Reply to Referee #1 on our manuscript "Oblique geographic coordinates as covariates for digital soil mapping"

5   We thank Referee #1 for the well-thought and qualified comments on our manuscript. In the following, we will address the referee's comments and describe the changes that they have occasioned in the manuscript.

COMMENT

The present paper is well written and structured. Moreover, the study aims to make a
10   contribution to the field of DSM by providing a novel methodological framework based on the usage of coordinates. This is something that can be considered as rather 'out of the box thinking', because most attention in the international literature goes either to the use of advanced geostatistical methods (e.g. capturing the spatial autocorrelation through kriging) or external drift fitting based on '(environmental) co-variates' or a combination of both. Hence,
15   the work certainly merit respect for its originality and the methodological framework seems to provide useful thoughts to be considered in future DSM-studies. However, I also see some shortcomings which should be addressed/considered in order to maximize its potential to be applied widely, and as such, I am looking forward receiving the authors replies on the associated comments and suggestions presented below.

20   REPLY

We thank the referee for seeing the value in our research. We will address the shortcomings listed by the referee in the following.

COMMENT

Major Comments: I believe that the main issue with this research is that it considers only one
25   rather small field characterized by 1 remarkable / specific spatial structure as regards the variation of SOM (i.e. one spot/area with clearly higher values) in order to test the validity of the present new methodology, whereas the authors claim that the method will be highly useful for mapping soil properties in larger areas. Hence, I believe that the present methodology requires further testing by considering larger areas (e.g. catchment-regional
30   scale) with more complex spatial patterns in SOM in order to prove the validity of the statements that have been made in this respect. Moreover, it would also be interesting to consider other key soil variables (besides SOM) to check whether the usages of oblique geographic coordinates as covariates could be seen as a universal DSM approach. In this context, I believe that using a national soil inventory database could be a good way forward. I
35   may understand that this might not be possible in this study, but I still believe that this should

1

be mentioned clearly (as a critical note) in the discussions (and maybe be picked up by the authors in future research).

REPLY

40 We agree with the referee that it is a shortcoming that we only test the method for one soil property in one area. We believe that it is a vital next step to test the method in other areas and for other soil properties. We already mention this issue in the conclusion (L323 – L325). However, we see that we could add further emphasis on the subject.

CHANGES

In order to increase emphasis on the necessity of testing OGC on more datasets, we will
45 rephrase L323 – L325:

"One of the main shortcomings of this study is the fact that we have only tested OGC and compared it to other methods for one soil property in one agricultural field. This means that our study could not fully assess the relative usefulness of the method. A vital next step is therefore to test OGC on a range of soil properties in a larger area. It will especially be
50 relevant to test it for large areas, as the spatial patterns of soil properties in large areas are typically more complex."

COMMENT

When I have a look at the performance of the different mapping methods (as presented in the Violin plots in figure 7), it seems to me that your new OGC (+AUX) method only results in
55 (very) small improvements as compared to some other (more commonly used) methods such as Kriging. Hence, I was wondering whether this improvement is statistically significant? And if this might still be the case when either (i) another field (characterized by a different spatial pattern), (ii) another soil variable or (iii) larger geographical extent are considered?

REPLY

60 As we state in L149 – L153 and the caption for Table 2, some of the differences in accuracy are statistically significant and some are not. We used the same 100 repeated training/test splits for all methods, and this allowed us to carry out pairwise t-tests between the accuracies of the methods. We then ranked the methods using the results of these t-tests. Methods that did not have statistically significant differences for a given metric received the same rank for
65 that metric, but methods with statistically significant accuracies received different ranks. For example, OGC + AUX and RFsp + AUX always received the same rank, as there were no statistically significant differences in the accuracies. Meanwhile, kriging always received a higher rank than AUX, because the differences in their accuracies were statistically significant.

70 Kriging, RFsp + AUX and OGC + AUX all received the highest rank for two out of three accuracy metrics. We therefore regard these three methods as equally accurate. We already

2

state in the manuscript that we regard these three methods as most accurate (L210). However, we see that we have not explicitly stated that we regard them as equally accurate.

Furthermore, we cannot know the results of t-test carried out for accuracies in other/larger areas and for other soil properties, as these results currently do not exist.

CHANGES

We will rephrase L210 to explicitly state that we regard the three methods as equally accurate:

"We therefore regard these three methods as the most accurate methods. Furthermore, we regard these three methods as equally accurate, as none of them were universally more accurate than the other two methods."

COMMENT

Minor Comments: I'm not too sure if it is entirely appropriate to use R2 as a measure to compare the different methods, because (i) a very high R2 value may also mean an 'overfit' and (ii) each method has it own degree of (model) complexity. Hence, I guess that it could be a good idea to take (also) another statistical measure into consideration that specifically aims to evaluate the methods' performance taking into account its complexity (in order to avoid overfitting)?

REPLY

We understand the referee's concern, as a very high R^2 on a training dataset can indicate overfitting of a model. However, we report R^2 for 25% holdout datasets not used in the models. Our R^2 values therefore indicate the predictive capabilities of the models rather than their fit on the training data. Furthermore, we are not aware of any measures of accuracy that account for complexity in Random Forest models. We are even less aware of any accuracy measures capable of comparing complexities of conceptually very different models, such as Random Forest and kriging. We think most readers will be aware that kriging is much simpler method than Random Forest. In fact, we explicitly state this in the manuscript (L295 – L296).

COMMENT

Figure 1 - Subpanel C: Showing hill shade is not enough to give the reader an insight into the topographical configuration of the field. Hence, I suggest adding contour lines.

REPLY

We thank the referee for this helpful comment. We agree that adding contour lines improves the visualization of the topography of the study area.

CHANGES

Due to the referee's comment, we have prepared a new version of Figure 1, where we have added 2 m contour lines. We will include this updated figure in the final version of the manuscript: