Digital soil mapping approaches to address real world problems in southern Africa

George van Zijl

University of the Free State, Nelson Mandela Avenue, 9301 Bloemfontein, South Africa

ARTICLE INFO

Handling Editor: A.B. McBratney
Keywords: Commercial use	Cost effective	Environmental Impact Assessments	Expert knowledge	Land type disaggregation	Machine learning

ABSTRACT

Soil spatial information is increasingly sought after for various agricultural, environmental and developmental uses, but is often unavailable, also in southern Africa. Digital soil mapping (DSM) can provide the tools to fill this gap, but the uptake in developing countries has been slow. Local research is required to adapt internationally developed methodologies to unique local situations. In southern Africa DSM research have reached the level where DSM tools can now be used in commercial soil related projects. Several DSM case studies, conducted across Southern Africa, have provided the platform from which this work is presented. These case studies were done for a range of situations, including environmental settings with variations, size, data availability and aim of the soil map. Three different approaches have been identified as useful DSM tools, with varying costs and level of information it provides. Land type disaggregation is the cheapest, as it is largely desktop based, but can only produce small scale soil association maps. The expert knowledge approach is the most widely used commercially. Large scale soil associations can be mapped, and 30 soil observations per homogeneous soil distribution area are required. Machine learning methods can map soil properties, but rely on large data sources, consequently it is the costliest. Machine learning is therefore used to produce large scale maps large areas, where cost can be diluted. This paper gives an outline of DSM research in southern Africa and presents a case study of each of the DSM approaches, showing the methodology, potential and limitations of the approach within a commercial context. Specific case studies presented in this paper include the agricultural assessment of 166 km of pipeline for a water distribution project in Limpopo (land type disaggregation), a land capability assessment of a 15,000 ha open coal mining area in Mozambique (expert knowledge) and hydropedological mapping in Johannesburg (machine learning).

1. Introduction

There is an increased demand for soil information, as the importance of soil in eco-systems is more widely recognized. In agriculture, it is estimated that 95% of food produced worldwide is dependent on the soil (FAOSTAT, 2003). In hydrology, the soil-water interaction has given rise to the exciting sub-discipline of hydropedology (Lin, 2003; Van Tol et al., 2013). In ecology, the important contribution of soil to ecosystem services are increasingly recognized (Dominati et al., 2014; Frank et al., 2014; Baveye et al., 2016; Bouma and Montanarella, 2016). One of the ecosystem services which soils provide, is its role in mitigation of climate change (Brevik, 2012; Wiesmeier et al., 2016), as soils contain the largest pool of active terrestrial carbon (Lal, 2010). Bouma (2014) mentions that soils directly influence five of the 10 sustainable development goals of the UN’s Sustainable Development Solutions Network (UN-SDSN, 2013). However, in another paper (Bouma and McBratney, 2013) he argues that soil science does not feature in high profile reports related to these goals, such as a CGIAR report on food security (Beddington et al., 2011) and the UNEP (2011) report on the need for a green economy. The main reason mentioned for this omission is that the authors of these reports do not have access to assessments of soil conditions, or bluntly put, soil scientists have not been able to provide the soils data necessary.

The situation is roughly the same in southern Africa. Due to the importance of soil being increasingly prioritized, soil information is increasingly sought after, and even required by law. Environmental Impact Assessments (EIA’s), which generally includes soil investigations, are required for developments in South Africa (RSA, 1998). The relatively new Spatial Planning and Land Use Management Act (SPLUMA; RSA, 2013) requires wall to wall zoning of municipal areas, which often requires soil surveys. Furthermore, high-value agricultural land will be protected from development for agricultural production when the draft Preservation and Development of Agricultural Land Framework Bill (PDALFA; DAFF, 2013) is passed, but a soil map will be required to know where this high-value agricultural land is situated. Outside of South Africa, funding agencies such as the International...
Funding Corporation insist on EIAs for any development to be funded, even if not required by law in the specific country. Large agricultural developments, such as the Farm Block developments in Zambia require soil maps for hundreds of thousands of hectares to finalize their feasibility studies. Often these requirements cannot be met with the current conventional grid surveys, as the areas are too large, or the developments requiring the soil surveys are of a linear nature, such as overhead power cables or water pipes, stretching over vast distances and many soil forms. What is different in southern Africa is that the governments do very little in terms of soil surveys, which opens the opportunity for soil scientists to conduct commercial soil surveys.

Commercial soil surveys are unique in that they need to be done with a competitive budget within a strict deadline, with the products directly affecting investment often times orders of magnitude costlier than the cost of the soil map. The sites are often also located within areas with very little soil information available, necessitating a fine balance between data acquisition (which drives up cost and time required) and level of detail and accuracy required in the final product. This is in stark contrast to high profile digital soil mapping (DSM) projects, such as the soil map of Africa (Hengl et al., 2015) and the soil map of Australia (Viscarra Rossel et al., 2015), which relies on legacy soil data, could be updated from time to time, shows regional trends and is not site specific or for a specific use.

Digital soil mapping provides a timely and cost-effective way of producing the soil information required. McBratney et al. (2003) observed in 2003 that digital soil mapping is moving from the research to the production phase. Globally there has been a drive to fulfill the need for soil information, with projects such as the Global Digital Soil Map (Sanchez et al., 2009) and national and continental soil maps which has been created for various soil parameters (Viscarra Rossel et al., 2015; Hengl et al., 2015). Due to soil-landscape interactions which vary at different geographic locations, methodologies used in regional soil mapping might not be applicable at a different scale or location (Minasny and McBratney, 2010), which necessitates the need for local DSM research, before it can be used productively to address real world issues. However, as evidenced in the attendance register of the Pedometrics Conference held in 2017 in Wageningen, the Netherlands, the uptake of the tools provided by DSM has been slow in developing areas such as Africa. This in turn is partly due to the little local research on DSM being done in these areas.

In southern Africa, DSM research and uptake thereof have been sparse and quite slow. At the South African Agricultural Research Council - Institute for Soil, Climate and Water (ARC-ISCW) Schoeman (2005) compiled a review on the theory of pedometrics, while his colleagues used the land type point database to map soils of the Kwa Zulu Natal province (Van den Bergh and Weepener, 2008; Van den Bergh et al., 2009). Wiese et al. (2015) accounted for soil carbon using vertical distribution functions and kriging. At Stellenbosch University, Stalz (2007) mapped salt affected soils in the Orange River irrigation scheme by mapping salt affected plants with remote sensing. Mashimbye et al. (2012) also mapped salinity using hyper spectral remote sensing data. Atkinson et al. (2017) works on pre-processing of the attribute data, with specific focus on DEM quality. Outside of South Africa, Chabala mapped the Zambian soil acidity (Chabala et al., 2014) and soil organic carbon (Chabala et al., 2017). In Mozambique Cambule (2013) assessed soil organic carbon stocks (Cambule et al., 2014), while creating a DSM methodology for poorly accessible areas (Cambule et al., 2013) which included rescuing legacy data (Cambule et al., 2015) and building a spectral library (Cambule et al., 2012).

Apart from these studies, a concerted effort to create commercial DSM projects has been conducted by a research group based at the University of the Free State (UFS) and related institutions over the past ten years (see for example, Hensley et al., 2007; Van Zijl et al., 2014; Van Tol et al., 2014; Botha, 2016), which has led to the use of DSM in commercial projects (e.g. Le Roux et al., 2013; Van Zijl, 2015; Van Zijl and Du Plessis, 2016). This paper discusses the major lessons learnt during the past ten years of research, and highlights its commercial application through case studies.

2. Application of DSM in general

To get a broad view of how DSM tools are applied, the DSM related projects done at the UFS and related institutions were analysed for certain trends. Fig. 1 shows the locations of these DSM projects, while a summary of the projects where the relevant information is available is given in Table 1.

The versatility of DSM is shown through the widespread geographical locations where the projects have been conducted (Fig. 1). The projects are spread over seven provinces in South Africa, and as far as Mozambique. This spread ensures that the methodology has been tested in various environmental settings, with differing climate zones, parent materials, vegetation types and topographies. Table 1 shows that DSM methodology has been used to create soil maps for a variety of uses, including EIAs, agriculture, and hydopedological assessments. The successful application of DSM methodology for commercial soil surveys is shown by the fact that 60% of the projects listed in Table 1 were done as consultancy projects. These projects are generally on fairly large areas of land (4000–20,000 ha; 45–166 km for linear features), with little legacy soil data available necessitating the collection of soil data as part of the project. Tough terrain often makes the gathering of soil data difficult. Therefore, minimizing the need for soil data is a high priority. As such, sparse observation densities (74–216 ha·obs 1) are observed in the projects. The advantage of DSM lies in the fact that it can extract the soil-environment relationship from relatively few soil observations, which enables the soil to be mapped using continuous environmental covariates. DSM produced soil maps has an accuracy assessment included, which allows the land user to use the map with a known confidence.

There are three typical approaches to commercial DSM in southern Africa, which have all been applied repeatedly. The land type disaggregation approach and expert knowledge approach is similar, in the way that both rely on soil-environment rules to be created for each soil map unit. This idea was first developed by the research group of A-Xing Zhu (for example: Zhu, 1997; Zhu et al., 2001; Qi et al., 2006; Yang et al., 2011). The approaches differ, since the land type disaggregation approach relies on soil-environment information contained in a land type inventory, while the expert knowledge approach uses the soil surveyor's knowledge on soil distribution, as well as soil observations as the basis to create the rules. Machine learning methods rely on large databases which are used to build statistical soil-environment relationships. These relationships are applied throughout the study area to create the soil map.

3. Specific DSM approaches

3.1. Land type disaggregation approach

The only source of soil information which covers the whole of South Africa is the Land Type survey (Land Type Survey Staff, 1972–2006). A land type mapping unit is defined as “a homogeneous, unique combination of terrain type, soil pattern and macro climate zone” (Paterson et al., 2015). Each land type mapping unit is accompanied by an inventory, which shows an estimated percentage which each soil form comprises on each terrain unit. Terrain units are defined on a simple, topo-sequence sketch. Therefore a relationship between soil and terrain is contained in the inventory. With some knowledge of the specific soil distribution patterns within the land type unit and some terrain analysis, this information could be extracted to produce a soil map. Botha (2016) investigated a method of land type disaggregation whereby the dominant soil form on each terrain unit is assigned to the terrain unit, which is determined with terrain analysis of a DEM. In complex land types, where there is no clear dominant soil form on a terrain unit, the
area is divided into hillslopes, and the average and standard deviation of the slope and profile-and planform curvature are determined. Based on the hillslope shape, a soil form is assigned to each terrain unit of the specific hillslope, based on a soil surveyor's knowledge of soil distribution within the land type. Botha (2016) concluded that the method worked well in simple landscapes, but in more complex landscapes failed to give adequate results. She further mentioned that even in simple land types expert soil distribution knowledge proved

Table 1
Summary of projects done with DSM methodology.

<table>
<thead>
<tr>
<th>Project</th>
<th>Application</th>
<th>Approach</th>
<th>Area</th>
<th>Training Obs</th>
<th>Obs density ha·obs⁻¹ km·obs⁻¹</th>
<th>Class type</th>
<th>Nr of classes</th>
<th>Map accuracy</th>
<th>Map acceptable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tete</td>
<td>EIA</td>
<td>EK</td>
<td>15,000 ha</td>
<td>131</td>
<td>115</td>
<td>Stockpiling groups</td>
<td>4</td>
<td>69</td>
<td>Yes</td>
</tr>
<tr>
<td>Madadeni</td>
<td>Research</td>
<td>LTD</td>
<td>6865 ha</td>
<td>0</td>
<td>N/A</td>
<td>Soil associations</td>
<td>9</td>
<td>36</td>
<td>No</td>
</tr>
<tr>
<td>Namarrua</td>
<td>Agriculture</td>
<td>EK</td>
<td>20,000 ha</td>
<td>151</td>
<td>132</td>
<td>Forestry production</td>
<td>10</td>
<td>80</td>
<td>Yes</td>
</tr>
<tr>
<td>Skukuza</td>
<td>HP</td>
<td>EK</td>
<td>4001 ha</td>
<td>54</td>
<td>74</td>
<td>Hydropedological response</td>
<td>7</td>
<td>73</td>
<td>Yes</td>
</tr>
<tr>
<td>CP</td>
<td>Research</td>
<td>LTD</td>
<td>9500 ha</td>
<td>0</td>
<td>N/A</td>
<td>Soil associations</td>
<td>3</td>
<td>89</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EK</td>
<td>248 ha</td>
<td>37</td>
<td>7</td>
<td>Soil associations</td>
<td>6</td>
<td>68</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ML</td>
<td>248 ha</td>
<td>37</td>
<td>7</td>
<td>Soil associations</td>
<td>6</td>
<td>64</td>
<td>Yes</td>
</tr>
<tr>
<td>Metsimatala</td>
<td>EIA</td>
<td>LTD</td>
<td>45 km</td>
<td>0</td>
<td>N/A</td>
<td>Agricultural production</td>
<td>4</td>
<td>N/A</td>
<td>Yes</td>
</tr>
<tr>
<td>Limpopo</td>
<td>EIA</td>
<td>LTD</td>
<td>166 km</td>
<td>0</td>
<td>N/A</td>
<td>Agricultural production</td>
<td>4</td>
<td>75</td>
<td>Yes</td>
</tr>
<tr>
<td>Ntabelanga</td>
<td>Research</td>
<td>LTD</td>
<td>13,796 ha</td>
<td>0</td>
<td>N/A</td>
<td>Soil erodibility classes</td>
<td>5</td>
<td>37</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EK</td>
<td>13,796 ha</td>
<td>64</td>
<td>216</td>
<td>Soil erodibility classes</td>
<td>6</td>
<td>65</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ML</td>
<td>13,796 ha</td>
<td>64</td>
<td>216</td>
<td>Soil erodibility classes</td>
<td>6</td>
<td>65</td>
<td>Yes</td>
</tr>
<tr>
<td>Secunda</td>
<td>EIA</td>
<td>EK</td>
<td>127 km</td>
<td>27</td>
<td>5</td>
<td>Agricultural production</td>
<td>2</td>
<td>80</td>
<td>Yes</td>
</tr>
<tr>
<td>CoJ</td>
<td>HP</td>
<td>ML</td>
<td>12,538 ha</td>
<td>96</td>
<td>130</td>
<td>Hydropedological response</td>
<td>5</td>
<td>70</td>
<td>Yes</td>
</tr>
</tbody>
</table>

CoJ = City of Joburg.
CP = Cathedral Peak.
HP = Hydropedology.
EK = Expert knowledge.
LTD = Land type disaggregation.
ML = Machine learning.
* Consultancy projects.
invaluable, and could either be obtained from the original land type surveyor, someone who conducted a soil survey within the land type, or making one’s own field observations. The method is similar to that of Thompson et al. (2010), as the soil environment rules are derived manually by extracting expert knowledge from the map to be disaggregated, but it lacks the automation of other disaggregation methods (e.g. Bui and Moran, 2001; Odgers et al., 2014; Vincent et al., 2018), where decision trees are created with computer algorithms. It is thus a time-consuming process.

Land type disaggregation has been used commercially for soil surveys for EIA’s of linear structures such as water pipe lines or power lines. The case study for this approach is the soil part of an EIA for 166 km of water pipelines for the “Upgrade of the Olifantspoort & Ebenezer Water Schemes” in the Limpopo Province, subcontracting for MSW Consulting on behalf of Lepelle Northern Water.

3.1.1. Upgrade of the Olifantspoort & Ebenezer Water Schemes

A soil map was required for a 166 km linear feature of a water pipeline in the Limpopo Province. Agricultural potential and soil erodibility were to be derived from the soil map. A land type disaggregation approach was followed, due to the linear extent of the site. The pipeline covered 21 different land types. To obtain site specific soil information 50 soil observations were predetermined using the conditioned Latin hypercube sampling method of Minasny and McBratney (2006), with altitude, profile curvature, planform curvature, slope gradient, topographical wetness index and multiresolution index of valley bottom flatness as covariates. Auger soil observations to 1.2 m or refusal were made at the locations determined by the sampling method, as well as at 11 other sites based on the soil surveyor’s discretion. Thus 61 soil observations were made. Soils were described and classified according to the South African soil classification system (Soil Classification Working Group, 1991). After the field work, the soil forms were grouped into soil associations based on their agricultural potential and erodibility (Table 2).

The soil map (Fig. 2) was created by combining the soil information contained in the land type inventories with the soil-environment relationship learnt while doing field work, upholding the principle of Botha (2016), that in-field knowledge is a necessary addition to the land type inventory. Broad land types were used to divide the area into homogeneous soil-environment relationship areas. The land type inventories as well as the soil dataset were consulted to determine which soil associations were present in each broad land type. Soil-environment rules were quantified in Solim Solutions software (Zhu, 1997) for each soil association, for each broad land type. The covariates used were slope, terrain wetness index and altitude above channel network. The soil observations and the covariate values at their locations were not used to create the soil-environment rules. Applying the rules to the entire area (also in Solim Solutions) created the soil map. Point validation was done using 60 of the soil observations (one observation was clearly disturbed by human impact and thus omitted). The point accuracy was an acceptable 75%, with a Kappa statistic value of 0.603, indicating that the map is in substantial agreement with reality at the validation points. Agricultural potential and soil erodibility values could be added to each soil mapping unit, creating the required maps for the EIA.

The advantage of the land type disaggregation approach is that fewer observations were necessary than with the other approaches, as the knowledge contained in the land type inventory could be used to supplement in field observations. The observation density of this project was 2.7 km per observation or three observations per land type. The expert knowledge approach would have required 30 observations per land type (Van Zijl, 2013), a 10-fold increase. A machine learning approach would have required even more. If working on a very coarse 500 m interval, 334 observations would have been necessary with the conventional grid survey approach. Thus, disaggregating the land types produced an acceptable soil map, while requiring fewer observations.

On the downside, only four mapping units were mapped, and although sufficient for this project, this approach could not be used for studies where more information is required, such as complex land types or hydromorphological studies.

3.2. Expert knowledge approach

Bui (2004) argued that “soil survey is a ‘knowledge system’, and soil maps are a representation of structured knowledge about the distribution of soils in the landscape.” In this paper, such knowledge held by soil surveyors is defined as expert knowledge. Thus, in the expert knowledge approach, soil observations are made with the aim of understanding the soil-environment relationships. Both the soil surveyor’s expert knowledge and the observation points are used to quantify the soil environment relationships as rules, which are applied across the mapping area to produce the soil map. Van Zijl (2013) found that one needs at least 30 soil observations per area with a homogeneous soil distribution pattern to map up to six soil map units per such area. Additionally, he also concluded that for areas larger than 1000 ha it would be advantageous to use the expert knowledge approach instead of the conventional grid survey method. The approach is used the most for DSM consultation purposes in southern Africa, as the maps produced are useful soil associations, based on soil morphology. The approach has however been used to map soils in more detail, for example to USDA soil series level (Zhu et al., 2001). Grouping of soil forms into the soil associations varies according to the aim of the map. The expert knowledge approach works best in areas with a strong soil-terrain relationship, as this relationship is best understood by the soil surveyor.

In areas with a weak soil terrain relationship, such as aeolian and fluvial landscapes, the success of the method will be limited. The case study of this approach is the Tete project, where a soil map for 15,000 ha was produced within 16 man days, in an area potentially infested with land mines in Mozambique. This study is fully published in Van Zijl et al. (2012).

3.2.1. Tete

A soil map for 37,000 ha was required as part of an EIA for an open cast mine near Tete, Mozambique. The aim was to evaluate the agricultural potential of the area so as to determine the rehabilitation after mining ceased, and to create stockpiling guidelines to protect the topsoil from decay while mining. To complicate the matter there was a time deadline of eight days, an uncrossable river running through the area and restrictions on where soil observations could be made, due to the land mine threat. Within the accessible area seven homogeneous

---

Table 2

<table>
<thead>
<tr>
<th>Soil association</th>
<th>Grazing potential</th>
<th>Dryland cultivation potential</th>
<th>Irrigation cultivation potential</th>
<th>Soil erosion potential</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shallow</td>
<td>Low</td>
<td>Very low</td>
<td>Very low</td>
<td>Low</td>
</tr>
<tr>
<td>Wet</td>
<td>Medium</td>
<td>Very low</td>
<td>Very low</td>
<td>High</td>
</tr>
<tr>
<td>Structured</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Apedal</td>
<td>High</td>
<td>Medium</td>
<td>Medium to very high</td>
<td>Medium</td>
</tr>
</tbody>
</table>
areas were identified where the geology and terrain pattern were similar. These homogeneous areas comprised 15,000 ha of the 37,000 ha and stretched into the inaccessible part of the site. This was also the final area to be mapped (Fig. 3). One hundred and thirty-one soil observations were made by soil auger to a depth of 1.5 m or refusal across these seven homogeneous areas. Soils were classified according to the Soil Classification Working Group (1991), and converted to WRB reference soil groups (IUSS Working Group, 2007) as the project occurred outside of South Africa. Soil-environment rules were derived from the soil surveyors’ expert knowledge and 131 training soil observations, for each of the mapping units, within each of the homogeneous areas. Co-variates used included slope and profile curvature. The final map was created in Solim Solutions (Zhu, 1997) based on these rules. Validation of the soil maps was done with a new independent set of 52 soil observations, made at the soil surveyors discretion in the accessible area. The points were mapped with a 69% accuracy, which had a Kappa statistic value of 0.55, indicating a moderate agreement with reality. Agricultural potential could be assigned to each soil map unit, and areas which should be stockpiled separately were indicated.

This case study was unique due to the near impossibility of creating

Fig. 2. The final soil map for the Limpopo water scheme. The insert shows the detail of the map.

Fig. 3. The soil map for the Tete project (from Van Zijl et al., 2012).
a soil map under such challenging circumstances and exhibits the immense potential of digital soil mapping. The observation density of one observation per 115 ha is much lower than conventional grid surveys would require (a very coarse 500 m grid has an observation density of one observation per 25 ha), but the main advantage which was showcased is the extrapolation potential of DSM produced maps. A map could be produced for areas which could not be surveyed, even though no direct measure of the map accuracy for the unsurveyed areas could be made.

3.3. Machine learning approach

With the machine learning approach, each soil observation is a value which will be added to an algorithm, with the aim of statistically deciphering the soil environment relationship. As with all statistics, there is a need for a large enough data set before this approach will yield acceptable results. Generally, compiling a large dataset is costly, so this approach is used for exceptionally large areas, in order to dilute the cost per hectare, or where very detailed soil information is required. The advantages are that human bias can largely be negated and desktop work is relatively little. It is also the only of the three approaches with which continuous soil properties are generally mapped. The case study that will be discussed is the City of Johannesburg hydropedological soil map.

3.3.1. City of Johannesburg

The challenge with this project was to create a hydropedological soil map which could assist the city officials when making developmental decisions. In Johannesburg, there have been several issues with developments causing damage to wetlands further downhill, as they cut off the hillslope water supply, and of infrastructure requiring increased maintenance due to it being built in soil water flow paths, which causes accelerated degradation. Hydropedology is the study of water movement within the soil, and generally uses the hillslope as landscape unit upon which assessments are based (Van Tol et al., 2013). The aim of the project was to create a soil map whereby the city officials could deduce the water flow paths within the soils from, allowing them to make improved developmental decisions. Thus, the final product is an interpreted hydropedological map, rather than a standard soil class map.

Firstly, the area was defined as the largely unbuilt area by visual inspection of satellite imagery. The study site was divided into hillslopes with the method described by Le Roux et al. (2014). Hillslopes are defined by first determining the channel network of the site, then delineating the subbasins around the channel network. Lastly the subbasins are divided into hillslopes, by “cutting” them with the channel network layer. Thirty hillslopes were selected with the cLHS method of Minasny and McBratney (2006). Soil observations were made on transects at the discretion of the soil surveyor on the selected hillslopes. One hundred and forty-two soil observations were made by soil auger to refusal or 3.2 m. Soils were described and classified into 17 soils forms according to the Soil Classification Working Group (1991). From the soil descriptions, observations were divided into five conceptual hydrological response classes based on Le Roux et al., 2011. Generally, there are three conceptual hydropedological response classes: Recharge, responsive and interflow. With the recharge soils water is expected to move vertically through the soil profile and recharge the ground water. The responsive soils are soils where water will flow on top of the soil surface, either due to the soil being very shallow, or saturated with water. A flow hydrograph on these soils would respond very quickly. In the interflow soils water flows laterally in the soil.
profile. The interflow hydrological class was divided into three units, shallow, deep and multi, based on whether the soil morphology indicated a water flow path shallower than 500 mm, deeper than 500 mm or both shallower and deeper than 500 mm, respectively. Ninety-six of the observations were used as training data to create the map (Fig. 4) with the multinomial logistic regression function in R. A hydropedological soil map and a soil form map were created. Validation was conducted with the remaining 48 observations, with the one pixel buffer method as proposed by Van Zijl et al. (2012). Soil forms with less than three observations in the original dataset (before dividing training and validation data sets) were omitted when creating the soil form map. The hydropedological map obtained a point accuracy of 69% and a Kappa statistic value of 0.59, indicating a moderate agreement with reality at the validation points. The soil form map only achieved a point accuracy of 24% and a Kappa statistic value of 0.17, indicating no agreement with reality. Thus, it was deemed as not being useful. The reason for this is that there were not enough observations per soil form to statistically determine the soil-environment relationship.

Without DSM, it would be impossible to create a hydropedological soil map for such a large area, due to the cost constraint thereof. Observations for hydropedological assessments are required to bedrock, and thus the time cost of each observation is potentially much larger than for other types of soil surveys, increasing the cost thereof. The extrapolation of hillslope responses to unsurveyed hillslopes allowed for the map to be created with acceptable cost.

4. Conclusions

Digital soil mapping methods have been used commercially to provide soil information in large areas (4000–20,000 ha). This have been done within competing budgets and time frames, by using the environmental setting to predict the soil distribution, necessitating less soil observations and thus time spent. Three different approaches have been used to obtain the soil information. The choice of approach is dependent on the soil information needed for the project.

Land type disaggregation is the cheapest method, since it is largely a desktop study, whereby the soil-environment relationship is extracted from the land type inventory with additional field work. These are some constraints due to the small scale and restricted depth of the land type survey observations. Hydropedological assessment methods should be used for this purpose. Land type disaggregation has been used successfully by EIA's of linear features and baseline soil maps for new projects.

The expert knowledge approach is the most used commercially, as it produces useful soil association maps at a reasonable cost. A moderate amount of field and desktop work is required. Thirty soil observations are needed and six soil associations can be mapped per homogeneous area. Crop production potential, soil erosion risk, EIA soil maps and landform assessment maps have been created with this approach.

Due to the large requirement for data, the machine learning approach is only used to map very large areas at a large scale, in order to “dilute” the cost per hectare. The machine learning work required is substantial, but the desktop work can be automated. This approach can map soil properties provided enough samples have been taken. It has been used in town planning, land degradation projects and precision agriculture.

Future work should focus on DSM training, in order to broaden the base of DSM skills available in southern Africa. This will ensure that soil information would be included in addressing an increasing amount of real world problems. In South Africa, DSM training is currently confined to post-graduate studies at a few institutions, including the University of the Free State, Fort Hare University and Stellenbosch University. Beyond South Africa DSM research is also conducted at the University of Zambia.

Acknowledgements

I would like to thank the following researchers who all contributed to the different studies mentioned in this paper: Malcolm Hensley, Pieter Le Roux, Johan van Tol, Darren Bouwer, Hendrik Smith, David Turner, Cobus Botha, Christina Botha, Martin Tinnefeld, Nancy Job and Cassie du Plessis. Darren Bouwer proofread the manuscript. Also, I would like to thank MSW Consulting Engineers, Lepelle Northern Water and the City of Johannesburg for allowing data to be shared in this publication. Special thanks should go to the reviewers, who's comments increased the quality of this publication.

References


Land Type Survey Staff. 1972–2006. Land types of South Africa: Digital map (1:250 000 scale) and soil inventory datasets. ARC-Institute for Soil, Climate and Water, Pretoria.


