



Comment on soil-2021-80

Bradley Miller (Referee)

Referee comment on "How well does Digital 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-RC2>, 2022

The spatial analysis techniques presented in this manuscript introduce important evaluations that need to be considered when comparing soil map products. However, the manuscript would benefit from a more thoughtful consideration of the meaning of the evaluations chosen. The metrics selected offer a variety of calculations, but it seems possible that they may be reflecting some of the same differences between the maps. Even after attempting to mute the effect of resolution, some patterns in the results remain. For example, some smoothing should probably be expected from models that – at their core – rely on regression fitting. So there appears to be some opportunity here to go beyond just reporting differences in the metrics. The metrics are intended to measure map characteristics other than relative smoothness. Can all the differences detected by the metrics be attributed to relative smoothness, and if not, what other map characteristics might these metrics be picking up on? I advocate for this because if others are to be convinced to apply these evaluations elsewhere, they will likely want to have some sense of how the results should be interpreted and/or what should be learned from them.

The spatial analysis of map properties promoted in the present manuscript reminds me of the books "Soil and Landscape Analysis" by Hole and Campbell (1985) and "Pattern of the Soil Cover" by Fridland (1977). Although dated in their focus on analyzing polygons, these books suggested several ways additional information could be evaluated from the spatial patterns in the map. Then again, the methods proposed here essentially go back to a polygon type of analysis since they require the map to be classified.

The way in which the mapping methods for the different products are described is sometimes unsupported and potentially not fully relevant to the spatial analysis applied. The distinctions between machine-learning models (categorized as PSM) and traditional soil mapping seem questionable. If the authors want to keep these assertions, then

citations or better explanations for why they think they are true need to be added. A building concern in the description of traditional soil survey is a potential failure to recognize what the two approaches being compared have in common, which is important for understanding the differences in the respective products.

To improve the manuscript, I would encourage the authors to consider how the strategies of the different mapping approaches may be connected to the results of the spatial evaluation metrics applied. As an example, if we were to recognize traditional soil survey as a predictive map product, then the covariates would largely be what the mapper could see in the available airphoto. Although the airphoto bases used for traditional soil survey leave a lot to be desired compared to modern covariates, one can see more detail of landform shapes than with a 100m digital terrain data set. Although resolution is already partially alluded to in the text, this kind of context may help sort through what may be driving some of the differences detected with the various map evaluation metrics.

While I offer considerable critique on the characterization of traditional soil survey and push for some consideration about what the selected metrics really describe about the maps being evaluated, I applaud the spatial analysis approach to evaluating different map products. Thinking about how to evaluate maps beyond the prediction of points is an important contribution to the advancement of soil mapping as a science.

General Comments

- Please consider being more consistent with terminology and abbreviations. For example, traditional soil survey versus conventional soil survey and map scale versus design scale. They appear to be used interchangeably, but the change in term makes the reader wonder if there is a difference being implied and if so, what that difference may be. Regarding terminology, I disagree with several of the terms used in the present manuscript but attempt to use the terms most frequently used by the authors in the present manuscript to facilitate communication in this review of the manuscript.
- Information presented in sections 1 through 2.3 sometimes circles back on itself. Consider reorganizing to avoid repeating some information. I think this would also help readers understand and compare the processes by which these maps are made along with the resulting map characteristics.
- Most figure captions are more like titles. Please elaborate in the figure captions to guide the reader in what to look for in the figures.

Specific Comments

- L53-55 – The acknowledgment of the Scull et al (2003) paper using the term ‘predictive soil mapping’ is appreciated. However, I’m concerned that the use of this term could lead to confusion over what differentiates traditional soil survey from the new

approaches that utilize computational algorithms to produce soil maps. The potential issue here is the perception that traditional soil survey is somehow not predictive. Of course, it is not possible to observe soil everywhere, which requires some level of spatial prediction for any kind of soil map with exhaustive coverage.

- L62 – How fair is it to say that machine-learning models can be implemented with fewer locations visited when the machine-learning models presented for comparison in this manuscript rely on a database of observations collected by the activity of traditional soil mapping? In statistical theory, it makes sense that some optimized sampling design should be able to capture the needed information. But has this been the experience of soil mapping?
- L64 – Please explain how PSM has a greater ability to map inaccessible areas than traditional methods.
- L73 – The assumption that mapping scale drives resolution is largely a concept held over from paper maps. In many disciplines (e.g., geology) we are seeing newer generations of maps adding detail without changing the extent of the map. Adding those details (making the resolution finer) has become functionally possible because the producer and the user can ‘zoom’ in and out of the map in a GIS. This technological development renders the question of mapping scale nearly mute. However, the remaining question is if there is sufficient information to support the finer resolution.
- L80 – This paragraph appears to present a non sequitur. The argument is made that point evaluations of PSM do not consider the spatial pattern of predictions. However, it is not clear to me how the subsequent information presented about traditional soil survey methods shows how that approach does more to evaluate spatial patterns of predictions.
- L83-88 – This section appears to be building an assumption that traditional soil survey does not include any kind of model that uses input data to make predictions. Although the authors point out real shortcomings in the “paradigm” of traditional soil survey described by Hudson (1992), they have left out how that approach uses ‘mental models’. This omission obscures what machine-learning and traditional soil survey methods have in common.
- L91-95 – The cartographic reasons that traditional soil survey is constrained to the levels of detail it has is a useful explanation here. However, for translating polygons delineated in USA soil survey maps, there is a more direct approach. The USA soil survey program has a strict protocol for minimum delineation areas. The USA’s “Field Book for Describing and Sampling Soils” specifies minimum-size delineations of 0.6 ha for 1:12,000, 1.0 ha for 1:15,840, and 2.3 ha for 1:24,000. Now, we should note that those are minimum delineation sizes, and the mean delineation size will be larger than that. The mean delineation size can vary by landscape and by the style of the mapper.
- L126 – The use of STATSGO2 is interesting here. With the possible exception of some areas where only an order 5 map has been made, STATSGO is a purposefully generalized map product that is aggregated by expert knowledge. At first, I questioned if it made sense to include STATSGO in this evaluation, but then STATSGO was not evaluated. Considering STATSGO is not evaluated in this manuscript, it does not seem relevant.
- L128 – “State” should not be capitalized.
- L162-169 – This is a nice, succinct description of the state of SSURGO.
- L175-183 – I think it may be misleading to state that SG2 does not use any information from SSURGO when SG2 uses profiles from the NRCS pedon database. Yes, SG2 is not using SSURGO itself, but they are both using a set of training points that they have in common for the USA area. This overlap in source information is even more so for the SPCG, which makes use of additional pedons that were produced from the activities of the USA Soil Survey. In the case of the RaCA dataset, it is new enough that it probably has not strongly influenced the SSURGO map. Nevertheless, the role of these training points in all the map products should be explained clearly. Specifically, I disagree with the idea that SSURGO is independent of the data points managed by the NRCS.
- L199 – Add space after the period.

- L247-252 – Libohova et al. (2014) explored the validity of these ranges to represent uncertainty. That evaluation seems relevant here.
- L267 – Consider a rubric here to define how the expert judgement will be evaluating the maps. This will help the reader understand the value system being applied in this evaluation and communicate a more structured approach to how the qualitative comparison will be made.
- L283 – Where does variability come from for any single point in these maps. Won't there be a single value for a grid cell, or is there something else being brought in here? Is this using the uncertainty ranges? In any case, please explain clearly to help the reader know the basis for the proportional nugget.
- L383-L388 – This paragraph drifts into results by beginning the evaluation. Recommend keeping the description of methods separate from the results found by implementing them.
- Table 1 – Please be consistent in abbreviations.
- L408 – Change 'distributions' to 'distribution'
- L429 – Add missing 's' after 'PSP'
- Figure 3 – Consider including the r values in the boxes.
- Table 2 – Again, please be consistent in abbreviations, both for matching abbreviations used in the text and previous tables.
- L446 – add 's' at end of 'indicate'
- Figure 4 – The 'SoilGrids' and 'SPCG100' maps show large areas of pH lower than shown in the 'gNATSGO' in the high elevation portions of the east and south areas. Could this be a case of extrapolation in the feature space? If so, how might this be reflected in the evaluation metrics presented in the manuscript?
- L447-448 – The authors suggest that the smoothing effect was caused by the spatial continuity of the covariates, which seems reasonable considering the resolution. It seems to me that any kind of fitted model is, almost by design, going to smooth out some patterns in the training data. Would the authors mind commenting on this possible additional factor?
- Figure 10 – Did the 'SPCG100USA' semivariogram actually reach a sill?
- L462-463 – These sentences are a little unclear; please consider rewording and/or expanding upon the explanation.
- L468-469 – If the difference between gNATSGO and PSP is going to be called out, maybe it is worth mentioning that the difference between gNATSGO and SPCG is even more. Some discussion about why the difference between gNATSGO and PSP captured the authors' attention may also be warranted.
- L473-474 – Just to be clear, which PSM product is being referred to here?
- L478 – This is the first mention of silt concentration! This switch makes for a mismatch between the methods described and the results presented.
- L480-482 – This content would be better suited in the figure caption.