EGUsphere, author comment AC2
https://doi.org/10.5194/egusphere-2022-246-AC2, 2022

Reply on RC2

Jonathan J. Maynard et al.

Author comment on "Accuracy of regional-to-global soil maps for on-farm decision-making: are soil maps "good enough"?" by Jonathan J. Maynard et al., EGUsphere, https://doi.org/10.5194/egusphere-2022-246-AC2, 2022

Response to anonymous reviewer (RC2)

Reviewer Comment:
"This manuscript compares soil maps (conventional and digital soil mapping) in terms of their prediction accuracies and their accuracies for soil suitability assessment for maize farms across Ghana. It aims to assess whether these soil maps could provide useful information for guiding farm management. This is an important question, and the manuscript is well written and provides a useful analysis, finding limitations in the ability of the soil maps to predict constrained soils. I have a few suggestions and queries below, but overall would think it could be published with some revisions."

"Section 2.5 describes field-support validation of predictions and I like the fact that this is being done, but am not convinced by the details of how it has been applied. The described process seems to be (for each field) first inferring a field boundary, second extracting from the soil map the predictions from all pixels within the field boundary, and third averaging these predictions; the result is compared with the average of the three sampled soil data from the field. But given your field-support validation data here are the average of three known points, shouldn't your associated prediction be the average of the predictions (as extracted from the soil maps) at these three same points? This could still be referred to as 'field-support validation', but would be comparing like with like, rather than two slightly different definitions of 'field support'. (While I don't expect this to make a huge difference to your results, I think it would be a more justifiable approach.)

Author Response:
The reviewer poses an important question about how to conduct a field-support validation. The intent of validation at a field-support is to estimate and evaluate the average soil property values within a field. Thus, validation samples are collected that represent the field average, and the map accuracy is evaluated within the entire area contained within the field. Since soil maps have different spatial support (ISDA=30m, SoilGrids=250m), the number of predicted values intersected by a field will vary, but the goal is to obtain an average predicted value for each field. If we were to only take the predicted value at each validation point, the average estimated field value would not reflect the average value predicted by the map. The figure below (Fig.1) illustrates this, which shows a map of predicted clay percentage (0-10 cm) from the ISDA map in southern Ghana. The three small black circles are validation point that have a clay surface texture (0-10 cm). If we were to only take the pixels that intersected our validation points, we would conclude that

the map and the validation points agreed at the field-scale. Whereas, when we take the average value for the field (i.e., all pixels within field, all points within field), we see that the map predictions are not in agreement with the validation data. For additional clarification on conducting validation at different spatial support, we direct the reviewer to the following papers:
https://bsssjournals.onlinelibrary.wiley.com/doi/full/10.1111/sum.12694,
http://dx.doi.org/10.1016/j.geoderma.2014.11.026.

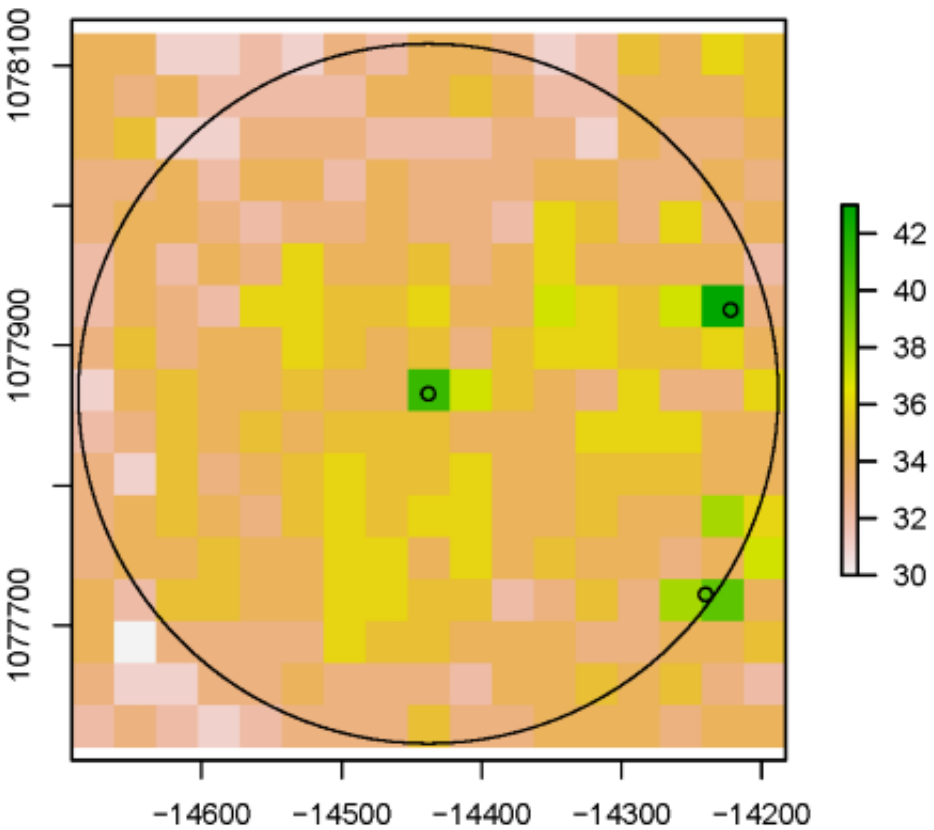

Figure 1. ISDA map of predicted clay percentage (0-10 cm). Small black circles show location of validation sites with clay surface textures (0-10). For reference, a clay textural class requires >=40% clay.

Reviewer Comment:
"Line 237: Is the Balanced error rate the (unweighted) average of the Producer's accuracies for the classes? I'm not convinced/sure about what this measure is saying – if there is a class (e.g. silty clay/sandy clay) with very few data (presumably because it is quite rare in the region of interest), won't the balanced error rate place too much importance on predictions of this class? Perhaps add a sentence after line 237 to clarify why use the balanced error rate. (I could see why a BER as a weighted average of PA values could be useful, for instance where you know that for some reason some classes were over-represented in your data, and had some knowledge of what the real proportions of the classes across the region of interest should be – but to use equal weights in the averaging seems to me to be saying that your initial expected proportions were equal for

all classes, which seems unlikely to me.)"

Author Response:
The balanced error rate (BER) is the average of the errors in each property class. This includes both the errors of omission or false negative rate (FNR) (i.e., 100%-Producer's accuracy) and the error of commission or false positive rate (FPR) (i.e.,100%-User's Accuracy). BER is calculated as: ½(FPR +FNR). BER is useful because it accounts for both types error associated with Producer's and User's accuracy. When datasets have high class imbalance the overall error rate is not very informative. For example, in models that tend to overpredict the dominant class, the Producer's accuracy can be very high while the User's accuracy is low due to a high false positive rate. The BER will reflect this.

Reviewer Comment:
"I wonder if another test might be helpful to provide further context about the quality of predictions from the maps. With the dataset that has 3 points sampled in each field, you could test how accurate predictions would be if you used just one of the sampled points from the field as 'representative' of the field, and evaluated how accurate this was at predicting the two other points in the field (to evaluate how good management would be based on a single 'representative' sample from the field). If there is a lot of within-field variation, even this might give poor prediction accuracies. This might add further insight into what is written on line 562 – "…fail to meet the accuracy requirements…" – would the use of a sample from the field also fail?"

Author Response:
The reviewer poses an interesting idea and we agree that in-field variability is an important consideration for field sampling but this was beyond the scope of this project. This would be an interesting analysis for future work.

Reviewer Comment:
"Line 431: I can't see the numbers you are referring to here in Table 2, they seem to me to be a lot higher than this – from my interpretation of the numbers in Table 2, it seems that OA-adj was over 80% in many cases (eg iSDA, OA-adj, soil texture: 0.9). Also check the numbers on line 593, and check the numbers throughout."

Author Response:
Yes, these numbers were incorrectly reported. We have reviewed and corrected all numbers reported in the text.

Reviewer Comment:
"Regarding everything being predicted as 'no constraint' (eg by iSDA, Fig 8) – I think that predicting the more extreme values of soil constraints is always going to be difficult, and the 'expected value' as extracted from the soil map will rarely give the extreme values – I guess a bit like predicting rare events. A possible point for discussion is that you could use the map of predictions + its uncertainty to give (for each pixel/prediction location) a probability of constraint, and this could be a more appropriate way of informing management decisions. (Although this would very rarely be done in practice, could note that tools to implement this type of analysis of the soil map + its uncertainty map could be something worth looking at in future?)"

Author Response:
This is a very good point and the idea to use the prediction uncertainty is a good one. We agree that predicting the extremes can be challenging, especially when dealing with global soil maps. However, the risk or cost associated with incorrectly identifying these soil constraints can be extremely high at the smallholder scale. The financial costs of a false negative result (i.e., failure to detect constraint – Type 2 error) is often much higher than a false positive result (i.e., false detection of constraint – Type 1 error) for cash-

constrained farmers. The uncertainty information could help with this type of risk analysis. We agree that this type of analysis is worth looking at in the future and we have added this point to the discussion.