SOIL Discuss.,
https://doi.org/10.5194/soil-2020-65-AC3, 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.

> About the title: I think that assessing the quality of the resulting soil information from digital soil mapping is a very common practice. The assessment can only be a cross validation of prediction accuracy. The "quality-assessed" seems not an obvious difference from other works. I guess that the biggest progress with regards to the SoilGrids1.0 is the addition of uncertainty estimations. I suggest revising the tile as ". . .producing global soil information with spatial uncertainty.

Printer-friendly version

Discussion paper

Thank you for the suggestion, indeed this is a key point. We will change the title to: "SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty".

> The use of the cross validation procedure based on spatial stratification can guarantee a balanced spatial distribution within each validation fold. It is an improvement. However, in the model calibration, the same imbalance problem of spatial distribution of soil observations was totally neglected, which may lead to biased predictive models for the mapping and consequently poor performance in areas with very limited samples or even without samples.

This is an interesting point. We agree that large differences in sampling density may also affect the calibration and that it would be worthwhile to develop methods that take differences in sampling densities into account when calibrating a RF model. However, this was beyond the scope of the current work. Please note that it is far from obvious how the calibration should be modified, because for calibration the distribution of the points in feature space is more important than their distribution in geographic space. Note also the link to a comment of Reviewer 1, who mentioned that reducing the number of observations in densely sampled areas would negatively influence calibration (see comment Reviewer 1 to L. 334 and further). We addressed this highly relevant issue in the Discussion of the revised manuscript.

> The soil observations used in the work were from legacy soil surveys conducted at least three decades ago I think. But, some data of environmental covarites were derived from remote sensing observations of recent years, for example, the land use/cover data. The inconsistence in time may have significant influence on the predictions of easy-to-change soil properties such as SOC, N and PH.

Thank you for this observation, which particularly concerns soil properties that are readily affected by changes in land use or management practices, and ideally age of the observations should be taken into account (see Batjes et al. (2020), p. 303). However, for dynamic soil properties such as pH and soil organic matter content, we considered that the spatial variation will be much greater than the temporal variation. Not explicitly considering time, in this study, should not affect the predictions greatly. For the future, we will look at space-time mapping at global scale, elaborating on the work of Heuvelink et al. (2020) for Argentina. We included these important considerations in the Discussion (Section 3.1).

> I understand that the general way of modelbuilding in this work is still the same as SoilGrids1.0 version of Hengl et al (2017), simple 3D approach, which just takes observation (mid-point of horizon) depth as a covariate. I know this way is convenient in operation, one time modelbuilding can produce a soil property map of any depth. But there are some issues. One is that samples with high correlation at a same (profile) location may violate statistical modelling principles and lead to bias. Another is that taking depth as covariate may complicate model and make the model failed to focus its resources on capturing details of soil spatial variation but mainly the trend. This would lower the quality of the soil prediction overall. Some case studies also found that this way may tend to produce unrealistic soil predictions (Ma et al, 2021 and Nauman and Duniway, 2019).

We are aware of the limitations of using depth as covariate as described in the studies cited by the reviewer. We think using depth as covariate is a practical approach in global modelling. We absolutely agree that further research is needed to assess the implications, especially when using global legacy datasets with varying densities. We have added text to discuss this point at some length.

Printer-friendly version

Discussion paper

I suggest that the authors also used R2 to express the performance from cross validation because it is very commonly used in soil mapping community and would be convenient to compare with other soil prediction studies.

We understand the point made by the reviewer but we purposely refrained from reporting the R2 because this metric has caused a lot of confusion in the digital soil mapping literature. The problem is that R2 can refer both to the coefficient of determination of a linear regression between predicted and observed (i.e., the square of the Pearson correlation coefficient) as well as to the amount of variance explained by the model. The first evaluates how close predicted and observed are to a fitted straight line, while the second evaluates how close they are to the 1:1 line. It is the latter that we are after in statistical validation, and this is properly assessed using the (Nash-Sutcliffe) Model Efficiency Coefficient (MEC). We believe that the digital soil mapping community is increasingly aware of these issues and that it will not take long until the use of R2 becomes obsolete and the community uses the MEC instead.

The section of “Conclusions and future work” should be rewritten. I have difficulty to find a conclusion of this work. This section is not brief and clear. I feel that the discussion and points of this section is too general and not going towards the SoilGrids Project. Readers may get nothing from it. The first paragraph of this section (Lines395-396) repeated the aim that already stated in the introduction section and should be deleted. The second paragraph does not belong to this section and should be removed. It would be better to condense the fourth paragraph (Lines405-411) into say two sentences. The last paragraph (Lines424-426) stated a very general point which can be applicable to any digital soil mapping work. With or without it does not make sense. I suggest putting words on the new progress of SoilGrids2.0, limitations and what next version would look like.

The conclusion section has been re-written. Most of the content of the discussion for the submitted version was indeed part of the discussion and it has been moved out of the conclusion section into a discussion of limitations and future work.

> I suggest adding a map to show the time of the legacy soil survey projects of different countries if the time is very different among different regions. It is importance information for readers and data users to know what time/period of soil status the soil maps actually reflect.

Thank you for this suggestion. However, considering the large number of profiles we are working with, such a map would not be legible. Therefore, we have addressed this by adding statistics about the time periods the different point data were collected. Please note that we have also added text in the discussion section (3.1), where we advocate the development of a global model for space-time modelling as a follow up to the present research.

> Some expressions are confusing. The word "Quality-assessed" was frequently and somewhat widely used in the manuscript, which may have different specific meanings. For example, "producing quality-assessed soil informtion", "quality-assessed soil profile data", "Following data quality assessment and control", "Ultimately, upon final consistency checks, the quality-assessed and standardised data". This would make readers confused. Another similar problem is the use of "standardised", for example, "standardised soil profile data", does the "standardised" mean that all profile data were converted to the GSM depth intervals, as we saw ". . .six standard depths intervals" and ". . . standard depth interval for each soil property" in the title of Table2.

We used 'quality-assessed' to indicate the various stages of checking the heterogeneous source point data, including consistency checks on lat-lon and depth of horizon/layer, flagging of duplicate profiles, providing measures for geographic and attribute accuracy, as well as time stamps; we also checked for possible erroneous entries (i.e. min, max checks). Details are provided elsewhere (Batjes et al., 2020; Ribeiro et al., 2018).

Standardisation here refers to making the soil analytical data comparable using 'operational definitions' that describe key elements of each method (see Ribeiro et al. (2018)); it also includes standardisation of the units of measurements and geo-referencing of the point locations.

As such, standardisation as used in this paper does not refer to adhering to the GSM depth specifications, as this 'standarisation' is part of the mapping process itself.

We removed this source of possible confusion from the manuscript, and added some clarification to Section 2.1.

> Line 14: "up to date information on world soil resources. . .is required to address . . .". SoilGrids was based on legacy soil samples and the produced soil maps in fact reflect the status of soil conditions at least three decades ago, which is not up to date soil information. This work cannot respond to the demand.

We see your point and accordingly have rephrased this as follows: " ... based on the currently 'best-available' (shared) soil profile data."

> Lines 143-145: The time information of the remote sensing data is missing.

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.

Printer-friendly version

Discussion paper

Lines 168-185: it is not clear that which model was used in the RFE processing.

We have modified the manuscript to explicitly mention the method used in the RFE:

In a first step, the RFE procedure from `caret` was run independently on each set with default model hyper-parameters for RandomForests algorithm as implemented in the package `ranger` (i.e. ntree as 500 and mtry as the rounded square root of the number variables)

Lines 217-218: ". . .considered constant for the whole depth interval". this practice is different from that of SoilGrids1.0 (Hengl et al 2017) which generated prediction of a depth interval through calculating mathematical integral over depths. Is it better?

The reviewer is correct that in the previous version of SoilGrids (Hengl et al. 2017) predictions were made at the boundaries between GlobalSoilMap standard depth intervals. In this version of SoilGrids we went back to our original approach used in SoilGrids1km (Hengl et al. 2014), by predicting at the centre of the standard depth intervals and using these predictions as predictions of the interval average (i.e., by assuming that soil properties are constant within the interval). Because of this there was no need for numerical integration. We do not know which approach is better, but we do know that the approach that we now use again is much more commonly adopted in digital soil mapping. Both approaches have deficiencies because they do not handle the differences in vertical support of soil observations and predictions very well. A more solid approach is presented in Orton et al. (2016, Geoderma 262, 174-186), but this is much too demanding for high-resolution global soil mapping using machine learning.

Table 2: I do not understand this table, actually confused. I guess that the the depth intervals of the soil horizons data used in this work are not uniform

among the profiles and mostly not same as the GSM depth intervals. So, how to assign a sample of for example 20-50cm to a GSM depth interval, 15-30 or 30-60? Or maybe the authors standarised all horizons data into GSM depth intervals before modelling, then get the number of observations of each depth interval.

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. See also our reply to a comment by RC1.

Figure 5: one graph example is enough.

Figure 5 has been modified to show only one graph. The caption is now more informative.

Table 5: I also do not understand this table. As mentioned above, the horizon observations are not uniform in depth intervals. How did you compared with the predicted values at 2.5cm (0-5cm) to calculate the performance metrics MEC?

The text has been expanded to explain how the cross-validation took into account the differences in depth.

Line259: What means the “standarised data”?

As indicated in an earlier reply, "Standardisation here refers to making the soil analytical data comparable using ’operational definitions’ that describe key elements of each method (see Ribeiro et al. (2018)); it also includes standardisation of the units of measurements and geo-referencing of the point locations."

> Lines299-300: what do you want to express? large observations and co-variates lead to better predictive performance?

This was also pointed out by RC1. 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. We have added a brief explanation in the text to clarify why we point this out.

> Line315: evaluation=validation?

We have added a reference to the argument of Oreskes (1998) and Rossiter (2017) on this, also explained our preferred usage. We link to the common use of "validation" in these contexts, which we refer to as "numeric evaluation". We have checked the entire text for consistent usage of these terms.

> Figure6: please add a legend to the maps

Added.

> Line379: some representative papers may be useful for reference to illustrate your point about national soil mapping: Liu et al. 2020, High-resolution and three-dimensional mapping of soil texture of China, Geoderma, 2020, 361: 114061; Liu et al 2020, A soil colour map of China. Geoderma, 2020, 379: 114556

Thank you for the suggestion, indeed this fits into the list of papers we already cited for this point. Since no paper showing a regional map of China was included in that list, we have added the suggested reference.

References

Batjes, N. H., Ribeiro, E., and van Oostrum, A.: Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019), Earth System Science Data, 2020, 299–320, https://doi.org/10.5194/essd-12-299-2020, https://doi.org/10.5194/essd-12-299-2020, 2020.

Heuvelink, G. B. M., Angelini, M. E., Poggio, L., Bai, Z., Batjes, N. H., van den Bosch, R., Bossio, D., Estella, S., Lehmann, J., Olmedo, G. F., and Sanderman, J.: Machine learning in space and time for modelling soil organic carbon change, European Journal of Soil Science, n/a, https://doi.org/https://doi.org/10.1111/ejss.12998, 2020.

Ma, Y., Minasny, B., McBratney, A., Poggio, L., and Fajardo, M.: Predicting soil properties in 3D: Should depth be a covariate?, Geoderma, 383, 114 794, https://doi.org/https://doi.org/10.1016/j.geoderma.2020.114794, http://www.sciencedirect.com/science/article/pii/S0016706120325490, 2021.

Meyer, H. and Pebesma, E.: Predicting into Unknown Space? Estimating the Area of Applicability of Spatial Prediction Models, arXiv:2005.07939 [cs, stat], http://arxiv.org/abs/2005.07939, 2020.

Nauman, T. W. and Duniway, M. C.: Relative prediction intervals reveal larger uncertainty in 3D approaches to predictive digital soil mapping of soil properties with legacy data, Geoderma, 347, 170 – 184, https://doi.org/https://doi.org/10.1016/j.geoderma.2019.03.037, http://www.sciencedirect.com/science/article/pii/S0016706118319347, 2019.

Poggio, L. and Gimona, A.: Assimilation of optical and radar remote sensing data in 3D mapping of soil properties over large areas, Science of The Total Environment, 579, 1094 – 1110, https://doi.org/https://doi.org/10.1016/j.scitotenv.2016.11.078, http://www.sciencedirect.com/science/article/pii/S0048969716325177, 2017.

Ribeiro, E., Batjes, N., and Van Oostrum, A.: World Soil Information Service (WoSIS) - Towards the standardization and harmonization of world soil data. Procedures Manual 2018, Report ISRIC Report 2018/01, ISRIC - World Soil Information, http://dx.doi.org/10.17027/isric-wdcsoils.20180001, 2018.

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