validation.

While the preferred approach for determining accuracy is a statistically-based, independent, external validation exercise, existing site data may also be used to assess survey accuracy. This technique is relevant to DSM methods in which site data is not used as part of the model (Minasny and Bishop 2008), but it is also relevant to historical surveys. Thus, in some areas, thousands of existing sites may be available for use in validation – for example, Odgers et al. (2014) in north Queensland and Holmes et al. (2015) in Western Australia. Polygonal data has also been used for validation e.g. Zund (2014) in Queensland and Vincent et al. (2018) overseas. A detailed study of this approach by Bazaglia Filho et al. (2013) in Brazil however recommended against it, on the basis that a survey map cannot be regarded as a singular point of truth – it is merely a particular realisation of the landscape, in the same way that DSM can yield many realisations. The literature describing validation does however overwhelmingly specify that validation exercises must be unbiased and that validation sites must not be used within any model that they are intended to validate (in the case of DSM). The latter principle is a well acknowledged concept associated with validation of modelling in general, not just in relation to DSM.
Achieving validation using personnel external to the survey is difficult if the activity is not specifically planned and costed within the survey, whether validation occurs as a post hoc exercise or within the mapping phase of a survey. Hewitt et al. (2008) observed that the process of validation ‘...may be threatening to individuals responsible for the initial mapping’. Some of the few surveys (conventional and DSM) in Queensland that have involved validation, such as Powell (1979), Bartley et al. (2013) and Thomas et al. (2018), appear to have not met the criteria of independence, as the validation sites were collected by the survey team after the mapping boundaries were determined. Other surveys, such as Ellis and Wilson (2009) and Smith and Crawford (2015) involved both independent and project staff. Validation of two recent conventional soil surveys (Brown et al. 2020; Smith and Calland 2020) in Queensland, did attempt to use a more independent approach.

Site selection

The statistically-based selection of unbiased sampling locations in DSM and the remote sensing discipline has been well explored in recent years, with most recommending area-weighted random sampling (Brus et al. 2011; Stehman and Foody 2019). Biswas and Zhang (2018) recommended that a validation sampling design should be free of model related assumptions, and the use of an independent dataset is preferred, compared with internal model validation methods. They do however note the need for further evaluation of the pros and cons of different approaches. The most common method used for site selection in DSM surveys in recent years has been a conditioned Latin Hypercube (cLHC) approach (Minasny and McBratney 2006). It has been used for the selection of sites within the mapping phase and for accuracy assessment.

Most surveys involving validation in Queensland have also used constrained approaches i.e. validation sites have been limited to a specified distance from roads for reasons of simplifying access and reducing validation effort. Even with the use of a statistically-based sampling strategy, there is a risk that constrained approaches will lead to inadequate sampling of the survey area due to poor road access (Clifford et al. 2014; Kidd et al. 2015). Road networks are frequently strongly biased by topography and road density typically decreases in steeper lands. The number of validation sites may also be restricted (perhaps unduly) by other factors such as project timeframes and budgets.

Sufficient validation

While The Guidelines describe the importance of validation, they give little guidance on the minimum number of validation sites. Minasny and Bishop (2008) suggested a sample size of 50–200 sites should be sufficient, but acknowledged it would vary with the scope of the study. Geostatistical methods can be used to determine the necessary sample size to achieve a desired confidence level (Wilding and Drees 1983), but one of the challenges is that the sample size will potentially vary with each attribute of concern. Some soil and landscapes attributes are inherently more variable than others within a given area e.g. surface electrical conductivity in a tropical environment versus in a semi-arid environment. One method is to continue to sample until the validation result ceases to change significantly (Yang et al. 2020). Such an approach would however be difficult to administer from a practical and time/cost perspective within a survey, as the quantity of validation is not known at the survey design stage.

Validation site numbers within surveys in Queensland have varied considerably. Powell (1979) undertook slightly more validation sites than mapping sites (351 vs 342) in a conventional 1:25 000 survey covering 5000 ha in south-east Queensland, but this is unlikely to be a normal scenario. Smith and Calland (2020) described 82 validation sites for a conventional soil survey of 258 000 ha in south-east Queensland, representing ~5% of their total number of sites. In a similar validation exercise, Brown et al. (2020) collected 34 sites in a 62 473 ha survey near Bundaberg, representing ~4% of their total number of sites.

In an early example of validation of a DSM survey in Queensland, Ellis and Wilson (2009), collected 81 validation sites, but for only a limited suite of attributes, in a study area of 127 000 ha. Smith and Crawford (2015) described 34 validation sites with an associated 266 survey sites in a hybrid DSM survey covering 26 000 ha on the Sunshine Coast. In a larger area DSM study (~155 000 km²) in northern Queensland, Bartley et al. (2013) collected 111 validation sites, which represented 25% of the total field effort in that study. In another large DSM survey in north Queensland, Thomas et al. (2018) collected 40 validation sites over a 72 000 km² study area. Interestingly, Bartley et al. (2013) also created extra ‘virtual’ validation sites via satellite imagery. There is no doubt that certain landscape features can be attributed with a high degree of confidence by experienced soil surveyors via high resolution imagery, but the creation of sites in such a manner is obviously open to subjectivity. Overall, there has been considerable variation in the number of validation sites collected and no clear guidance for surveyors in relation to determining an appropriate number of validation sites in a practical, consistent manner.

Validation criteria

A common problem for those undertaking validation is the response design i.e. how to determine a match between predicted and observed parameters (Stehman and Foody 2019). Many authors who have undertaken validation of soils surveys have commented upon the issue, whether the entity being validated is a soil type or an individual soil attribute (Bartley et al. 2013; Holmes et al. 2015; Thomas et al. 2018; Brown et al. 2020; Smith and Calland 2020). The window for determining a match – for example a binary test (yes/no) or a degree of tolerance around a central value – can significantly influence the accuracy result. Some surveys, such as Powell (1979) and Bartley et al. (2013), did not indicate how a match was determined and the reader is left to assume that a complete match was required. Zund (2014) validated and reported accuracy against varying levels of the Principal Profile Form (Northcote 1979).
In the more recent work of Thomas et al. (2018), validation of individual attributes was deemed as correct, accept or fail, which is a useful approach that may be applied to either soil types or attributes. As with other approaches though, the outcome is strongly influenced by the boundaries set for each category. For example, in the case of that survey, the authors did not assess the accuracy of the soil attributes directly, the acceptance criteria were a function of a secondary rule i.e. whether or not the difference between the measured value and the predicted value influenced the outcome of a land suitability framework in which the data was used. While there is some logic to this approach (does the difference between predicted and observed really matter?), it relies inherently on an assumption that the land suitability framework is correct. This is not necessarily a safe assumption, nor an objective measure of the prediction of pedological attributes. Furthermore, the accuracy assessment is tied to the intended land use, which means it is of no value for other uses of the data.

Brown et al. (2020) used narrow rules relating to a validation match for soil profile classes (SPCs) but also applied a geographic tolerance. A validation site was defined as a match if the SPC of the site matched any SPC assigned to the unique mapping area (UMA) within the site was located or any entity of a UMA within 100 m. The SPC assigned to the validation site should also occupy a large enough area to meet minimum mapping area requirements for the survey scale of the project. These rules did not consider the similar characteristics between some of the SPCs within the project. For example, the split between SPCs occupying the same landscape position and geomorphic setting was as small as half a pH unit, which puts it within the error margin of field pH methods.

Despite the direction provided in literature, practices concerning the use of validation data have varied and at times appear to have contravened guidance principles. For example, validation sites within Bartley et al. (2013) were added back into their model, thus the validation statistics they presented did not reflect the final outputs. The justification provided by the authors was a lack of site data in the survey area. In the similar survey of Thomas et al. (2018), the model for some attributes was modified after the validation process and consequently the final models for those attributes were not validated.

**Accuracy benchmarks**

How good is good enough? Hewitt et al. (2008) did not set a benchmark for ‘good quality’, partly because of the paucity of existing data for survey accuracy. Many decades ago, Pomerening and Cline (1953) in the USA suggested that the combination of stereo air photograph interpretation and ground validation should yield an accuracy of 80–90% for soil survey. Hewitt et al. (2008) suggested that if validation occurs, then over time the statistics generated will suggest the benchmark. Other mapping disciplines have been more proactive in setting a minimum standard for mapping accuracy. For example, the accuracy benchmark for vegetation mapping in Queensland is >80% (Neldner et al. 2019) and that for land use mapping in Australia is also >80% (ABARES 2011), although these are both ‘above-ground’ mapping disciplines. A commonly used benchmark for accuracy of remote sensing related mapping such as land cover is 85%, however in some instances it may not be appropriate (Stehman and Foody 2019).

**Published accuracy assessment results**

Published results for the accuracy of soil surveys in Queensland have varied considerably (Table 1), perhaps in part because of the variations in approach used. In terms of conventional surveys, Powell (1979) observed an 87% accuracy of any SPC mapped in a UMA i.e. a validation site in a UMA matched any of the SPCs mapped within that UMA (in Queensland, up to four SPCs may be mapped and their proportions described within a UMA). Powell’s accuracy of allocating the dominant SPC was 60% i.e. 60% of the time the validation site recorded the SPC mapped as the dominant SPC in the UMA. Such a low number is not necessarily surprising. There is automatically a 50% chance that a randomly located site may encounter a sub-dominant SPC, assuming that a dominant SPC represents 50% or more of a UMA (in some cases, it may be a value as low as 30% of the UMA e.g. 30/30/30/10%). Smith and Calland (2020) reported an accuracy of 80% for the validation of any SPC listed in a UMA, while Brown et al. (2020) reported an accuracy of 69% for the same assessment of their survey. Brown et al. (2020) also assessed the accuracy of allocating the field texture of the A and B horizon, permeability and drainage and achieved results greater than 85% for all attributes.

An international example that used a similar approach to Powell (1979) was Brevik et al. (2003) in the USA. They observed an accuracy of 63% in mapping of the major soils, although it was for a very small area (25 ha), and the existing survey was of 1:1970 scale. It is therefore questionable whether the results are comparable to broader scale surveys that cover large areas. In another small area validation in the USA, Drohan et al. (2003) observed an 80% accuracy for 1:20 000 scale conventional mapping, using detailed independent site investigations.

In the DSM discipline, reported accuracy statistics have also been quite variable and primarily concerned with soil attributes rather than soil types (SPCs). Ellis and Wilson (2009) described an accuracy of 62% and 60% for prediction of permeability and drainage respectively, while Smith and Crawford (2015) reported lower accuracies for the same attributes. They did however record high accuracy (82%) for predicting A horizon pH. Their prediction of plant available water capacity (PAWC) was poor, but this is no surprise in any circumstance, as PAWC is a derivative of other attributes such as soil depth and clay content. These are often difficult to predict, and any derivative of them is also likely to predict poorly. The accuracy results in Bartley et al. (2013) are difficult to fully provide here as they gave an indication that 10 attributes were validated, but only 6 were reported. The overall accuracy (on the basis of their error matrices) for surface permeability, rockiness, soil surface texture and simplified soil group were 63%, 93%, 78% and 57% respectively. Bartley et al. (2013) also provided graphical
Table 1. Summary of accuracy statistics reported for soil surveys in Queensland (noting that the statistics are not necessarily directly comparable due to the variations in accuracy assessment methods used)

| Survey                  | Type                | Scale     | Area (km²) | n     | Km²/site | Attribute/entity validated        | Reported accuracy | Comments                               |
|-------------------------|---------------------|-----------|------------|-------|----------|-----------------------------------|-------------------|----------------------------------------|
| Powell (1979)           | Conventional        | 1:25 000  | 50         | 351   | 0.14     | SPC (any in UMA) SPC (dominant)   | 87%               |                                        |
|                         |                     |           |            |       |          | Permeability Drainage             | 60%               |                                        |
| Ellis and Wilson (2009) | DSM                 | 1:100 000 | 1276       | 81    | 15.75    | Permeability Drainage             | 62%               |                                        |
| Zund (2014)             | DSM (DSMART)        | 1:250 000 | 195 142    | 11937 | 16.35    | Land unit geomorphic category     | 42–93%            | Land unit mapping, not soil mapping    |
|                         |                     |           |            |       |          | PPF                               | 20–24%            |                                        |
|                         |                     |           |            |       |          | PPF class level                    | 37–47%            |                                        |
|                         |                     |           |            |       |          | PPF section level                  | 44–67%            |                                        |
|                         |                     |           |            |       |          | PPF subdivision level              | 49–72%            |                                        |
|                         |                     |           |            |       |          | PPF division level                 | 61–83%            |                                        |
| Odgers et al. (2014)    | DSM (DSMART)        | 1:250 000 | 68 000     | 285   | 238.6    | Dominant SPC Top 3 predicted SPCs | 23%               |                                        |
|                         |                     |           |            |       |          | SPC                               | 49%               |                                        |
|                         |                     |           |            |       |          | pH (A horizon)                     | 74%               |                                        |
|                         |                     |           |            |       |          | pH (B horizon)                     | 82%               |                                        |
|                         |                     |           |            |       |          | Permeability                        | 79%               |                                        |
|                         |                     |           |            |       |          | Drainage                           | 41%               |                                        |
|                         |                     |           |            |       |          | Soil depth                          | 56%               |                                        |
|                         |                     |           |            |       |          | PAWC                               | 68%               |                                        |
|                         |                     |           |            |       |          | Maximum rockiness                   | 44%               |                                        |
|                         |                     |           |            |       |          | Rockiness                           | 63%               |                                        |
|                         |                     |           |            |       |          | Rockiness                           | 93%               | Presence/absence                       |
|                         |                     |           |            |       |          | Surface texture                     | 78%               |                                        |
|                         |                     |           |            |       |          | Permeability                        | 63%               |                                        |
|                         |                     |           |            |       |          | Surface pH                          | 0.67H             |                                        |
|                         |                     |           |            |       |          | Soil depth                          | 0.46H             |                                        |
|                         |                     |           |            |       |          | Simplified soil group               | 57%               |                                        |
| Smith and Crawford (2015)| DSM hybrid         | 1:50 000  | 260        | 34    | 7.65     | Depth of A horizon Permeability     | 81%               | Not assessed as independent accuracy of the attribute Statistic reported here was sum of ‘correct’ and ‘accept’ |
|                         |                     |           |            |       |          | Rockiness                           | 94%               |                                        |
|                         |                     |           |            |       |          | Surface pH                          | 93%               |                                        |
|                         |                     |           |            |       |          | Soil depth                          | 93%               |                                        |
|                         |                     |           |            |       |          | Drainage                            | 64%               |                                        |
|                         |                     |           |            |       |          | Microrelief                         | 90%               |                                        |
|                         |                     |           |            |       |          | Surface salinity                    | 96%               |                                        |
|                         |                     |           |            |       |          | Simplified soil group               | 100%              |                                        |
|                         |                     |           |            |       |          | Surface condition                   | 68%               |                                        |
|                         |                     |           |            |       |          | Surface structure                   | 94%               |                                        |
|                         |                     |           |            |       |          | Surface texture                     | 94%               |                                        |
|                         |                     |           |            |       |          | SPC (any in UMA)                    | 77%               |                                        |
|                         |                     |           |            |       |          | SPC (dominant)                      | 80%               |                                        |
|                         |                     |           |            |       |          | Permeability                        | 58%               |                                        |
|                         |                     |           |            |       |          | Drainage                            | 97%               |                                        |
|                         |                     |           |            |       |          | pH (A horizon)                      | 88%               |                                        |
|                         |                     |           |            |       |          | pH (B horizon)                      | 97%               |                                        |
|                         |                     |           |            |       |          | Soil depth                          | 87%               |                                        |
|                         |                     |           |            |       |          | SPC (any in UMA)                    | 96%               |                                        |
|                         |                     |           |            |       |          | SPC (dominant)                      | 96%               |                                        |
|                         |                     |           |            |       |          | Permeability                        | 96%               |                                        |
|                         |                     |           |            |       |          | Drainage                            | 96%               |                                        |
|                         |                     |           |            |       |          | pH (A horizon)                      | 67%               |                                        |
|                         |                     |           |            |       |          | pH (B horizon)                      | 86%               |                                        |
|                         |                     |           |            |       |          | Soil depth                          | 79%               |                                        |
| Bartley et al. (2013)   | DSM                 | 1:250 000 | 155 000    | 111   | 1396.4   | Rockiness                           | 93%               |                                        |
|                         |                     |           |            |       |          | Surface texture                     | 78%               |                                        |
|                         |                     |           |            |       |          | Permeability                        | 63%               |                                        |
|                         |                     |           |            |       |          | Surface pH                          | 0.67H             |                                        |
|                         |                     |           |            |       |          | Soil depth                          | 0.46H             |                                        |
| Thomas et al. (2018)    | DSM                 | 1:250 000 | 72 000     | 40    | 1800     | Depth of A horizon Permeability     | 81%               |                                        |
|                         |                     |           |            |       |          | Rockiness                           | 94%               |                                        |
|                         |                     |           |            |       |          | Surface pH                          | 93%               |                                        |
|                         |                     |           |            |       |          | Soil depth                          | 93%               |                                        |
|                         |                     |           |            |       |          | Drainage                            | 64%               |                                        |
|                         |                     |           |            |       |          | Drainage                            | 90%               |                                        |
|                         |                     |           |            |       |          | Micrelief                           | 96%               |                                        |
|                         |                     |           |            |       |          | Surface salinity                    | 100%              |                                        |
|                         |                     |           |            |       |          | Simplified soil group               | 68%               |                                        |
|                         |                     |           |            |       |          | Surface condition                   | 94%               |                                        |
|                         |                     |           |            |       |          | Surface structure                   | 94%               |                                        |
|                         |                     |           |            |       |          | Surface texture                     | 77%               |                                        |
| Smith and Calland (2020)| Conventional        | 1:50 000  | 2580       | 82    | 31.46    | SPC (any in UMA) SPC (dominant)     | 80%               |                                        |
|                         |                     |           |            |       |          | Permeability                        | 58%               |                                        |
|                         |                     |           |            |       |          | Drainage                            | 97%               |                                        |
|                         |                     |           |            |       |          | pH (A horizon)                      | 88%               |                                        |
|                         |                     |           |            |       |          | pH (B horizon)                      | 97%               |                                        |
| Brown et al. (2020)     | Conventional        | 1:50 000  | 624        | 34    | 18.35    | SPC (any in UMA) SPC (dominant)     | 74%               |                                        |
|                         |                     |           |            |       |          | Permeability                        | 65%               |                                        |
|                         |                     |           |            |       |          | Drainage                            | 93%               |                                        |
|                         |                     |           |            |       |          | pH (A horizon)                      | 96%               |                                        |
|                         |                     |           |            |       |          | pH (B horizon)                      | 96%               |                                        |
|                         |                     |           |            |       |          | Soil depth                          | 86%               |                                        |
|                         |                     |           |            |       |          | Soil depth                          | 79%               |                                        |

ADSMART = disaggregation and harmonisation of soil map units through re-sampled classification trees (Odgers et al. 2014).

R² Result reported as of predicted versus observed.

Statistic reported here was sum of ‘correct’ and ‘accept’.
representations of predicted versus observed for some attributes, as did Thomas et al. (2018). As we discussed earlier however, the reported accuracy statistics for Thomas et al. (2018) are not directly comparable to other surveys, as their assessment was not an assessment of the soil attributes specifically, but in the context of the land suitability.

Analysis of historical surveys using existing site data

In the absence of independent validation of historical surveys, site data may be used as an approximation of accuracy, while being cognisant of the potential for inherent bias within such an exercise, if the site data was collected as part of the mapping process. The approach does not meet all necessary criteria for an accuracy assessment, but does provide statistics that are indicative of accuracy, as it provides an indication of the likelihood that the mapped soils do actually exist within the delineated units (polygons). The approach may be considered as a proxy assessment of the ‘truthfulness’ of the map. We have undertaken such an assessment of the degree of correlation between recorded SPCs for sites versus UMAs for a variety of historical surveys in Queensland.

Methods

The 12 surveys assessed (Table 2) encompass the most common scales of soil survey in the state i.e. 1:50 000 and 1:100 000. The surveys span from southern inland Queensland to far northern Queensland and were conducted by different staff from the Queensland Government and CSIRO at different stages in their careers, and in a wide range of geomorphic and climatic environments. For example, the Goodar survey (Ross and Crane 1994) encompassed a significant sedimentary rock derived fluvial landscape in a semi-arid environment and the Windeyers Hill survey (Burgess 2003) encompassed a broad, gently undulating landscape derived from igneous and sedimentary rocks of the Bowen Basin in a sub-tropical environment. The Ingham survey (Wilson and Baker 1990) mapped a coastal landscape in a high rainfall, tropical environment. All site and polygonal data were extracted from the Queensland Government Soil and Land Information (SALI) database and spatially joined in ArcGIS to determine the number of sites within a survey boundary that matched the dominant or subdominant SPCs of the corresponding UMAs.

Results

With one exception, the percent match between the SPC of any site in a UMA and the SPCs listed for the UMA was >80%. The weighted mean of all surveys for the same statistic was 86%. Detailed results are provided in Table 2. There were no obvious trends in relation to location, climatic zone or age/experience of the surveyors.

Discussion and recommendations

It is evident that the soil survey discipline needs to take the assessment of survey accuracy more seriously – to quote Stehman and Foody (2019), ‘If accuracy assessment is worth doing, then it most certainly is worth doing well.’ The vagaries of historical assessments cannot be remedied but certainly used as pointers towards best practice in the future. The discipline does not need to reinvent the wheel, but adopt a more collaborative approach around the country and learn from historical experiences, both within and outside of the discipline. This will no doubt take time and resources, hence the need for a more strategic approach to the issue around Australia, not just within Queensland. We provide here some discussion and several recommendations for future validation exercises in the hope that this will promote further discourse and investigations into this topic.

Analysis of historical surveys

The analysis of historical survey data is not an assessment of accuracy but it is certainly indicative of thematic accuracy i.e. how correct is the map. The issue of bias is easily raised in relation to the fact that the sites were described by those undertaking the mapping. This however is a non-issue. Thematic accuracy is a measure of how accurate the map is.

Table 2. Historical surveys assessed for estimated accuracy using intersection of site and polygon data

| Survey                | Authors                    | Scale  | Area (km²) | No. sites | Km²/site | Dominant SPC | Any SPC in UMA | % match | % match |
|-----------------------|----------------------------|--------|------------|-----------|----------|--------------|---------------|---------|---------|
| Bundaberg             | Donnollan and Searle (1999)| 1:50 000 | 1233.81    | 3163      | 0.39     | 74           | 83            |         |         |
| Cardwell-Tully-Innisfail | Cannon et al. (1992)    | 1:50 000 | 2840.14    | 1449      | 1.96     | 53           | 73            |         |         |
| Toowoomba-Oakey       | Biggs et al. (2001)       | 1:50 000 | 508.75     | 840       | 0.61     | 66           | 81            |         |         |
| Logan-Albert          | Smith and Calland (2020)  | 1:50 000 | 2589.04    | 1428      | 1.81     | 72           | 97            |         |         |
| Mossman               | Murtha (1989)             | 1:50 000 | 1169.54    | 226       | 5.17     | 66           | 82            |         |         |
| Atherton              | Malcolm et al. (1998)     | 1:50 000 | 1679.65    | 1972      | 0.85     | 78           | 85            |         |         |
| Miara-Winfield        | Brown et al. (2020)       | 1:50 000 | 625.08     | 821       | 0.76     | 73           | 95            |         |         |
| Goodar                | Ross and Crane (1994)     | 1:100 000 | 2717.31    | 640       | 4.25     | 81           | 90            |         |         |
| Windeyers Hill        | Burgess (2003)            | 1:100 000 | 2843.56    | 1793      | 1.59     | 81           | 94            |         |         |
| Childers              | Wilson (1997)             | 1:100 000 | 2560.76    | 1085      | 2.36     | 70           | 82            |         |         |
| Ingham                | Wilson and Baker (1990)   | 1:100 000 | 3196.20    | 652       | 4.90     | 67           | 84            |         |         |
| Kilcummin             | Shields and Williams (1991)| 1:100 000 | 2853.74    | 1592      | 1.79     | 71           | 84            |         |         |

Note 1: Some project sites fall outside of final mapping areas and have been excluded from the analysis. Only sites with an SPC were used. Note 2: Sites from other projects that intersect the survey area have been included in this analysis.
does the map represent what is truly in the ground? The process of mapping is primarily driven by stereo-air photo interpretation and sites are used to validate what the soil surveyor has predicted is there. There is of course a high degree of correlation between sites and the map because the map is to a large degree correct – it has been extensively validated on ground. Consequently, it is not surprising to see statistics in excess of 80% and why the approach may be considered as a reasonable estimate of thematic accuracy.

An obvious issue associated with the analysis we undertook is positional error. We did not buffer the location of sites before analysis, despite the potential for positional error in either the site or polygon boundary locations. An analysis of the impact for one survey (Cardwell-Tully-Innisfail) was undertaken. Including a site buffer of 50 m radius increased the match to dominant SPCs by 2% and the match to any SPCs by 3%. For surveys with numerous sites in transitional zones, buffering will have a larger impact.

Validation approaches – recommended practice

There is no question that the most appropriate method for determining accuracy is statistically-based, independent, external validation – it should be taken as a given for any modern survey (DSM or conventional). Independent parties undertaking validation for accuracy assessment must however be suitably skilled (preferably of a high skill level) and operating to a clear set of guidelines that encompass all aspects, from the determination of validation sites to the criteria for acceptance. Guidelines must be explicit enough to reduce the likelihood of an ambiguous or incorrect outcome at all levels within the validation process. The soil survey discipline must strive towards a common approach to such guidelines so that statistics are comparable across surveys. The literature suggests that error matrices are an invaluable tool for analysing and communicating results of accuracy assessments. These, coupled with other visual communication methods and relevant statistical approaches, will significantly improve the ability to convey soil survey accuracy to a wide range of clients.

As we have alluded to throughout this text, guidance for accuracy assessment of soil survey has been described in the literature previously. Soil survey is not alone in its need to undertake map validation and other disciplines such as land use mapping, vegetation mapping and remote sensing have already made progress in this area. The work of Stehman and Foody (2019), while derived for remote sensing, provides a good basis to follow for soil survey, given our discipline relies inherently upon remotely sensed data. Those authors described six good practice criteria to guide accuracy assessments for remote sensing. Their criteria are similar to those listed by Hewitt et al. (2008) and Minasny et al. (2008) in The Guidelines, and should be regarded as key pointers to follow in any soil survey accuracy assessment. Stehman and Foody (2019) suggested that an accuracy assessment must be: map relevant, statistically rigorous, quality assured, reliable and transparent. In the following text we explore these concepts and provide associated recommendations in the context of soil survey.

Map relevant

The central concept of map relevance as described by Stehman and Foody (2019) was that the accuracy assessment must account for the spatial distribution (proportional areas) of the entities being validated. The use of a cLHC to achieve this outcome is well established but further evaluation and refinement is needed, in particular in relation to the assessment of the impacts of constraints to road buffers. There is however a secondary component to map relevance, which is that the entity being assessed must be conceptually the same as that mapped. For example, a validation exercise should not deliberately set out to assess a soil attribute, such as surface pH, for a map that was created to depict soil types – while there is likely to be an inherent relationship between the two that can be appropriately described, one entity is a soil attribute and the other is an informal taxonomic construct. A soil attribute being assessed may have no bearing on the defining characteristics of the conceptual entity being mapped. The accuracy of specific attributes may be estimated as a by-product of validating an SPC, but this is a different outcome to conducting a validation exercise targeted at a specific attribute, for the primary reason that an area-weighted sample distribution for an attribute may be different to the area-weighted distribution derived for SPCs. An attribute accuracy assessment derived from validation of an SPC, while useful, can only be regarded as an approximation of accuracy unless analysis is undertaken to determine the statistical rigour of the assessment of the attribute.

In a similar vein, it is inappropriate to validate a formal taxonomic class. In Queensland (and Australia in general), soil surveyors do not map to a formal taxonomy. The entity being mapped is an informal soil type or other landscape unit which often encompasses a formal taxonomic range. A formal taxonomy such as the Australian Soil Classification (Isbell and NCST 2016) is an artificial construct based on a particular knowledge base. SPCs are also an artificial construct, but are taken from the landscape being mapped. They are therefore highly location specific, and as a consequence are appropriate to validate (given they are the entity mapped).

The statistic of the degree of match between the mapped soil type for a polygon and that observed at a validation site within it is a measure of thematic accuracy. As discussed earlier, there are two metrics that may be reported – a match with the dominant soil type (which would typically appear as the mapcode on the published map) or a match with any of the described soil types in the UMA. We regard the latter as the primary statistic and the former as the secondary statistic, although it is acknowledged that the dominant soil type is invariably what is communicated on a map. Similarly, in DSM outputs, comparison may be made against the most likely predicted soil type or several of them. It is unrealistic to expect that the thematic accuracy can achieve 100%, other than for surveys at scales more detailed than 1:10 000. At less detailed scales, spatial units delineated are typically not homogenous. Furthermore, minor soils comprising <10% of a UMA are typically not recorded. Thus it is quite possible for a site, whether it be an original one by the soil surveyor or an
independent validation site, to occur on a minor soil type that is not recorded for the UMA. For example, a soil surveyor may have specifically chosen to investigate a small, atypical feature (minor soil type) in a UMA, and not describe a site in the dominant soil, due to a high degree of confidence in their determination of the dominant soil through other information (vegetation, geology, photo-pattern, land use etc.). In such circumstances, the soil description site in the UMA is by definition non-representative of the UMA.

In DSM surveys, it is common to present a measure of the confusion index for any pixel e.g. Odgers et al. (2014) and Thomas et al. (2018). While a potentially useful statistic, it only compares the two most likely soils and does not account for the possibility that the top two soils predicted are both incorrect. This illustrates one of the key challenges within accuracy assessment – ensuring that methods are comparative between DSM and conventional surveys. If this is not achieved, the value of each approach cannot be truly evaluated and more importantly, consumers of soil survey products will continue to be challenged by the differing terms used. Arrouays et al. (2020a) proposed the solution to this challenge is that clients require training in understanding the use of uncertainty statistics – we suggest that instead, a long-understood paradigm of soil extension be revitalised i.e. the need to communicate in the language of the client.

**Statistically rigorous**

A validation sampling strategy should implement a probability sampling design (Brus et al. 2011; Biswas and Zhang 2018). Many authors, including the aforementioned and McKenzie et al. (2008b), have reviewed the methods available. As we discussed earlier, there can be issues when such designs are translated into practice. Authors such as Cambule et al. (2013), Clifford et al. (2014) and Kidd et al. (2015) have provided suggestions for improvements to sampling designs. Thomas et al. (2018) also used a novel approach, in which their validation was specifically targeted using their model error. It is apparent though that further evaluation of the impact of constraints to road corridors and the point of diminishing returns in validation must occur in order to develop robust, reliable approaches to determining the number and location of validation sites.

Determination of a validation sample size must account for survey size as well as the variability in the environmental space and map entities concerned. The number of sites by itself is not a relevant metric. In Table 1, we included the ratio of sites to survey size as this is a common metric used for describing the ‘intensity’ of a survey. The data illustrates a wide variation, from 1 site per 0.14 km² in the case of Powell (1979) to 1 site per 1800 km² in the case of Thomas et al. (2018).

In the absence of a more detailed understanding of the minimum number of sites required for validation, we recommend the use of site density as a determinant, as it has a logical basis, is easy to derive in a transparent manner and is familiar to practitioners. In terms of an achievable number of sites, we suggest between 5 and 10% of the theoretical number of sites for a survey area, computed from the minimum acceptable values in table 14.4 of Schoknecht et al. (2008). Further analysis of existing site data and validation efforts in new surveys will no doubt test whether this initial suggestion is valid. While some may baulk at the number of sites, consideration must also be given to the type of sites used in validation. Rapid mapping observations with limited attributes described (Class IV sites as per table 14.2 of The Guidelines) may be suitable for validation, in combination with full profile descriptions (Class I sites), leading to an increase in the number of validation sites practically achievable. The conditions under which Class IV sites are used in validation should however be clearly described.

**Quality assured**

Quality assurance (QA) and quality control (QC) are generally poorly documented in soil survey reports, particularly in relation to validation exercises. Failure to observe quality assurance protocols can lead to significant errors and ambiguity in accuracy assessment. There are many elements within accuracy assessment practices that must be more rigorously documented – for example, ensuring (and reporting) that field equipment such as pH and electrical conductivity meters record accurate, repeatable and reliable results. While all field practitioners in Australia follow the Australian Soil and Land Survey Field Handbook (NCST 2009), skill level does vary with individual, consequently efforts must be made to standardise those undertaking the accuracy assessment before commencing. Surprisingly little formal evaluation has been published in Australia in relation to the variance between practitioners, but it is an acknowledged issue – for example variance in determination of field texture and assessment of site permeability and drainage. Unpublished work within the Queensland Government and exercises undertaken in Australian Collaborative Land Evaluation Program workshops during the 1990s has highlighted that even among experienced practitioners, there are occasionally some ‘unusual’ variances. More effort should be expended to both assess and document the QA/QC of soil surveys in general and more specifically in relation to accuracy assessments.

The question of whether the parties undertaking the accuracy assessment should be external to the survey is worthy of investigation. From first principles, an unbiased outcome (in a personal rather than statistical sense) is more likely if those conducting the validation were not involved in the mapping. It may be argued that as long as the sites are selected in a statistically unbiased manner then it does not matter who does the accuracy assessment. There is also an argument, in the case of validating soil types, that external parties may not be able to determine the correct SPC. Others may suggest it is too difficult to find the necessary staff to undertake the work. We suggest however that all of these points lack substance, on the basis of practical experience in recent Queensland surveys that included external validation. Use of external parties for accuracy assessment is possible if designed into the survey and should be regarded as best practice.

**Reliable**

Achieving reliable accuracy assessment involves controlling those features of the assessment protocol that
are subject to variability (Stehman and Foody 2019). Apart from those matters discussed above, under the topics of QA and statistical rigour, this subject particularly concerns acceptance criteria, whether they be for attributes or taxonomic classes.

The tiered acceptance criteria (correct, accept or fail) used by Thomas et al. (2018) provides an obvious solution to one of the key challenges created in validation. It is a useful approach that could be used more widely in soil survey validation. The bounds of the criteria should however be clearly defined and documented before validation occurring, to ensure those conducting the validation are not faced with ambiguous choices. The criteria should not be set using an unduly wide window, as this will lead to a misperception of the accuracy. Constraining validation to a yes/no or presence/absence proposition, while technically correct and useful in some circumstances, can create statistics that are perhaps not terribly informative and overly generalised. There are multiple ways in which the window of acceptance around a central value may be determined. For example, conventional statistical approaches, such as ±1 standard deviation or ± a specified percentage. For categorical values, it may be ±1 class. This is an area that requires further evaluation of the implications of differing approaches.

Validation criteria must also be inherently objective and a function of the feature rather than a secondary outcome. We discussed earlier the approach used by Thomas et al. (2018), in which the determination of acceptance was not related to inherent properties of the soil attribute concerned. Instead, it was dictated by the relationship of the difference the value created in a land suitability framework, and specifically whether the difference between predicted and observed values led to a difference in the calculated land suitability. This ties the validation to the end-use of the data, rather than being an independent assessment of the quality of the underlying soil data. Although the evaluation of the implications of accuracy (does a difference matter?) are certainly valid and to be encouraged, this should not be regarded as a substitute for appropriate validation methods. It should instead be regarded as an additional form of accuracy assessment.

Validation of SPCs creates greater challenges as the entity of concern is not a singular value – an SPC is represented by a modal soil profile description with an associated range of values. Thus, in a validation exercise, some attributes may match and some not – for example the texture, horizonation and segregations may match, but not the colour. While the variation between SPCs is regarded as greater than the variation within (Powell 2008), the reality is that the difference between SPCs can sometimes be very minor – for example a difference in the size of structural units. There is also the potential for overlap between SPCs at their outer ranges. This presents a significant challenge for validation and there is no simple solution other than deriving clear guidelines before validation in order to avoid ambiguous outcomes. For example, one option may be to amalgamate SPCs using clustering algorithms (Chamberlain et al. 2020). In detailed surveys, the difference between some soil types may be a function of soil chemistry rather than morphology, thus determining a match requires extra cost and effort in relation to soil sampling. Such requirements should be identified early in a project to ensure that sufficient resources are allocated to meet this need. In some cases, a field surrogate may be used e.g. field pH for laboratory pH, but this can only occur if the necessary evaluation of the relationship between the parent attribute and surrogate is confirmed to be robust and reliable.

**Transparent**

It is imperative that accuracy assessments of soil surveys are fully described in the relevant survey reports with all aspects (methods and results) clearly documented. No survey to date appears to have achieved this goal but it is one that must be aspired to. Hewitt et al. (2008) observed that surveyors may find validation a challenging experience. It should however be taken in the same manner as effective peer review of literature – an opportunity to evaluate and improve. Surveyors should avoid the temptation to gloss over or fail to report poor accuracy results and should not make overly generous statements in relation to the statistics presented. Transparency also involves more fulsome evaluation of the implications of assumptions or compromises that are frequently made in accuracy assessments. The increasing trend towards open data publication is likely to also assist in aspects of transparency.

There is a growing trend, within DSM projects, of recommendation of the use of ‘expert opinion’ in validation (Holmes et al. 2015; Thomas et al. 2018; Arrouays et al. 2020b; Bui et al. 2020). The scrutiny of a soil map by colleagues before publication has long been standard practice in conventional soil survey. The use of experts in the review process of a survey before publication should be regarded as standard practice but expert opinion cannot be regarded as a form of accuracy assessment. It is not a transparent process nor is it repeatable, reliable or quantitative, which seems at odds with the paradigm of DSM being a more objective, unbiased approach than conventional survey (McBratney et al. 2003).

**Reproducible**

The criteria for reproducibility is essentially about documentation, thus it interacts with the criteria for transparency. All methods should be documented sufficiently that a third party could undertake a parallel validation and achieve a similar outcome. In particular, it is critical that a detailed description of the response design is provided (Stehman and Foody 2019). To date, surveys in Queensland involving validation have not met this goal. Stehman and Foody (2019) list the following as key elements to be documented: (1) definition of spatial assessment unit (2) definition of classes; (3) sources of reference data; (4) specific information collected from each source; (5) rules for assigning reference values; (6) specification of how agreement between the map value and reference value is achieved. Documenting these elements is well and truly within the capacity of any modern survey.
Communication of accuracy results

The criteria described above lead to a discussion on the topic of communication. It is clear that a key challenge in relation to accuracy is the plethora of terms that can be used to describe it. Each term has a specific nuance that can be highly relevant and informative, but the use of many terms can lead to confusion. Hewitt et al. (2008) notably advised that accuracy must be reported ‘...in a manner the user can understand’, but in general, conventional surveys have not reported accuracy directly. Most DSM surveys report multiple accuracy and uncertainty related statistics, such as root mean square error, confusion index, \( r^2 \) etc. While there are logical reasons for this, few of these statistics can be understood by ‘average users’ and they are not always an assessment of accuracy (rather an estimate of model uncertainty). While DSM surveys offer the ability to indicate numerous statistics regarding the level of uncertainty associated with predictions, this is of little value if clients cannot understand the terms used or don’t understand how the uncertainty affects the usefulness of the data. Arrouays et al. (2020a) noted the problems faced by the DSM discipline in relation to the communication of uncertainty maps. Similar problems exist for representations of accuracy.

The simplest representation of accuracy is a graphical display of predicted versus observed values, such as provided by Bartley et al. (2013). There is considerable value in such simple representations as they are easily understood. Error matrices have been widely explored and used in the remote sensing discipline for decades (Story and Congalton 1986; Congalton and Green 1999; Stehman and Foody 2019). They are also applicable to the validation of soil surveys (Brus et al. 2011) as they are a simple, yet informative tool. In an error matrix, the overall accuracy is derived from the primary diagonal of the matrix but other statistics such as errors of omission and commission, users’ accuracy and producers’ accuracy may also be derived (Story and Congalton 1986). For example, the producers’ accuracy would be the percentage of times a given SPC validation site was mapped correctly out of the number of instances of that SPC among all of the validation sites. The users’ accuracy would be the percentage of times a specific map unit was correctly attributed out of the number of instances of that map unit. Both Bartley et al. (2013) and Thomas et al. (2018) have included them, although a more fulsome use of them is likely to be beneficial.

A value of error matrices is that they may be explored at multiple levels, from the broad/geomorphic scale down to specific attributes. For example, given conventional soil maps use a map reference with a geomorphic construct, the highest order of validation is whether or not the mapped soils fit within the correct geomorphic group e.g. only floodplain SPCs are mapped on a floodplain. High order validation statistics are very useful to convey in simple terms which parts of the landscape were mapped better/worse than others. This approach was used by Smith and Crawford (2015) in their hybrid DSM survey in south-east Queensland and also explored by Zund (2014). The merits of and ways in which error matrices can be used clearly requires more attention from the soil survey community.

Rosenfield and Fitzpatrick-Lins (1986) promoted the use of the kappa statistic in relation to error matrices. Some DSM soil surveys, such as Thomas et al. (2018), have reported the statistic. Stehman and Foody (2019) and some earlier authors have however recommended unmistakably against the use of kappa, indicating it is poorly founded. This suggests that the usefulness of the kappa statistic also requires further investigation in the context of soil survey.

We have not canvassed here all possible ways and means of visually representing accuracy, but merely touched upon a few examples. A concerted effort is obviously required in many elements of this topic. Irrespective of whether the measure is accuracy or uncertainty, or it is derived from DSM or conventional survey, it is important that the purpose and relevance of all metrics are clearly explained in plain English. The importance of plain English communication in science is well acknowledged (Grealish et al. 2015; Rakedzon et al. 2017). The provision of geostatistical terms in spatial metadata is not an adequate approach to cater for the needs of all clients. All users should be clearly informed of how each statistic or statement was derived. The need for the use of technical terminology where appropriate is not disputed, but this must not be regarded as the only means by which accuracy (and related terms) are communicated.

A benchmark

The benchmark for thematic accuracy among other mapping disciplines (vegetation and land use) in Queensland is 80%. These two disciplines have an advantage in comparison to soil survey, in that they map surficial features, thus it should be easier to derive accurate maps. Our review of published accuracy assessment results and the analysis we have undertaken of existing survey data suggests however that this benchmark is readily achieved in conventional soil survey (acknowledging the limitations of our analysis). While there is an argument that the benchmark could be adjusted with the scale of the survey (a higher benchmark for more detailed surveys), we recommend that surveys in the range of 1:25 000 to 1:100 000 should use 80% as the accuracy benchmark until such time as further analysis suggests otherwise. Although several historical surveys have not achieved this level of accuracy, it should be regarded as a minimum standard to be met for future surveys, consummate with advances in available mapping inputs e.g. DEM and imagery.

Although our analysis was primarily related to the mapping of soil types (SPCs), data from both conventional and DSM surveys suggest it is also possible to reach this level of accuracy for attribute mapping. It is acknowledged that the mapping of soil types (SPCs) may not automatically lead to accurate mapping of attributes. However, the fact that SPCs are concepts created around a modal set of attributes points towards a likelihood that they may successfully represent the distribution of attributes in an accurate manner. The degree to which this is true must be further evaluated rather than assuming that the two are mutually exclusive.

Some may argue that benchmarks for attribute mapping are not needed as long as the accuracy or uncertainty of a map is
spatially represented – the client can choose what to use or not to use e.g. Thomas et al. (2018). We would argue however that this is a short-sighted view that takes no account of clients who are not au fait with the intricacies of pedology, mapping, accuracy and geostatistics. Moving forwards, it is in the interest of the discipline to produce mapping products of a high quality that clients can automatically rely upon, rather than create a perception that it is acceptable to publish inaccurate material as long as the uncertainty is described.

Utilisation of a benchmark relies upon further exploration and clarification of a range of issues, in order to develop consistent methods for validation. Most importantly, standards for acceptance criteria must be developed, as whether or not a benchmark is reached is related more to the acceptance criteria used in validation than what the benchmark value is.

**Other relevant accuracy-related terms**

The discussion above has concerned accuracy assessment in its purest form, but as we noted at the beginning of this text, accuracy has many meanings to a range of people and the general use of the term has historically encapsulated many related terms. While these are not strictly a component of accuracy, it is worthwhile to comment upon their use within this text, in the hope that it leads to a more objective and consistent use of them in conjunction with explicit statistics of accuracy. Many of these terms have been in use in soil survey for decades and they are also described in The Guidelines. A more consistent approach to their use will however undoubtedly assist clients. We recommend that statistics concerning the following should be included as a minimum for any published land resource survey.

- Positional accuracy
- Minimum resolvable limit
- Scale/equivalent scale
- Observation density

Positional accuracy is rarely described in land resource surveys, but should be, as it is a relatively simple concept that is highly informative to users of the information. In the case of vector data, it is a function of the way in which the linework was captured. In most historical soil surveys in Queensland, this involved primary capture (wax pencil on aerial photo) and secondary capture (digitisation). Estimates using historical linework suggest the positional accuracy is in the order of ±1 mm, which translates to ±50–250 m for 1:50 000 to 1:250 000 surveys. This does not allow for inherent error associated with the choice of location of the line. The positional accuracy of linework was once tied closely to the scale of mapping, as the scale of aerial photographs used for primary capture was concomitant with the scale of the survey. The wide availability of high resolution imagery today means that positional accuracy has the potential to be significantly improved in coarse scale surveys e.g. it is possible to use 1:25 000 equivalent imagery to locate boundaries precisely for a 1:250 000 survey. As a consequence, positional accuracy is no longer inherently tied to survey scale and a definitive statement should be made for a survey. This is also relevant to validation as the positional error of polygon boundaries should be considered when determining the spatial window for validation – as in Brown et al. (2020) and Smith and Calland (2020). The vegetation mapping discipline in Queensland explicitly describes the confidence of the location of polygonal boundaries (Neldner et al. 2019). While this is a very useful approach and a practice that could easily be adopted for soil survey, we recommend that a numerical expression of linework positional accuracy is also reported.

Scale remains highly relevant as the simplest plain English concept that conveys a general sense of accuracy. Users of maps generally have lesser expectations of a 1:500 000 scale survey than a 1:50 000 survey, as there is an implied accuracy that is associated with scale and the dimension of the mapped entities. Modern surveys such as Brown et al. (2020) and Smith and Calland (2020) have embraced the concept of the graphical depiction of variable scale within the survey – we have reproduced an example here as Fig. 1. This practice is to be encouraged as it provides clear, simple communication to clients.

![Fig. 1. Visual depiction of variable scale in a conventional soil survey, coupled with an indication of site locations. Reproduced from the soil map accompanying Brown et al. (2020) published by Department of Natural Resources Mines and Energy, CC BY 4.0. Available at https://www.publications.qld.gov.au/dataset/soils-and-agricultural-suitability-of-the-miara-winfield-area/resource/33107a80-f7bf-41e4-a172-cc0c32e36df5](https://www.publications.qld.gov.au/dataset/soils-and-agricultural-suitability-of-the-miara-winfield-area/resource/33107a80-f7bf-41e4-a172-cc0c32e36df5)
Raster products from DSM surveys provide an extra challenge, as there is a need to derive an equivalent scale, which is an accepted term within the ANZLIC metadata framework. Lagacherie and McBratney (2006) provided a suggested correlation between pixel size and cartographic scale, but their calculations are based upon an assumed minimum resolvable limit that is one twentieth of the norm accepted in Australia (Reid 1988), and thus appear overly optimistic. There is also an associated issue in relation to the representation of single pixels compared with the concept of a minimum legible area. In remote sensing, there is an understanding that no single pixel by itself represents a valid resolvable entity, partly due to the potential for random noise (Congalton and Green 1999). Lagacherie and McBratney (2006) suggested a 2×2 minimum resolution, but in published DSM works to date, there is inherently an assumption that each individual pixel is valid (Grunwald et al. 2011) in terms of both inputs and outputs. While there may be an uncertainty associated with a specific pixel, this is not obvious to the user. This matter requires further evaluation and standardisation in the interests of improved transparency and communication.

Minimum resolvable limit and site density are both concepts that have been in use for many years and described in both editions of The Guidelines (McDonald et al. 1984; McKenzie et al. 2008a). Despite them being simple concepts, they are often not reported in surveys. The minimum resolvable limit is very much dictated by conventional cartography and what the human eye can see on a hardcopy map. It is essentially fixed, irrespective of map type. The minimum resolvable limit is regarded as ~0.2 cm², with some variation in dimensions depending on the shape of the unit (Reid 1988). There is an argument that a smaller single entity is more visible on a conventional thematic map than on a raster product due to the crisp boundaries associated with thematic maps. The digital era has however removed the self-limiting nature of the minimum resolvable limit, consequently clear statements should be made for any survey (DSM or conventional) as to the size of the minimum mappable area (which is a function of the minimum resolvable limit and the scale of the survey).

Site density has long been used to make inferences about accuracy of soil surveys because of an obvious relationship – the greater the density of observations the greater the on-ground validation of what has been mapped and therefore an implied improvement in accuracy. Site density is generally reported as a singular metric of sites per unit area (or the inverse statistic) for the whole of the survey area. This is appropriate where sites are approximately evenly distributed across a survey area, but becomes less relevant in surveys where there are significant internal variations in site distribution. There is potentially greater value therefore in generating spatial representations of site density across survey areas rather than a singular statistic. Site density is not a measure of survey accuracy, but a more effective representation of it can assist with the holistic communication of the factors influencing the accuracy of a survey.

Key issues for attention

Of the matters we have discussed, it is apparent that the two key issues most urgently requiring investigation and standardisation are methods for determining the minimum acceptable number of sites for accuracy assessment and the nature of acceptance criteria. Exploration of these matters within the academic literature will not be sufficient to turn theory into practice. A concerted effort must be made to partner with a range of soil survey teams (both conventional and DSM) to turn theory into practical guidelines that can be used on a day to day basis. Soil surveyors must also make greater effort to encapsulate accuracy assessment with the design of surveys.

Conclusions

The need for accuracy assessment of soil survey has been well described historically, but it remains poorly implemented in either conventional or DSM surveys in Queensland and elsewhere in Australia. While some progress has been made in the last 20 years, considerable effort is still required in the further development and application of consistent methods of accuracy assessment. Data from limited historical assessments, our assessment of several conventional surveys and industry standards for similar disciplines, suggest that a benchmark of 80% accuracy is realistic for any soil type or attribute mapped within a spatial entity. We recommend this as a minimum standard for any soil surveys between 1:25 000 and 1:100 000 scale and potentially for landscape surveys at 1:250 000. Greater care and effort should be expended to ensure that validation is integral in modern surveys (either conventional or DSM) and to ensure that their accuracy is independently and appropriately assessed. No survey is going to be perfectly accurate, but it is essential that the accuracy of surveys is assessed and communicated correctly, consistently and appropriately.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

This manuscript has benefited from discussions with David Morrison, Sue-Ellen Dear (both Queensland Government) and comments from reviewers. This research did not receive any specific funding.

References

ABARES (2011) Guidelines for land use mapping in Australia: principles, procedures and definitions. 4th Edition. Australian Bureau of Agricultural and Resource Economics and Sciences, Canberra.
Arrouays D, McBratney A, Bouma J, Libohova Z, Richer-de-Forges AC, Morgan CLS, Roudier P, Poggio L, Mulder VL (2020a) Impressions of digital soil maps: The good, the not so good, and making them ever better. Geoderma Regional 20, doi:10.1016/j.georeg.2020.e00255
Arrouays D, Poggio L, Salazar Guerrero OA, Mulder VL (2020b) Digital soil mapping and GlobalSoilMap. Main advances and ways forward. Geoderma Regional 21, doi:10.1016/j.georeg.2020.e00265
Bartley R, Thomas M, Clifford D, Philip S, Brough D, Harms B, Willis R, Gregory L, Glover M, Moodie K, Sugars M, Eyre L, Smith D, Hicks W,
Petheram C (2013) Land suitability: technical methods. A technical report to the Australian Government for the Flinders and Gilbert agricultural land resource assessment (FGARA) project. CSIRO.

Bazaglia Filho O, Rizzo R, Lepsch IF, do Prado H, Gomes FH, Mazza JA, Demattê JAM (2013) Comparison between detailed digital and conventional soil maps of an area with complex geology. Revista Brasileira de Ciência do Solo 37(5), 1136–1148. doi:10.1590/0100-06832013000500003

Biggs AJW, Clark J, Hall IR (2001) Soil survey of the Eastern Darling Downs (Westbrook–Highfields–Oakey area). Queensland Department of Natural Resources and Mines.

Biswas A, Zhang Y (2018) Sampling Designs for Validating Digital Soil Maps: A Review. Pedosphere 28(1), 1–15. doi:10.1016/S1002-0160(18)60001-3

Brevik EC, Fenton TE, Jaynes DB (2003) Evaluation of the accuracy of a central Iowa soil survey and implications for precision soil management. Precision Agriculture 4(3), 331–342. doi:10.1023/A:1024960705861

Brown W, McClurg J, Sugars M (2020) Soils and Agricultural Suitability of the Miara–Winfield Area, Wide Bay–Burnett, Queensland. Queensland Department of Natural Resources, Mines and Energy, Bundaberg.

Brus DJ, Kempen B, Heuvelink GBM (2011) Sampling for validation of digital soil maps. European Journal of Soil Science 62(3), 394–407. doi:10.1111/j.1365-2389.2011.01364.x

Bui EN, Searle RD, Wilson PR, Philip SR, Thomas M, Brough D, Harms B, Hill JV, Holmes K, Smolinski HJ, Van Gool D (2020) Soil surveyor knowledge in digital soil mapping and assessment. Geoderma Regional 22, e00299. doi:10.1002/geo3.2020.e00299

Burgess JW (2003) Land resource assessment of the Windyeyers Hill area, Isaac-Connors and Mackenzie River catchments, central Queensland. Queensland Department of Natural Resources and Mines. QNRM02189.

Burrough PA, Van Gaans PFM, Hootsmans R (1997) Continuous classification in soil survey: Spatial correlation, confusion and boundaries. Geoderma 77(2–4), 115–135. doi:10.1016/S0016-7061(97)00018-9

Cambule AH, Rossiter DG, Stoovogel JJ (2013) A methodology for digital soil mapping in poorly-accessible areas. Geoderma 192(1), 341–353. doi:10.1016/j.geoderma.2012.08.020

Cannon MG, Smith CD, Murtha GG (1992) ‘Soils of the Cardwell-Tully area, North Queensland.’ (CSIRO Division of Soils: Townsville.)

Carré F, McBratney A, Mayr T, Montanarella L (2007) Digital soil assessments: Beyond DSM. Geoderma 142(1–2), 69–79. doi:10.1016/j.geoderma.2007.08.015

Chamberlain T, Silburn DM, Biggs AJW, Rattray D, Ellis R (2020) A soil parameter framework to support paddock-scale modelling in the Great Barrier Reef catchments, Paddock-Scale Modelling for Grain Cropping, Technical Report, Great Barrier Reef Reef Report Card 2018. Queensland Department of Natural Resources, Mines and Energy.

Clifford D, Payne JE, Pringle MJ, Searle R, Butler N (2014) Pragmatic soil survey design using flexible Latin hypercube sampling. Computers & Geosciences 67, 62–68. doi:10.1016/j.cageo.2014.03.005

Cohen J (1960) A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 20(1), 37–46. doi:10.1177/001316446002000104

Congalton RG, Green K (1999) ‘Assessing the accuracy of remotely sensed data: principles and practices.’ (Lewis Publishers: Boca Raton, FL).

Donnollan TE, Searle RD (1999) ‘Land resources of the Burnett region, Queensland Part 3 North Burnett.’ (Queensland. Department of Natural Resources, Qld.)

Drohan PJ, Ciolkosz EJ, Petersen GW (2003) Soil survey mapping unit accuracy in forested field plots in northern Pennsylvania. Soil Science Society of America Journal 67(1), 208–214. doi:10.2136/ssaj2003.2080

Ellis R, Wilson PR (2009) Modelled soil and landscape attributes of the Cooloola Area, South-East Queensland. Queensland Department of Environment and Resource Management.

Gooch MF (2011) Scale in GIS: An overview. Geomorphology 130(1–2), 5–9. doi:10.1016/j.geomorph.2010.10.004

Grealish GJ, Fitzpatrick RW, Hutson JL (2015) Soil survey data rescued by means of user friendly soil identification keys and toposequence models to deliver soil information for improved land management. GeoResJ 6, 81–91. doi:10.1016/j.grj.2015.02.006

Grunwald S, Thompson JA, Boettiger JL (2011) Digital soil mapping and modeling at continental scales: Finding solutions for global issues. Soil Science Society of America Journal 75(4), 1201–1213. doi:10.2136/sssaj2011.0025

Hewitt AE, McKenzie NJ, Grundy MJ, Slater BK (2008) 18. Qualitative survey. In ‘Guidelines for surveying soil and land resources. Second Edition.’ (Eds NJ McKenzie, MJ Grundy, R Webster, AJ Ringrose-Voase.) pp. 285–306. (CSIRO Publishing: Melbourne).

Holmes KW, Griffin EA, Odgers NP (2015) Large-area spatial disaggregation of a mosaic of conventional soil maps: Evaluation over Western Australia. Soil Research 53(8), 865–880. doi:10.1071/SR14270

Isbell RF, National Committee on Soil and Terrain (NCST) (2016) ‘The Australian soil classification. Second Edition.’ (CSIRO Publishing: Melbourne).

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. doi:10.1016/j.geodrs.2014.11.002

Lagacherie P, McBratney AB (2006) Chapter 1 Spatial Soil Information Systems and Spatial Soil Inference Systems: Perspectives for Digital Soil Mapping. In ‘Developments in Soil Science’. Eds P Lagacherie, AB McBratney, M Voltz.) Vol. 31, pp. 3–22. (Elsevier).

Malcolm, DT, Nagel, BKA, Sinclair, I, Heiner, IJ (1998) Soils and agricultural land suitability of the Atherton Tablelands, North Queensland. Department of Natural Resources. DNRQ980091.

McBratney AB, Mendoza Santos ML, Minasny B (2003) On digital soil mapping. Geoderma 117(1–2), 3–52. doi:10.1016/S0016-7061(03)00223-4

McDonald RC, Isbell RF, Speight JG, Walker J, Hopkins MS (1984) ‘Australian Soil and Land Survey Field Handbook.’ (Inkata Press: Melbourne).

McKenzie NJ, Grundy MJ, Webster R, Ringrose-Voase AJ (2008a) ‘Guidelines for surveying soil and land resources. Second Edition.’ (CSIRO Publishing: Melbourne).

McKenzie NJ, Webster R, Ryan PJ (2008b) 20. Sampling using statistical methods. In ‘Guidelines for surveying soil and land resources. Second Edition’. (Eds NJ McKenzie, MJ Grundy, R Webster, AJ Ringrose-Voase.) pp. 319–326. (CSIRO Publishing: Melbourne).

Minasny B, Bishop TFA (2008) 24. Analysing uncertainty. In ‘Guidelines for surveying soil and land resources. Second Edition’. (Eds NJ McKenzie, MJ Grundy, R Webster, AJ Ringrose-Voase.) pp. 383–394. (CSIRO Publishing: Melbourne).

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

Minasny B, McBratney AB, McKenzie NJ, Grundy MJ (2008) 22. Predicting soil properties using pedotransfer functions and environmental correlation. In ‘Guidelines for surveying soil and land resources. Second Edition’. (Eds NJ McKenzie, MJ Grundy, R Webster, AJ Ringrose-Voase.) pp. 349–367. (CSIRO Publishing: Melbourne).

 Murtha GG (1989) Soils of the Mossman Cape Tribulation area, North Queensland. The National Committee on Soil and Terrain. (CSIRO Publishing: Melbourne)
Neldner VJ, Wilson BA, Dillowski HA, Ryan TS, Butler DW, McDonald WJF, Addicott EP, Appelman, CN (2019) Methodology for surveying and mapping regional ecosystems and vegetation communities in Queensland. V5.0. Queensland Herbarium, Department of Environment and Science, Brisbane.

Northcote KH (1979) ‘A factual key to the recognition of Australian soils.’ (Relfilm Technical Publications Pty Ltd.: Coffs Harbour, NSW).

Odling NP, Sun W, Minasny B, Clifford D (2014) Disaggregating and harmonising soil map units through resampled classification trees. Geoderma 214–215, 91–100. doi:10.1016/j.geoderma.2013.09.024

Pomerening JA, Cline MG (1953) The accuracy of soil maps prepared by various methods that use aerial photograph interpretation. Photogrammetric Engineering 19, 809–817.

Powell B (1979) Soils of the Kalbar District, south-east Queensland. Queensland Department of Primary Industries, Brisbane.

Powell B (2008) 19. Classifying soil and land. In ‘Guidelines for surveying soil and land resources. 2nd Edition’. (Eds NJ McKenzie, MJ Grundy, R Webster, AJ Ringrose-Voase.) pp. 307–315. (CSIRO Publishing: Melbourne.)

Rakedzon T, Segev E, Chapnik N, Yosef R, Baram-Tsabari A (2017) Automatic jargon identifier for scientists engaging with the public and science communication educators. PLoS One 12(8), e0181742. doi:10.1371/journal.pone.0181742

Reid RE (1988) 5. Soil survey specifications. In ‘Australian soil and land survey handbook - guidelines for conducting surveys’. (Eds RH Gunn, JA Beattie, RE Reid, RHM van de Graaff.) pp. 60–72. (Inkata Press: Melbourne).

Rosenfeld GH (1986) Analysis of thematic map classification error matrices. Photogrammetric Engineering and Remote Sensing 52(5), 681–686.

Rosenfeld GH, Fitzpatrick-Lins K (1986) A coefficient of agreement as a measure of thematic classification accuracy. Photogrammetric Engineering and Remote Sensing 52(2), 223–227.

Ross DJ, Crane AJ (1994) Land resource assessment of the Goodar area, Queensland. Queensland Department of Primary Industries. QV94003.

Schoknecht N, Wilson PR, Heiner I (2008) 14. Survey specification and planning. In ‘Guidelines for surveying soil and land resources. Second Edition’. (Eds NJ McKenzie, MJ Grundy, R Webster, AJ Ringrose-Voase.) pp. 205–223. (CSIRO Publishing: Melbourne).

Shields P, Williams BM (1991) Land resource survey and evaluation of the Kilcummin area, Queensland. Queensland Department of Primary Industries. QV91001.

Smith D, Calland D (2020) Land Resource Assessment of the Upper Logan and Albert Rivers Catchment, South-East Queensland. Queensland Department of Natural Resources, Mines & Energy.

Smith D, Crawford M (2015) Land Resource Assessment of the Lake Coorarabu Catchment. Queensland Department of Natural Resources and Mines.

Stehman SV, Foody GM (2019) Key issues in rigorous accuracy assessment of land cover products. Remote Sensing of Environment 231, 111199. doi:10.1016/j.rse.2019.05.018

Story M, Congalton RG (1986) Accuracy assessment: a user’s perspective. Photogrammetric Engineering and Remote Sensing 52(3), 397–399.

Thomas M, Brough D, Bui E, Harms B, Hill J, Holmes K, Morrison D, Philip S, Searle R, Smolinski H, Tuomi S, Van Gool D, Watson, I, Wilson, PL, Wilson, PR (2018) Digital soil mapping of the Fitzroy, Darwin and Mitchell catchments. CSIRO.

Vincent S, Lemercier B, Berthier L, Walter C (2018) Spatial disaggregation of complex Soil Map Units at the regional scale based on soil-landscape relationships. Geoderma 311, 130–142. doi:10.1016/j.geoderma.2016.06.006

Wilding LP, Drees LR (1983) Chapter 4 Spatial Variability and Pedology. In ‘Developments in Soil Science. Volume 11, Part A.’ (Eds: LP Wilding, NE Smeeck, GF Hall) pp. 83–116. (Elsevier)

Wilson PR (1997) Soils and agricultural suitability of the Childers area, Queensland. Department of Natural Resources. DNROQ7158.

Wilson, PR, Baker, DE (1990) Soils and agricultural land suitability of the wet tropical coast of North Queensland: Ingham area. Queensland Department of Primary Industries Number QV90001.

Yang L, Li X, Shi J, Shen F, Qi F, Gao B, Chen Z, Zhu AX, Zhou C (2020) Evaluation of conditioned Latin hypercube sampling for soil mapping based on a machine learning method. Geoderma 369, 114337. doi:10.1016/j.geoderma.2020.114337

Zund PR (2014) Disaggregation of land systems mapping - Fitzroy Drainage Basin. Department of Science, Information Technology, Innovation and the Arts, Brisbane.

Handling Editor: Stephen Cattle