SOIL Discuss.,
https://doi.org/10.5194/soil-2019-36-RC2, 2019

Interactive comment on "Comparing three approaches of spatial disaggregation of legacy soil maps based on DSMART algorithm" *by* Yosra Ellili et al.

Caroline Chartin (Referee)

caroline.chartin@uclouvain.be

Received and published: 24 September 2019

General comments

This paper focuses on testing if, and in which way, disaggregating legacy soil map is improved by adding supportive data in the procedure, i.e. soil legacy data and soil-landscape relationships deduced from local expert knowledge. The purpose of this study is important given the lack of accurate soil information in many regions of the world. Those are particularly needed to better face the current process threatening/degrading soils. Moreover, some methods tested here could considerably help diminishing the time and cost for producing new accurate soil maps by reducing field-

work efforts.

In my opinion, the manuscript is mostly well structured, logical, and the language correct.

However, I have some concerns about the approach and the methodology.

Specific comments

My concerns about this study join those highlighted earlier in the discussion by the referee Madlene Nussbaum.

Indeed, the authors proposed to compare three methods of disaggregation, each based on the DSMART algorithm, and to test them on the Ille-et-Vilaine department. As far as I understand, the method 3 was proposed by Vincent et al. in 2018 who applied it to the entire Brittany (which includes the Ille-et-Vilaine department) using the same covariates (at the same resolution) and validation databases used here, but obviously at a bigger extent. Although Vincent et al. (2018) do not detail the results obtained by using the classical version of DSMART (i.e., the Method 1 here), they already visually compared the maps resulting from Methods 1 and 3 on a reduced area of Ille-et-Vilaine.

The maps obtained here and in Vincent et al. (2018) showed that only $\sim$ 20 % of the validation data had been correctly predicted. The authors of the latter study already highlighted that adding the soil-landscape relationships (Method 3) did not substantially improve the results accuracy but tend to produce a more pedologically coherent map. Hence, the authors of Vincent et al. (2018) proposed different coherent ways to optimize the disaggregation procedure and improve its performance (through improving soil data, covariates, and predictive models).

Here, the authors proposed to improve input data by combining DSMART algorithm to legacy soil data. Unfortunately, the legacy soil data are very largely outnumbered by the observations created artificially by the algorithm, limiting greatly their potential

effect on the model performance. In this context, applying Method 2 is almost the same as applying Method 1. A weighing procedure should be implemented in the procedure for the Method 2.

Considering the low performance of the different methods, I suggest the authors to dig in more in improving the procedure as suggested by Vincent et al. (2018) before applying a pairwise map comparative study. For example, the use of the legacy soil data could be optimized in Method 2 as proposed above and more complex and efficient ensemble tree methods could be tested (e.g., random Forest, cforest. . .) which have many advantages as, among others, integrated validation procedure and clear estimation of the respective variable importance in the model.

Technical corrections

l.143: Please, replace the underscore '_' by the dash '-' in "0_20 m" and "20_50 m".

l. 165-178: §2.2.2. 'Soil validation data'

- As the existing detail maps define one of the three validation datasets, l. 173-174 should be aligned with l.167-172.

- The dataset extracted from 'Sols de Bretagne' is used for validation of M1 and M3 but also used as calibration datasets in M2: it has to be clear somewhere in the text.

- Could you please precise what are the main characteristics considered by an expert to define a STU and how you converted legacy data points and vector maps to raster (l. 176-178)?

l. 277-291: The validation procedure should be more explicit and maybe improved by computing one or two more parameters in order to better apprehend the performance of the models.

l. 179-200: §2.3 'Soil covariates'

Please, could you quickly justify the choice of the covariates used in this procedure,

and maybe make a parallel with the characteristics considered for defining STU?

- The TPI and waterlogging parameters are categorized here. I understand that it facilitates the computation of the soil landscape relationships, but have you try to input the continuous versions of these parameters in the models?

- The landscape unit parameter is an aggregation of vegetation, land use and relief attributes. Why did you prefer to use one aggregated layer instead of more accurate maps about land use, vegetation and relief attributes? Is there a significant correlation between all of these parameters? Is it in order to take into account the landscape morphology at different scales, i.e. main features with the Landscape units and then local features within thanks to more accurate relief attributes layers?

l. 256: Could you precise which proportions of the 18,320 samples used in the Method 3 are derived from expert knowledge and from the random selection implemented in the DSMART algorithm?

Figure 1: Please, reduce the size of the dots or change to triangles. Precise the scale of the detail maps.

Figure 3: Please, could you precise the names of the STU in the legend or in the caption.

Figure 6: Please, add the x-axis labels.

Figure 7: Please, harmonize the covariate names with main text.

I wish my comments would be considered and helpful. I stay available for any discussion.

Best regards, Caroline Chartin

Interactive comment on SOIL Discuss., https://doi.org/10.5194/soil-2019-36, 2019.