SOIL Discuss.,
https://doi.org/10.5194/soil-2020-65-AC2, 2021

Interactive comment on "SoilGrids 2.0: producing quality-assessed soil information for the globe" *by* Luis M. de Sousa et al.

Luis M. de Sousa et al.

laura.poggio@wur.nl

Received and published: 10 February 2021

Thank you very much for your useful comments. Below you will find detailed answers.

> L43-44. "DSM consists primarily in building a qualitative numerical model between soil observations and environmental information acting as proxies for the soil forming factors." Not only, DSM may also use information acting directly as a proxy for a soil property (see for instance proximal sensing, remote sensing of bare soils, etc.)

The sentence has been modified to:

DSM consists primarily in building a quantitative numerical model between soil observations and environmental information acting as proxies for the soil forming factors. DSM can also integrate direct information as proxies for soil properties, for example proximal sensing measurements.

> L49 put e.g., "country (e.g. Mora-Vallejo. . ..)" or enlarge the list of citations, many countries are omitted: e.g., Australia, US, Denmark, France, China, India, South Korea. . .)

The manuscript has been modified to recognise that these are only examples. "e.g." has been added to all relevant references.

> L90. Not sure EU-Lucas (2013) can be considered as a soil profile DB. This is only topsoil, isn'it? May be find another way to say this.

The manuscript has been modified using soil observations database instead of soil profiles database.

> L. 93-94. The readers would be interested in knowing briefly what were these minor corrections.

The manuscript has been modified by adding some examples of the minor checks, "for example further depth congruence checks".

> L. 131-132. Not sure this is clear for all the readers, a scheme or a flowchart would be useful.

The text was expanded to make the procedure more clear. We do not think that a full flowchart is necessary to explain a 10-fold split of the observed data.

L. 147-148. “The long-term average and standard deviation of climatic variables and vegetation indices were computed from monthly data to capture their seasonal dynamics.” Be more precise: how long? From which date to which date?

The manuscript has been modified to include this information:

The average and standard deviation of climatic variables and vegetation indices over 15 years (2001 - 2015) were computed from monthly data to capture their seasonal dynamics.

L. 164. Why alphabetical order?

We think it is not that important which of the two covariates is retained. We used the alphabetical order to mimic a random choice of the covariates. Please note that we could not retain the covariate that has the strongest bivariate correlation with the dependent variable because we ran the de-correlation analysis as a preliminary step and not separately for each of the soil properties considered.

At the end of section 2.3.2 we would like to know how many covariates finally remained – this indicated in further tables but you could say that a number of covariates ranging between XX and xx were retained depending on the soil property.

The results mentioned are already in Section 3.3. Moving the results here would mix the results with the methods. We think it is better to keep them separate.

Table 2. I’m very surprised to see that sometimes the number of observations increases with depth. Logically it should be the reverse, no?

We carefully checked the numbers in the table and they are correct. Please note that the thickness of the six standard depth intervals is not constant and increases with depth. This could explain that the number of observations increases with depth.

Table 4. Missing units.

The units are summarised in Table 1. A reference to Table 1 has been added in all relevant table captions in the revised manuscript.

L. 299-300. So what? Needs a discussion. What is the most important? Think there is a paper by Samuel Rosa et al in Geoderma discussing these effects of the nb of covariates and the nb of points.

Here we point out an interesting result from the table. We did not do a study varying the number of points and covariates to come to a general conclusion on this. The paper referred to is, we think Samuel-Rosa, A., Heuvelink, G. B. M., Vasques, G. M. and Anjos, L. H. C.: Do more detailed environmental covariates deliver more accurate soil maps?, Geoderma, 243–244, 214–227, https://doi.org/10.1016/j.geoderma.2014.12.017, 2015. But this paper deals with the spatial detail of the covariates rather than the number of them, or the number of points. We have added a brief explanation in the text to clarify why we point this out.

L. 304-305. This seems contradictory with the observation on table 2 (see before), I believe that “weakened relationships between environmental layers and soil properties of the deeper horizons” is more likely.

Thank you for spotting this inconsistency. The text has been modified to make this more clear and focusing on weakened relationships between environmental layers and soil properties of the deeper layers.

> L. 334 and further. “The USA and large regions of Europe and Australia have very high numbers of observations that could be reduced to further strengthen the spatial robustness of the validation procedure”. That’s true for validation, but you reduce the number of calibration points. Is not here a kind of trade-off between the quality of predictions and the quality/robustness of the evaluation of the performance of the validation?

This was poorly-phrased. We should not suggest reducing the number of calibration (training) points; rather, they could be weighted according to the degree of clustering (sampling density). The folds used for evaluation would be similarly weighted. We did not investigate this yet, it will be considered in future work. Some text detailing this was added to the text.

> Lines 414-419. “This work described only the modelling of some of the primary soil properties, as defined and described in the GlobalSoilMap specifications. More work is necessary to obtain maps for soil thickness (rooting zone, solum or regolith), soil properties derived with pedo-transfer functions e.g. hydrological soil properties as saturated hydraulic conductivity (Pachepsky and Rawls, 2004) and complex properties that depend on multiple primary properties, e.g., carbon stocks. These layers are important inputs to model and map soil functions in the present and in the future as well as to support Earth System Modelling (Luo et al., 2016; Dai et al., 2019)”. I think this is more discussion than conclusion, it should be seen as a limitation of the study, you should explain why these properties could not be predicted and suggest ways to improve the situation. Future progress on how to predict these parameters should be proposed or taken from the literature where they exist.

The conclusion section has been re-written. We agree that most of the content of

the Conclusions section of the submitted version was indeed more discussion than conclusion. In the revision we moved these parts to the Discussion section.

Interactive comment on SOIL Discuss., https://doi.org/10.5194/soil-2020-65, 2020.