SOIL Discuss.,
https://doi.org/10.5194/soil-2020-65-RC1, 2020

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

Dominique Arrouays (Referee)

dominique.arrouays@inrae.fr

Received and published: 10 November 2020

Review of the paper SoilGrids 2.0

This paper shows maps of soil properties for the entire globe at medium spatial resolution (250 metres cell size) using state-of-the-art machine learning methods to generate the necessary models. It takes as inputs soil observations from about 240 000 locations worldwide and over 400 global environmental covariates describing vegetation, terrain morphology, climate, geology and hydrology. The aim of this work was the production of quality-assessed global maps of soil properties, with cross-validation, hyper-parameters selection and quantification of spatially explicit uncertainty.

The main improvement compared to the previously published paper by Hengl et al., is

Printer-friendly version

Discussion paper

that the quality of the maps is assessed with cross-validation and quantification of spatially explicit uncertainty. This steps were missing in the previous versions of SoilGrids. Therefore, this is a huge progress that merits publication in SOIL.

I have only a few minor comments on the MS.

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.)

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. . .)

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

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

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

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?

L. 164. Why alphabetical order?

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.

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

Table 4. Missing units.

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.

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.

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 que quality of predictions and the quality/robustness of the evaluation of the performance of the validation?

Good discussions in section 3.4 and 3.5!

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.

Overall a very nice piece of work that merits publication after minor changes and some

development in discussion. Looking forward to seeing it published in SOIL!

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