



Comment on soil-2021-80

Alex McBratney

Community comment on "How well does Predictive Soil Mapping represent soil geography? An investigation from the USA" by David G. Rossiter et al., SOIL Discuss., <https://doi.org/10.5194/soil-2021-80-CC1>, 2021

Mercedes Román Dobarco, Alex McBratney, Budiman Minasny, José Padarian, Alexandre Wadoux (The University of Sydney) [alphabetical]

Digital Soil Mapping

It is interesting to query why the terminology *predictive soil mapping* is used as opposed to the long-standing and more common term *digital soil mapping* (DSM). There is a cogent argument that **all soil maps are predictive** in the sense that they make statements about soil entities (classes, properties, functions) at unvisited or unobserved locations (usually on the basis of having some observed locations). This is certainly the case in conventional soil mapping. Conversely, we may argue that not all soil maps are digital, some being still in an analogue format and require digitisation prior to being used in a computer system. The authors make a few uses of the term digital in this manuscript when referring to digitised products based on field surveys (without statistical modelling, what we usually call conventional soil maps). Deviation from the more common term *digital soil mapping* brings some confusion, such as when citing Arrouays et al. (2020) at lines 255-257: Arrouays et al refer to digital soil map in the sense that the current authors use predictive soil maps, while the current authors call digital soil maps the digitised product based on field surveys. The use of the term *digital soil mapping* in the considerable literature of the past 20 years has been that the rasterised predictions are quantified in some way and have an associated uncertainty measure. We do not see the need for the confusion through the use of the term '*predictive soil mapping*' when a perfectly well used and understood term is available.

Comparisons are made between three kinds of digital soil mapping. The SoilGrids and its USA SPCG approach is the more classical one in which point observations across the United States are interpolated through the use of a set of covariates representing scorpan variables. The predictions are made by highly non-linear and multivariate prediction functions. Polaris+ is made by quite a different approach. It is two-stage procedure, both stages are firmly based on the SSURGO mapping, series descriptions and associated data. In the first stage soil mapping units are disaggregated in a probabilistic manner to soil taxonomic units (soil series). This is POLARIS – a 30 m probabilistic digital map of soil series across the United States. Covariates somewhat similar to that used for SoilGrids are

used in this process. In the second stage, properties are calculated from weighted averages of properties reported for each STU – using triangular distributions using the minimum, maximum and average. This gives a 30 m raster of soil properties over various depth intervals. This product is POLARIS+. The POLARIS products are a harmonised reinterpretation of SSURGO with a filling in of some areas where maps were not previously available.

Soil Geography

The paper talks about soil geography. What might we mean by that? Generally, it can be taken to mean the spatial distribution of soil entities or the evolution of the spatial distribution of soil entities. Which entities? Normally they are soil classes (taxonomic units or mapping units). In this paper soil geography is restricted to the distribution of level sets obtained by discretising the continuous maps of soil properties (with particular focus on soil pH). This can be done of course and described by the calculus of random sets but the patterns and divisions are somewhat arbitrary and would be better characterised by spatial methods that recognise continuity. (The variogram which is used is one of these – it is well known that the variogram of predicted variables is always lower than the variogram of soil observations – because of regression.) Description of the spatial distribution of soil classes remains an underdeveloped area of pedometrics.

Ground Truth

Considering that both DSM and “conventional” soil mapping have uncertainties and often a different focus (classes vs properties), why one should be used to measure the quality of the other? Would we not reach a similar conclusion if we take maps from different soil surveyors to compared them with a DSM product?

The authors propose methods to evaluate maps generated using DSM which is a valuable contribution. However they use “conventional” polygon maps as a ground truth as if the final goal of DSM is to recreate a polygon map (which is doable but generally not the goal). Both DSM and “conventional” soil mapping have an associated uncertainty (accounted or not). The authors mention some of the uncertainties of “conventional” mapping:

- L69-70: “multiple survey projects over time with inconsistent standards and mapping concepts, inconsistency among mappers, difficulties in objectively identifying boundaries, and indeed the need to identify boundaries”

The authors do not mention the intrinsic uncertainty of mapping units which are not homogeneous as a single polygon might suggest.

The paper uses as reality a soil surveyor’s expert knowledge of the soil landscape which is not to be denied. However that knowledge is generally in terms of the spatial distribution of soil classes rather than soil properties, and the North American mental model tends to focus most on soil topographic relationships whereas the digital soil mapping approaches are more explicitly multi-factorial. **In order to do a convincing comparison, it is important to have an independently observed dataset with which to compare the various representations else we might simply realise a self-fulfilling prophecy.**

Ideally such a comparison of the various maps with the independent observed dataset will be made in a statistically robust way, i.e. through the use of probability sampling and design-based inference. Statistical validation also benefits from summary diagrams that communicate various aspects of quantitative map quality. This way different aspects of map quality are represented, for example, the degree of smoothing, the precision, but also the spatial pattern. There are in the literature various examples of such diagrams, the

most common of which is the Taylor diagram, but there are also the target and solar diagrams that further include information about systematic differences between maps (bias).

The Way Forward

We agree that "there is no substitute for actually examining the soil and landscape", but given the various restrictions (in budget, space and time) and the need to address the relevant needs of end users who have demanded quantitative data and information, DSM will most often be the method of choice for efficiently tackling real-world problems.

Separately, more work is clearly needed on all aspects of the quantification of both soil classes and maps of such classes and the quantitative formulation of their spatial distribution and evolution all of which has developed only very gradually over the last forty years. We can see some initial glimpses of such developments in the current paper and we thank the authors for that.