EGUsphere, author comment AC1
https://doi.org/10.5194/egusphere-2022-246-AC1, 2022

Reply on RC1

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-AC1, 2022

Response to review by Dr. Colby Brungard (RC1)

Reviewer Comment:
“I have two concerns with this analysis. My main concern with the research regards the (im)precision of the validation data. The methodology that was used to evaluate the soil map accuracy assumes that there is no uncertainty in the labels of the validation observations (e.g., soil texture or rock fragment classes), even though the authors recognized the error/uncertainty in field collected observation in section “4.2.1 Sources of field sampling error”. I am uncomfortable with this because I feel that not accounting for measurement error in the validation data has the potential to obfuscate the ‘true’ accuracy of the soil maps (either higher or lower than is presented).”

Author Response:
While Dr. Brungard makes a valid point here, we feel that this issue is not confined to field collected validation data. It is often assumed that laboratory measured soil properties have both high accuracy and precision, and therefore any inaccuracy or imprecision is ignored. However, a recent interlaboratory comparison found considerable variability in lab measured soil property values among labs. For example, 6th FSCC interlaboratory comparison found that percent clay values had a coefficient of variation of 32% among 50 participating laboratories (Cools and Vos, 2010). Furthermore, our work found significant variation in soil texture measurements among 3 different soil laboratories in Ghana, with a CV of 47% for clay based on 10 soils with contrasting textures (2020, unpublished data). Such high imprecision complicates the ability to identify a true or standard value that can be used to evaluate accuracy. Based on these interlaboratory comparisons, it would be incorrect to simply assume a single laboratory-based texture class measurement represents the ’Gold Standard’ for analysis of soil map accuracy. In fact, a recent study by Vos et al. (2016; 10.1016/j.geoderma.2015.12.022) that assessed the variance of field-based texture class
estimates and their precision in determining a soil's particle size distribution, found that between 57-72% of deviations between field and laboratory derived texture estimates was due to laboratory measurement uncertainty and the fact that only texture classes were estimated in the field and not mass fractions.

So how do we deal with multiple sources of uncertainty? One approach is to determine the minimum level of measurement precision needed to inform a particular outcome. While soil particle size mass fractions (high measurement precision) are often required for soil modeling, soil texture classes (low measurement precision) are generally sufficient for on-farm soil management. Thus, using soil texture classes lowers our level of measurement precision which in turn minimizes the different sources of uncertainty (e.g., lab, field). Our use of the GAEZ soil suitability ratings and classes further lowered the measurement precision, which in turn further minimized the inherent uncertainty associated with lab and field-based measurements.

Reviewer Comment:
"The authors do attempt to account for the accuracy of the validation observation class labels using the results of several authors (Line 479) to argue that since sandy soils are often the most accurately identified and since most of the soils in the study are sandy then the validation observations are probably okay. The generalization of soil classes into soil suitability classes also helps address this issue; however, using the numbers provided by the authors may suggest that this is not an entirely robust assumption. For example, Salley et al., indicates that field technicians (and I assume that field technicians, not trained soil scientists collected the validation observations) are only able to correctly identify the true textural class 41% of the time (a rather dismal number), but that this improves to 78% if the adjacent textural class is accounted for. This suggests that the accuracy of the validation observations might be expected to be between 41% and 78%."

Author Response:
We agree with Dr. Brungard that a 41% correct classification rate is rather dismal, but it is incorrect to assume that the 59% misclassification is solely attributable to field technician. We acknowledge that a large portion of this error is likely attributable to the field technician, but we must also acknowledge that the laboratory measurements also have inherent inaccuracy. Widening the classification criteria to also include adjacent textural classes greatly improved the accuracy reported by Salley et al., (2019; https://doi.org/10.2136/sssaj2018.04.0137), and we also include this type of comparison in our analysis. Our study goes a step further by further generalizing our classification criteria by using the GAEZ soil suitability ratings and classes. Although this type of analysis was not evaluated in Salley et al., one would expect a further increase in accuracy well above 78%, and thus serves as a more reliable metric for accuracy evaluation given the inherent uncertainty associated with each data type (i.e., field vs. lab).

Reviewer Comment:
"A more in-depth, if rather crude, analysis* however suggests that the true accuracy of the
class labels for the validation dataset is between 64% and 82%. Thus, if the true accuracy of the validation observations is 82% then even if the soil maps were 100% accurate (which they are not) then the maximum accuracy that the soil maps could achieve would be 82%. This suggests that the reported accuracy metrics are not robust making a clear accuracy assessment difficult. Because the uncertainty of class labels in the validation data is not accounted for in the analysis I feel that the methods are not appropriate for this type of analysis."

Author Response:
We agree with Dr. Brungard's analysis but wish to clarify two important points:
1. It is incorrect to assume that laboratory data is 100% accurate and that any difference between field and laboratory data is due to inaccuracy in field measurements. In the best-case scenario, field and laboratory data disagree in only 18% of cases, however based on previous work (including our own), some portion of this error is attributable to laboratory uncertainty. Previous studies have shown that certain soils, due to their minerology or chemical make-up, are not well-suited for laboratory particle-size analysis and field-based estimates may be considered more reliable than lab data (Landon, 1988). For example, with highly weathered oxide-rich tropical soils traditional laboratory techniques often underestimate clay content due to the soil's resistance to dispersion (Silva et al., 2015). These types of highly weathered and/or oxide rich soils (Plinthisols, Ferralsols, Acrisols)

are common throughout our study area.
2. Second, the calculated accuracy range of 64% -82% for validation observations is based on an exact matching criteria. By widening the classification criteria to include adjacent classes or by using the GAEZ soil suitability ratings, the accuracy range for the validation observations will likely increase above the 64-82% accuracy which we feel is an acceptable level of accuracy given the inherent uncertainty associated with either type of validation data.

Reviewer Comment:
"There are several methods that might be employed to address the imprecision of the validation observations such as: 1) selecting a subset of the validation observations and running these through a lab (e.g., hydrometer) to verify class labels or 2) Use a method that accounts for measurement uncertainty (though I will admit to not knowing of a method to do such a comparison without retreating to a statistics textbook)."
Author Response:
We appreciate these suggestions and agree that future work is needed to address issues relating to the uncertainty of validation data. But more work is needed to quantify the uncertainty of the validation data, both field and lab data. There needs to be greater recognition of the uncertainty associated with laboratory-based soil particle size analysis, which for certain soils can be quite high. Therefore, we feel that Dr. Brungard's first suggestion does not solve the issue of validation uncertainty. We agree with his second suggestion about using a statistical method that accounts for measurement uncertainty, and while some of the soil maps evaluated in this study contain measures of uncertainty that could be used in the validation procedure, this would also require a validation sample set with quantified uncertainty that would allow for the calculation of probability distribution functions for both the map and validation datasets. Our solution to the uncertainty problem was to base our evaluation on several sets of validation labels with increasing degrees of generality (i.e., soil texture class, soil texture groups [includes adjacent classes], GAEZ suitability). While the 'true' accuracy of each set of validation labels is unknow, we can infer (as Dr. Brungard did for the soil texture class labels) an estimate of accuracy based on previous studies. As we stated earlier, it is important to understand the level of precision required to address a particular outcome, and in the case of agricultural management of rainfed maize, the GAEZ suitability labels provide an appropriate level of precision and inferred accuracy. While the accuracy of validation labels based on soil texture classes may be too low (64% -82%) to reliably evaluate the map sources, we feel that the GAEZ labels are sufficiently accurate and corroborate the results from the other labels.
Based on Dr. Brungar's comments we have added the following text to the manuscript to clarify our stance on validation uncertainty and the use of field-based validation data:
"Furthermore, some soils due to their minerology or chemical make-up are not well-suited for laboratory particle-size analysis and field-based estimates may be considered more reliable than lab data (Landon, 1988). This is true for highly weathered oxide-rich tropical soils where traditional laboratory techniques often underestimate clay content due to the soil's resistance to dispersion (Silva et al., 2015). However, difficulties with lab-based clay estimation extend beyond oxide-rich tropical soils, as was shown in the 6th FSCC interlaboratory comparison which found that clay content was one of the most difficult properties to consistently measure, with a coefficient of variation (CV) of 32% among 50 participating laboratories (Cools and Vos, 2010). A recent interlaboratory comparison among three soil laboratories in Ghana revealed significant variation in soil texture measurements, with a CV of 47% for clay based on 10 soils with contrasting textures (2020, unpublished data)."
Cools, N. and Vos, B. De: 6 th FSCC Interlaboratory Comparison 2009 Further development and implementation of an EU-Level Forest Monitoring System (FutMon), Life + Regulation of the European Commission , in cooperation with the International Cooperative Programme on Assessment a, , 32(0), 2010.
Landon, J. R.: Towards a standard field assessment of soil texture for mineral soils, Soil

Surv. L. Eval., 8(3), 161–165, 1988.
Silva, J. H. S., Deenik, J. L., Yost, R. S., Bruland, G. L. and Crow, S. E.: Improving clay content measurement in oxidic and volcanic ash soils of Hawaii by increasing dispersant concentration and ultrasonic energy levels, Geoderma, 237, 211–223, doi:10.1016/j.geoderma.2014.09.008, 2015.

Reviewer Comment:
"I am also uncomfortable with the conclusions. The authors state "results from this study highlight the need for on-site verification technologies… that can constrain the… site-based soil map predictions". I generally agree with this, but feel that this conclusion focuses too much on such technologies. What about improving the soil maps? If soil maps were 100% accurate then they would be a very useful source of soil information and we wouldn't need on-site verification technologies. I realize that no soil map (or any map) will every be 100% accurate, but the authors should consider this in their conclusions. Also, what about training more soil scientists? A broader cadre of local professional soil scientists could provide such site-specific information and might be more familiar with local soils and issues."

Author Response:
We thank Dr. Brungard for this insightful comment and we fully agree with these additional suggestions for improving the utility of soil maps for farm scale management. We have revised our conclusions to include these valuable suggestions.

Technical Corrections:
1. Figure 5 and 5. A-d. I believe this would be much more helpful if these were to show the resampled depths instead of the original depths. Also, do i-n show M2F observations at the pedon scale or farm scale?
Response: We have updated this figure to show resampled depths for the original maps. Panels i-n show M2F observations at point-support. This has been added to the figure caption.

2. iSDAsoil soil maps are generated from a suite of ensemble models. My own application of ensemble modeling techniques suggests that they can miss the extreme values (although this observation was likely a result of the specific ensemble method that I tested). It might be more informative to use the predictions from Hengl et al 2015. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0125814 If the authors want to test ALL available maps for Ghana, then this would be good to include, but I leave this to the authors discression.
Response: This is an interesting finding regarding ensemble modeling and while we agree it would be interesting to test for any differences between SoilGrids v1 and v2, our goal was to test current soil map data sources likely to be used for soil decision making. Since SoilGrids v2 supersedes v1, we only included the most recent data.

3. Please explain the balanced error rate in more detail. Is this (UA + PA)/2 or some other equation? What does this mean? Please also clarify that the error rate seems to be an error metric with the users and producer's accuracy rates are accuracy metrics.
Response: The balanced error rate (BER) is the average of the errors in each property class which includes both the error 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: (FPR +FNR)/2. We have added additional details on the calculation of the BER and clarified that it is an error metric vs an accuracy metric.

4. The paragraphs from lines 255-265 seem duplicated. Please revise.
Response: We have revised this section.

5. Please put the discussion of SQ ratings and associated equations in their own paragraph as it is confusing as currently written.
Response: We have separated this section into its own paragraph.

6. Line 339: How do you know the delineation procedure captured 48% of the fields area if you do not have farm field maps? (the lack of such maps was stated earlier as the justification for the delineation proceedure).
Response: It is correct that we do not have delineated farm boundaries but we do have two pieces of information: (1) the total farm area, and (2) the locations of three points within the farm boundary. By creating a convex hull (in our case a triangle since we only have three points) and applying a narrow buffer (10m), we were able to delineate an area within each field. Based on the known area of each field we were then able to compare our delineated area to the total field area, which on average captured 48% of the fields area.

7. Line 509: ... of the point soil data...
Response: Correction was made.

8. Line 525: scale does not really translate well to discussing gridded variables. I think that the neighborhood size, or distance-over, is a more accurate description. Also, maybe a salient point to this discussion: the source of the geospatial source data (DEM's in particular) is important. A 30m DEM from contours will be much different than a 30m DEM derived from upscaled LiDAR.
Response: For gridded variables scale refers to both the spatial extent or neighborhood size and the grid resolution. We have clarified this point. We agree that the accuracy and precision of the source data is important but our main point here is that increasing spatial resolution (i.e., grid size) doesn't automatically produce a more accurate map due to scaling effects.

9. Please check the following sentences for grammatical errors: 48, 89, 91, 308, 331, 369, 370, 591
Response: Grammatical errors have been corrected.

10. Line 153 should be grouped with the preceding paragraph.
Response: We have incorporated this change.

11. Please check references. Lines 169 and 172 are missing, also Ritchie and Roser does not seem complete.
Response: References have been checked

12. Line 230, what about user's accuracy?
Response: We have corrected this omission.