



Comment on soil-2021-80

Anonymous Referee #1

Referee 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-RC1>, 2021

This paper describes a new evaluation approach of DSM products that is based on qualitative and quantitative comparisons with local soil maps obtained from field survey. This is a valuable attempt for going beyond the evaluation approaches currently practiced in the DSM community, which consists in checking only the good reproduction of point observations without considering, as the author stated, "the soil landscape, as interpreted by the soil surveyor and as managed by the land user".

The evaluation approach includes a very large set of metrics. Some of them account for similarities/dissimilarities of soil patterns between DSM products and soil maps and between different DSM products (MD, RMSD, RMSD "adjusted", V-measures). Other metrics characterize separately some spatial properties of these products (variogram parameters, landscape metrics). These quantitative comparisons are completed by visual comparisons after overlaying classical soil maps and DSM products.

The authors present two examples of application of these metrics at two different scales and for two soil properties (comparison of regional pH patterns and local silt patterns). The overall conclusion of the paper is that all metrics show differences between DSM products and between DSM products and reference soil maps at the noticeable (and surprising) exception of SG2 vs SPCG. No clear and convincing hierarchy across the three DSM products is revealed by the analysis of the result.

As I said before, this paper is interesting because it explores alternatives to point-to-point comparisons of soil property measurements for evaluating DSM products. I fully support the idea of using local soil maps elaborated by experienced soil surveyors as an alternative (complementary) ground-truth, despite the well-known weakness of soil maps. However, the examples given in the paper do not fully demonstrate that the proposed metrics are necessary, relevant and sufficient to provide a comprehensive evaluation of the different DSM products that can help an unexperienced end-user in choosing between the different available DSM products he/she can get in a given local territory. In this perspective, I see several questions that should be supported by adequate examples and further discussed: which metric best account for the visual differences of soil patterns?, which metrics are redundant to each others? Which metrics best discriminate the different DSM products? I can understand that the authors have not studied enough examples for addressing today these questions but they should be presented in the discussion as the

way forward to obtain an evaluation method that should be communicable to users.

The paper is generally well organized. However, the authors used a fairly unusual way to present their data (section 2) i.e. instead of presenting separately each DSM and soil survey products, they chose to deliver the information progressively and "in parallel" along a set of sections that are not always straightforward for an external reader. Furthermore, this induces some redundancies and contradictions (see examples in my comments along the text). Finally, the study areas selected as examples for regional and local spatial patterns are, if not presented (the local), not presented with the necessary details (regional) for allowing the interpretation of the results (see my comment along the text). To conclude, I think that the presentation of the soil data considered in this paper (actually in sections 2 and 4) needs to be deeply reworked to improve the understanding by an external reader not familiar with US context.

Comments along the text:

Title and line 1: Why do the authors rename "predictive soil mapping" what is currently known as "Digital Soil Mapping"? This introduces a useless complexity. I therefore suggest to replace "predictive soil mapping" by "Digital Soil Mapping products" in the title and by "Digital Soil Mapping" or "Digital soil mapping products" along the text following that the authors refer to the technique or to its results.

Line 114: "data sources" is only the first part of the section (before 2.1.) isn't it? If yes you should replace "data source" by "soil data" and add a subsection "data source" immediately after.

Line 138: Contrary to what is suggested here "Polaris soil properties" is not further systematically replaced by "PSP". A lot of "POLARIS" remains in the text and in the figure. This should be corrected.

Line 152: gSSURGO has not been presented before

Lines 184-195 (Section 2.2.) :

- a) why mixing in a same section "environmental Covariates" and "geographic scope"? The relation between them is weak. The statements that deal with the latter (Lines 191-193 and 195) would be better located in the next "mapping methods" section (see my next comment)
- b) perhaps more comfortable for the reader to know first the set of covariates used by SG2 and also, as I understood, by the two other DSM products. Then cite the specific covariates that have been added to PSP and SPCG.

Lines 197-205 : this paragraph is largely redundant with lines 151-161.

Lines 218-224 : insert here the geographical scope and the number of samples used for training the models.

Line 231: I understand here that gNATSGO is a generalization of gSSURGO. However it was stated before (line 125) that "gNATSGO is a composite of [...] SSURGO and [...] STATGEO and [...] RSS. Please, explain more clearly what are the differences between gNATSGO and gSSURGO.

Lines 255-257: Even before looking at your further results, we could expect that soil survey products and DSM products do not converge toward similar uncertainties assessments. At best, we could expect that the level of uncertainties mapped by these two products could be ranked similarly, independently from their absolute values. You should select a metric for representing this.

Line 272: "RMSD adjusted for MD". Not clear. A mathematical formula would clarify.

Line 301: "region" is qualitative. How can we calculate the variance of a qualitative variable?

Lines 359-375: The sections "Regional patterns" and "local patterns" are redundant with the introduction of section 3 (lines 259-265). This should be re-organized.

Lines 377-349 (section 3.6.1. "visual method") Does this section refer only to "local patterns" (section 3.6.1.)? I don't think so since you provided further visual comparisons for both regional and local patterns. This should be clarified. Furthermore, this section looks redundant with section 3.1. ("qualitative methods")

Figure 1: This study area looks different from the ones considered further (Figures 4 to 12). This could explain why there is a so great discrepancy between the visual inspection and the quantitative results obtained further. Similar problem occurs with figure 13 (what is this study area?) . Please give the length and width of the rectangle for a better appreciation of the scale.

Lines 393-402 (section 4): It would be useful to have more information about the study area that is finally selected as example of regional spatial pattern comparisons (size of the rectangle, scale of the soil survey product gNATSGO at this location, average size of

polygons, pedology, landscape drivers of soil variability etc...). All these data would be very useful for interpreting the results. This comment applies also to the description of the study area selected further as example of local spatial patterns comparisons. I did not find any information on this area.

Line 412: provide the significance of the sizes of the circles in figure 3. Indicate what you mean by "well-correlated" (threshold?)

Line 464: There is an apparent contradiction with the concluding statement here "overall the agreement is fairly good" and what is written just before (line 459): "gNATSGO is considerably different from all other products"

Lines 539-541: I suppose you should replace "...and covariates limited in geographic scope to the USA" by "...and input soil data limited in geographic scope to the USA". To my opinion, this is the most surprising result of the paper. Have you any explanation ? to my opinion, the similarity of machine learning algorithms cannot be a convincing explanation.

Line 543-544. you cannot conclude that because the specifications for defining these confidence intervals are different between the DSM and the soil survey products. Furthermore, there is not any ground truth to identify what is the most realistic CI.

Lines 556-560. I disagree with your diagnosis on PSP. Figure 14 clearly shows that PSP does not bring more knowledge on soil variations than the initial soil survey product from which it was derived. Furthermore, you cannot say that PSP can be useful for unsurveyed areas since PSP requires a soil map as input