[ISBN: 978-2-9105-4510-5]

Exploring the uncertainty of soil water holding capacity information

Linda Lilburne*, Stephen McNeill, Tom Cuthill, Pierre Roudier

Landcare Research, Lincoln, New Zealand

*Corresponding author: lilburnel@landcareresearch.co.nz

Abstract

Soil water holding capacity is an important soil property for understanding how much irrigation water is required and the quantity of nutrients that are likely to leach into groundwater. This soil profile level property is derived from horizon level data including soil water content, stone content, thickness and horizon type. Soil water content data is expensive to measure so is often estimated in a model that is based on more readily collected soil information. A model of the soil water content (or hydraulic response) has been developed and tested. Its inputs include sand and clay content, profile and horizon classifications. Thus, uncertainties in the derived estimate of soil water holding capacity are due to variability in the inputs to the soil hydraulic model, error in the model itself, and variability of the other key soil properties (stone content, thickness). The combined uncertainty is estimated in a new Soil Profile Simulator tool. It is based on a simulation approach that draws upon the statistical error model of the soil hydraulic model and expert information held in a national soil survey database in the form of probability distributions. This expert information characterises the variability of clay, sand, horizon thickness, stone content, and uncertainty in the classification information associated with the soil polygons. This paper describes the tool, reporting on the progress that has been made in deriving and visualising quantified estimates of uncertainty of soil water holding capacity. Recent advances in technology, including new R packages (aqp, VGAM) and Rserve have been behind this progress.

Introduction

Information on soil hydraulic properties is essential for the sustainable management of irrigated agricultural land. Providing too much water is wasteful and contributes to contamination of water through leaching and runoff of nutrients. Too little water will impact negatively on yield. One of the key soil properties for managers of irrigated land is the soil's water holding capacity or profile available water (PAW), which represents the amount of water held in the soil that is readily accessible to plants. In New Zealand, this is taken to be the water content held between soil moisture tensions of 10 and 1500 kPa. Although this can be measured directly, it is an expensive process, so PAW is commonly predicted using other soil properties that are easier to measure or observe. These include texture, carbon, bulk density and soil morphology. A relationship between soil properties is known as a pedotransfer function (PTF).

The soil hydraulic response has a number of features that make model predictions difficult. First, the response is bounded (0–100%), so either a bounded-value regression (e.g. logistic) is required or a transformation is needed to an unbounded domain. Second, the response is monotonic with respect to the tension: that is, the response strictly decreases as the tension increases. Finally, the response at a given tension is correlated with the response at other tensions. This complexity also adds to the challenge of estimating the error of the predicted response.

The soil properties used as inputs by the PTF are also uncertain. This paper presents progress with estimating and communicating uncertainty of PAW, whereby the error model of the soil hydraulic PTF is combined with probabilistic information on variability of the key soil inputs of the PTF held in a soil survey database. A new visualisation R package 'aqp' is used to visualise the results.

Methods

Soil hydraulic error model

In forming an empirical model for soil hydraulic response, we use soil sample data available from the New Zealand National Soils Database (NSD), which provides the response at tensions of 0 (total porosity), 5, 10, 20, 40, 100, and 1500 kPa. For each sample, texture (sand, silt, clay fractions) data are available, as well as the soil classification, and other factors describing the soil sample.

Our methodology for response prediction uses a vector generalised linear model (VGLM). This, according to Yee (2015), can be thought of as a generalisation of the generalised linear model (GLM) with a vector of responses, which is free from many of the restrictions that the GLM method imposes. We use a vector of responses formed from the logit-transformed 1500 kPa response (transformed to give an unbounded range), as well as the difference between the logit-transformed 100 and 1500 kPa responses, the difference between the logit-transformed 40 and 100 kPa responses, and so on, up to the difference between the logit-transformed 5 and 0 kPa response. The response at one of the specified tensions is formed from the VGLM prediction for 1500 kPa, plus a succession of differences for lower tensions. The uncertainty of each marginal response (at 1500 kPa or a difference in response between tensions) is formed from the estimated VGLM model.

An error model for the difference between the 10 and 1500 kPa responses (i.e. estimates of total available water within a soil horizon or layer), or indeed for any other convenient combination of responses, has been verified using independent validation data (McNeill et al. in prep.). Uncertainty limits, calculated in terms of containment intervals, are estimated by simulation of the aggregated response.

Soil Profile Generator tool

Lilburne et al. (2012) described how information on soil variability and uncertainty was being incorporated in a national-scale soil database for New Zealand called S-map (Landcare Research, 2015). The very limited amount of soil sample data meant that an expert knowledge approach was used to record information on the confidence of classification and base property attributes, the variation of soils in a polygon and their proportional reliability, and the range of values of key soil properties (Lilburne et al. 2008). Variability of quantitative soil properties is stored in the form of probability distribution functions (pdf). The lack of point soil sample data precluded a geostatistical approach to modelling and simulating PAW.

A new Profile Simulator tool has been developed that creates realisations of profiles based on the expert-derived information on soil variability and uncertainty. The key information used in this study on PAW is the PTF error model; expected variability of the stone, sand and clay content; type and thickness of functional horizons within a soil survey polygon; confidence in the Soil Order classsification, rock type of the fine material and drainage class, as well the reliability of the proportions of soil types within a polygon. Values for horizon stone, sand, and clay content, and thickness were drawn from their respective pdfs. Each profile realisation is checked against a set of rules to ensure that it is still consistent with the soil definition. For example, the sum of the simulated horizon thicknesses must fit the pdf of the soil's depth. Realisations that do not fit the rules are discarded and regenerated. The PTF error for each soil profile realisation was simulated and the PAW calculated.

The tool is based on an architecture that links the S-map database in SQL Server with RServe – a server that responds to requests from clients by running a R script. Rserve enables R calculations to be performed on request without the need to start up an R session each time. A number of servlets have been developed that retrieve the stored uncertainty information from SQL Server, call a set of Monte Carlo functions to generate realisations of the key parameters required by the soil hydraulics error model, run this model on each set of parameter realisations, simulate model error, and finally, return the set of profile realisations with estimates of available water, in a range of useful formats (database tables, csv file, Rdata file). RServe was installed on a Linux-based server to allow for multi-threaded processing.

Visualisation

The 'aqp' package (Beaudette et al. 2013) is an R package designed for working with soil information. It contains a SoilProfileCollection class to simplify the process of working with the collection of data associated with soil profiles. It also includes tools for plotting soil properties by depth and their associated variability.

Results

Using a training dataset of 1007 points, a VLGM model was developed and tested on an independent dataset of 432 points (Figure 1). Histograms of the residuals for seven tension values, and the associated model diagnostics using training and validation data, showed similar patterns, indicating that the model was not overfitted. Simulated error for the 1500 kPa tension value and total available water for an example horizon is shown in Figure 2. Response distributions are formed by simulation, and can be processed to give containment intervals (e.g. 95%). While most response distributions appear symmetric, the response can in some cases be highly skewed where there is little training data (rare soil classes). The simulated water retention curves for the 500 realisations of a Barr_6a sibling (a moderately deep sandy loam) are shown in Figure 3.

The Soil Profile Simulator is fast, generating 10 000 realisations of a specified soil sibling as a Rdata file in 28 seconds. The tool can output the realisations as tabular data or as a SoilProfileCollection (the class used in aqp). Use of the R aqp package facilitates drawing graphs of the soil profile. Figure 4 compares the 500 realisations of estimated PAW with the PAW that is listed on the S-map fact sheets (123.17 mm). The fact sheet value is estimated using the mean soil hydraulic response with mean values for clay and sand content, along with mean horizon thickness, stone content, and modal horizon classification. The mean PAW of the 500 realisations is 122.96 mm with a standard error of 1.2356 and a standard deviation of 27.62. Figure 5 shows the 500 soil profile realisations in terms of stone, sand and clay content down the profile. The variability of PAW down the profile is shown in Figure 6, where the darker grey area is the 25% to 75 % quantile, and the dashed lines show the 95% confidence interval.

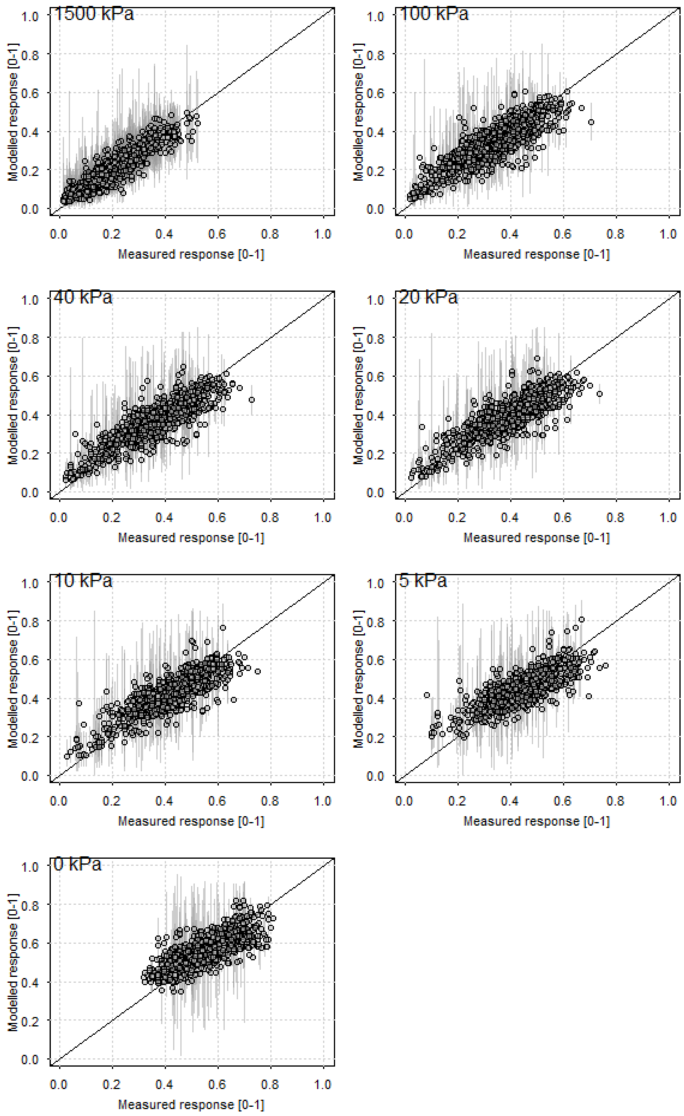

Figure 1: Measured-versus-fitted plots of the soil hydrological response for seven tension values. The vertical lines are plus and minus one standard error of the prediction uncertainty.

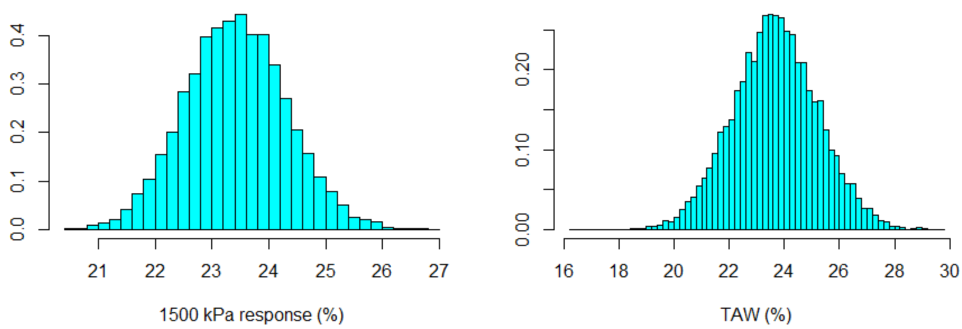

Figure 2: Histograms showing the simulated error (1500 kPa and total available water (TAW)) for an Allophanic soil with a loamy functional horizon (sand 15%, silt 57.5%, clay 27.5%).

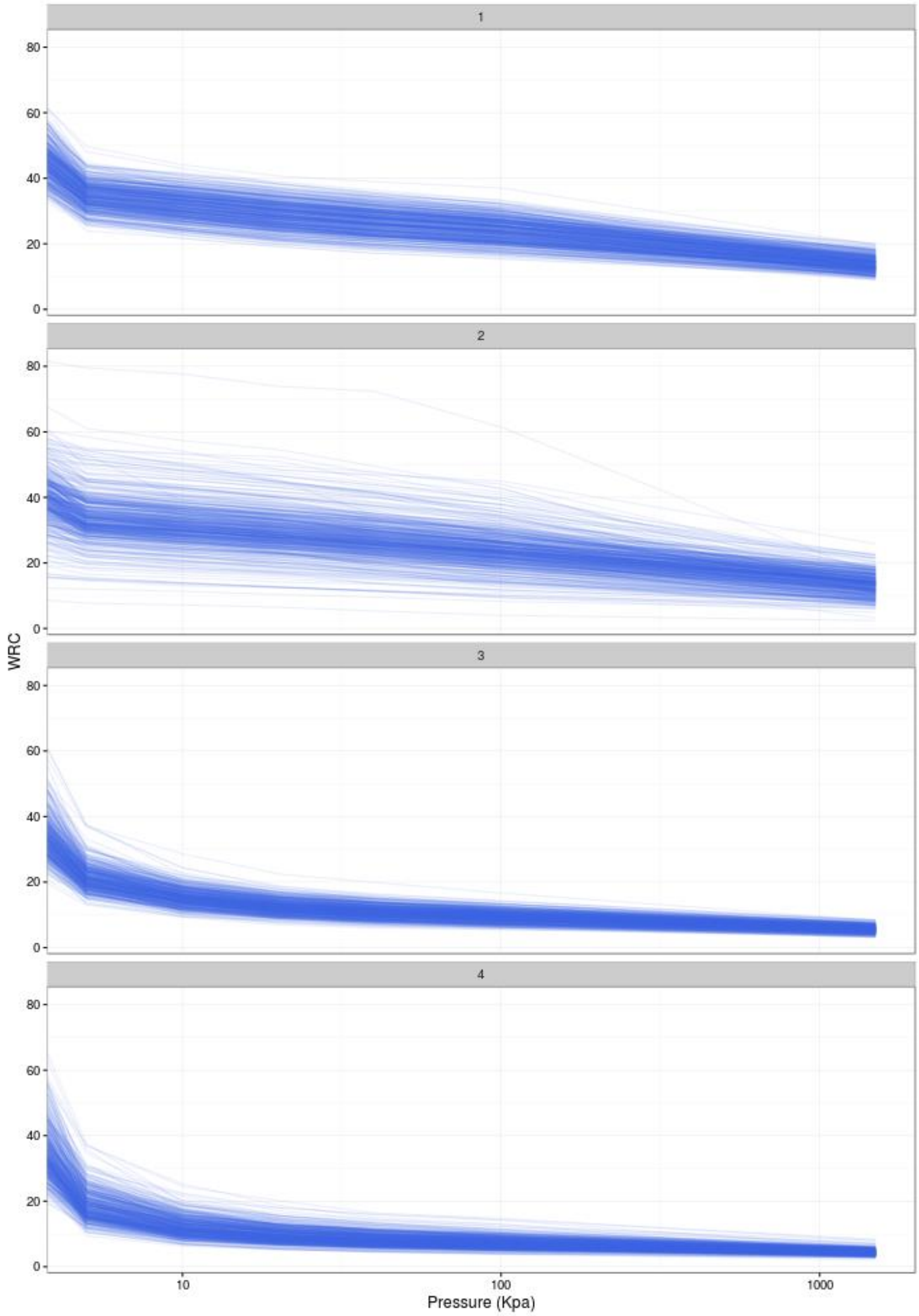

Figure 3: Simulations of the water retention curve for each of the four horizons of the Barr_6a sibling (a moderately deep sandy loam) as simulated by the soil hydraulic PTF error model.

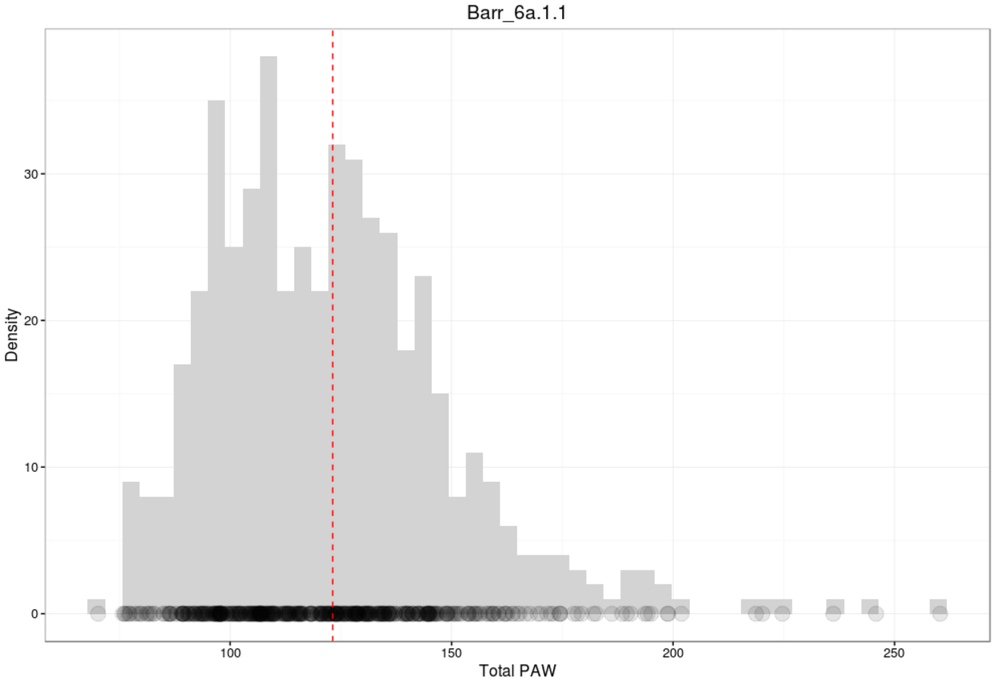

Figure 4: Histogram of the profile available water (PAW) of 500 realisations of the Barr_6a sibling. The red line shows estimated PAW where the mean values are used for the soil hydraulic response; clay, sand, stone content; and horizon thickness. The circles along the x axis indicate the distribution of the PAW estimates.

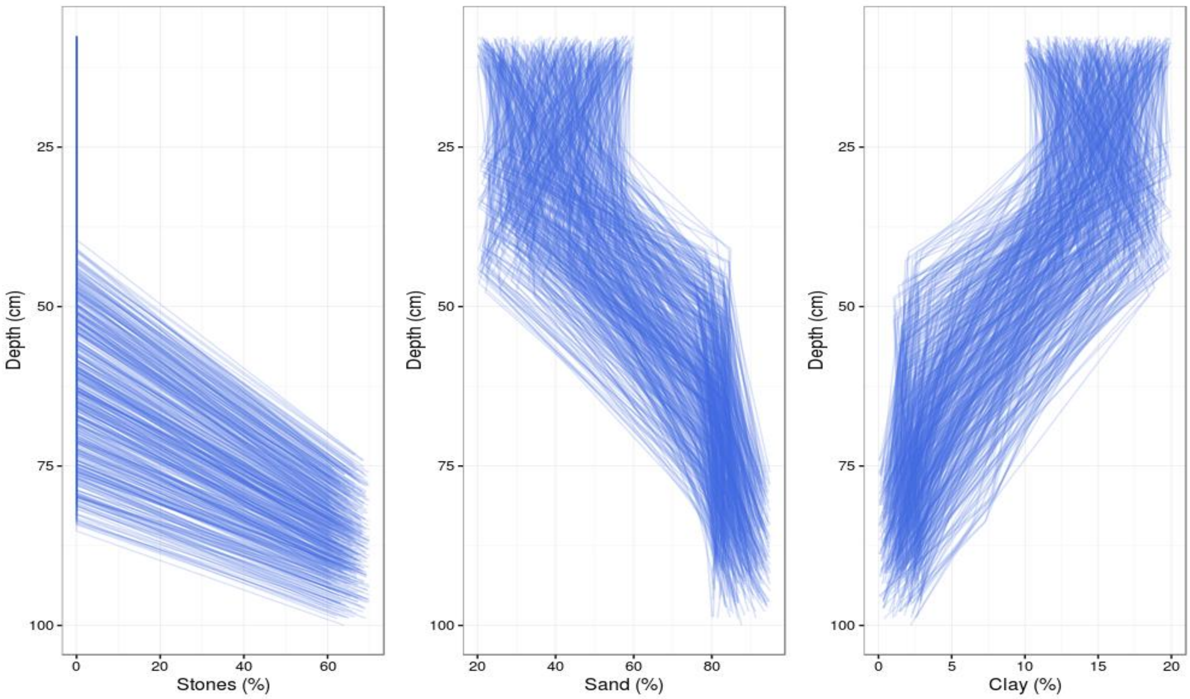

Figure 5: Soil profile realisations (n = 500) from the Soil Profile Simulator showing the variability of stone, sand and clay content down the soil profile of the Barr_6a sibling.

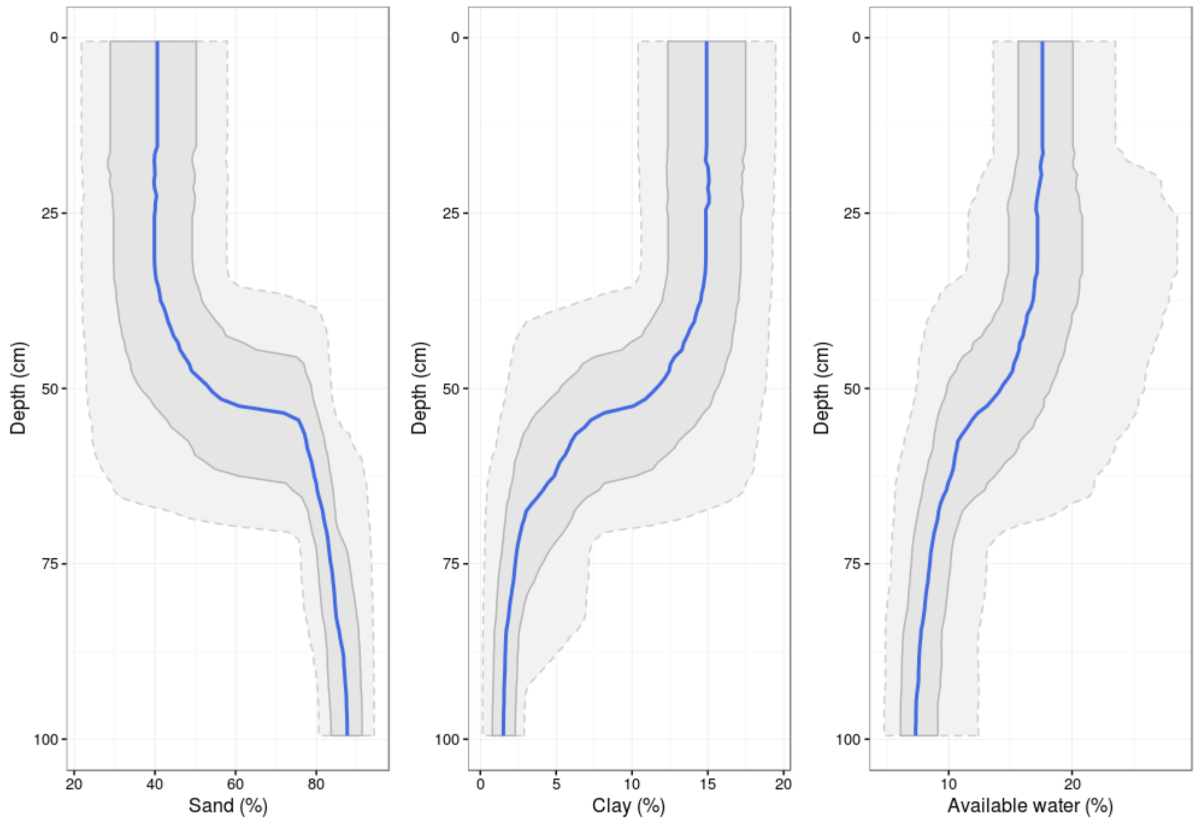

Figure 6: The variability of profile available water (PAW) down the Barr_6a soil sibling profile. The blue line is the median PAW from the 500 realisations, the solid and dashed grey lines indicate the 5% and 25% quantiles, respectively.

The speed of the Soil Profile Simulator tool also allows many siblings to be simulated according to the uncertainty of the proportions of each soil type within the soil survey polygons. Thus map realisations of PAW can be generated.

Discussion

Spatial uncertainty

The use of an expert approach to recording information about uncertainty within S-map is currently limited by the lack of any quantitative information about the spatial accuracy of the polygon line work. In the future, as we build in knowledge of soil-landform relationships, either explicitly or via digital soil mapping approaches, we will be able to represent this spatial uncertainty. For example, in a case where a soil type is linked to concave slopes and gullies, the weaker the membership of a DEM pixel in the concave-slope/gully class, the more uncertain the association with the soil class. This uncertainty can be added to the simulation tool. An alternative approach could be to use an expert approach to simplistically quantify the possible offset of each polygon line boundary.

Correlation

Another current limitation is the lack of information on both correlation between soil properties and spatial autocorrelation. The former was accounted for by using an approach whereby simulated profiles that did not meet known constraints were discarded. For example, S-map contains pdfs of horizon thickness. Some horizons will be inversely correlated – if one horizon is thicker than the mean, then at least one other horizon must be thinner in order for the overall depth (up to 1 m) to be maintained. If the sum of the simulated thicknesses was not consistent with the range of the depth pdf, then that realisation was discarded and another regenerated. The tool stops after one thousand attempts to generate a valid realisation. However, lack of information on the spatial autocorrelation of soil properties means that this aspect cannot be addressed at this point.

Conclusion

Conventional soil survey databases do not lend themselves to exploring the impact of uncertainty in soil information. However, progress has been made in deriving and visualising quantified estimates of uncertainty in key soil properties related to the soil water holding capacity. This has been enabled by recent advances in technology, including new R packages (aqp, VGAM) and Rserve.

Acknowledgements

This work was funded by Landcare Research Core Funding. We thank Stella Beliss and Varvara Vetrova who kindly reviewed the manuscript, and Leah Kearns for editing it.

References

Beaudette, D.E., Roudier, P., O'Geen, A.T. (2013). Algorithms for quantitative pedology: A toolkit for soil scientists. *Computers & Geosciences 52*, 258–268.

S-map - New Zealand's national soil layer. DOI: http://dx.doi.org/10.7931/L1WC7. Accessed: 2016-05-16.

Lilburne, L., Hewitt, A., Ferriss, S. (2006). Progress with the design of a soil uncertainty database, and associated tools for simulating spatial realisations of soil properties. In: M. Caetano, M. Painho (Eds.), 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. Instituto Geografico Portugues, Lisbon, Portugal, pp. 510-519.

Lilburne, L.R., Hewitt A., Webb, T. (2012). Soil and informatics science combine to develop S-map: a new generation soil information system for New Zealand. *Geoderma 170***,** 232–238.

McNeill, S., Lilburne, L., Webb, T., Cuthill, T. (in prep.) Pedotransfer functions for hydrological properties of New Zealand soils using S-map information.

Yee, T.H. (2015). *Vector generalised linear and additive models*. Springer.