



EGUsphere, referee comment RC2
<https://doi.org/10.5194/egusphere-2022-583-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on egusphere-2022-583

Anonymous Referee #2

Referee comment on "Mapping soil micronutrient concentration at national-scale: an illustration of a decision process framework" by Christopher Chagumaira et al., EGU sphere, <https://doi.org/10.5194/egusphere-2022-583-RC2>, 2022

This manuscript uses a case study on mapping soil micronutrients in Malawi to illustrate a decision process framework that accounts for uncertainty. I was eager to read and review it because the subject matter is highly relevant and has my great interest, and also because several of the authors are big names in this specific domain of science. Initially I was quite happy with the paper (the Abstract and Introduction read fairly well) but as I read on I got more and more disappointed. The manuscript is sloppy and often vague and imprecise. The Introduction does not clearly define the objectives and many of the sections do not match well. For example, Sections 2 and 3 are very descriptive and often unclear, while parts of Sections 4 and 5 dive into advanced mathematical-statistical approaches for modelling spatial variation, while these sections are much less on incorporating uncertainty in decision making. To me the paper is too much of a mishmash of subjects that are not clearly linked. In the end I am left with a manuscript that does not show me clearly how to deal with uncertainty in decision making. I am afraid that it did not advance my understanding, rather the opposite.

Detailed comments

(L27-28) What do the authors mean by 'scales' when they write "inherent variation of soil at multiple scales"? Does 'scale' refer to the extent of the study area, the spatial resolution, or the spatial support? Authors also mention spatial measurement error as a source of prediction uncertainty, but has this uncertainty source been covered in this study? And what do you mean with "uncertainty arising from predictive factors in our spatial models"? Are these 'predictive factors' the covariates or something else? If these are the covariates, how do covariates lead to uncertainty? Perhaps it is the limited ability of covariates to explain the spatial variation of soil properties, but that is not clearly stated.

(L31) Here it writes "organized", while L34 writes "recognise". Decide between UK, US or Oxford spelling and use it consistently throughout the manuscript.

(L39) Again, what do you mean by "... soil variation occurs ... at multiple spatial scale in space". Also sloppy formulation, is not 'multiple spatial scale' by definition something that refers to space? For instance, there is no such thing as 'multiple spatial scale in time', or is there?

(L45) Confusing text because it suggests that past surveys is the opposite of a systematic grid, as if past surveys could not use systematic designs.

(L67-68) This makes no sense. Variable selection methods are also used by ML methods (e.g. Recursive Feature Elimination). Moreover, kriging with external drift also weights covariates 'appropriately'.

(L77-78) This comes closest to a sentence that states the objective(s) of this paper, but it is not clearly formulated. It would help the reader if the aims were clearly communicated at the end of the Introduction.

(Introduction) The Introduction pays much attention to soil mapping, geostatistics en machine learning for soil mapping, including sampling design optimisation (L23-L68), giving the text a flavour of a review paper, while this is meant to be a research paper. Much less attention is paid to a review of how prediction uncertainty is quantified and communicated (L69-L77), while this is the focus of this paper. I sense a disbalance.

(L84) I am not convinced that they always require reliable evidence. Sometimes it simply is not possible to generate reliable information and so one has to make do with information that is not very accurate. But it may still be better than no information. Our task then is to communicate the uncertainty associated with the information to users and even better explain them how uncertainty can be incorporated in decision making. Is this not what this manuscript is all about? Stating that stakeholders can only work with reliable evidence to act undermines the message of this paper (namely that stakeholders can also make use of uncertain information).

(L93) Sentence says that there are four questions but the list has five.

(L94, I2) This has a link with the spatial support at which information is needed (if a decision is made at farm level, then we need information at farm level, or perhaps not?).

Authors will know that the uncertainty associated with soil predictions is strongly affected by the spatial support, but none of this is included in the manuscript. To me this is unacceptable. They should have addressed how uncertainty changes under a change of support and their methods should account for that.

(L94, I4) Not clear to me why the "given uncertainties" is included and what it means. Do we get different outcomes from the decisions in case of uncertainties?

(L94, I5) What do you mean by "potential legacy value"? How relevant is this?

(L97) Cryptic sentence, I do not understand what "The state is the state of affairs which our soil information predicts" means. Surely this can be formulated more clearly.

(L103-104) This needs a better explanation. Some farmers may rather apply extra lime if they are uncertain, just to make sure that there is no yield loss.

(L106-115) I did not find this text very clear. Perhaps including a figure might help?

(Figure 1) I like this figure, although it could perhaps be rearranged a bit so that it is clearer that it is a flowchart from I1 to I4. I also did not understand the "Opportunity cost. Nutrient deficiency" description, did not understand the difference between "Unnecessary cost" and "Opportunity cost" and between "Moderate yield loss" and "Some yield loss". It is also not explained in the main text.

(Section 2.2) The first paragraph gives the impression that authors did not make a thorough literature study and came up with the division in three types themselves. There are no references to that type of literature, but surely there must be lots of literature on defining and classifying stakeholders. I searched for publications in Web of Science that had both the word "soil" and "stakeholder" in the title, and got already 35 hits (1752 hits if both terms must appear in the abstract).

(L129) But there are so many factors other than sample size that influence uncertainty. Why focus on this one only. Why are other factors not mentioned and reviewed?

(L130) Is it really the sponsors and users who make decisions about information? In fact, what do you mean by this? Is it not the surveyor or producer of information that makes decisions about the uncertainty? I would expect that a surveyor informs the user how large the uncertainty will be given a sample size, and how much it will decrease if sample

size is increased, so that a user can take an informed decision about trade-off between costs and accuracy. This is not clearly explained in L130-134.

(L135) I wonder how many of these questions are understandable to the stakeholder, I don't think they should be addressed by engaging with stakeholders. These are questions that the surveyor/modeller must answer, the results of which may be shown to stakeholders (by providing them with trade-offs between accuracy and costs).

(L135) What I really miss here is the most important first question of any survey: what is the goal? None of the listed questions can be answered without it. See the excellent book by De Gruijter et al.

(2006, <https://link.springer.com/article/10.1007/s11004-008-9147-7>), which, unlike this paper, takes a very structured and comprehensive approach on how to design a survey while accounting for uncertainty.

(L137, V1) Of course we can, if we have the data. But do we need it, if decisions are taken at much larger supports (field, farm, district)?

(L137, V2) Other than what?

(L138-139) Data and surveys are costly, but is not a survey a means to get data? So why treat them as two different entities? And of course we should take rational decisions, that is a *sine qua non*.

(L142) This does not help me much. What do you mean by an "actual decision on sampling"? What is the difference between R2 and R3? What is the difference between R4 and O6? Covariates for what?

(L145) I agree with this but then shed some light on how this is done effectively. Is that not what this paper aims to do? I don't find it addressed and explained.

(L146) What do you mean by "value of uncertain information"? There is a rich literature on "value of information" but that is not mentioned.

(L147) What do you mean with "can the acceptable uncertainty be quantified? What is the difference between "acceptable uncertainty" and "tolerable uncertainty", how are these defined? Provide references to these terms because I would hope and expect that you build on existing theories.

(L152-157) Difficult to follow without proper explanation. Strange that from a rather general text you move to a very specific and somewhat complicated case. Same applies to L159-161: what is offset correlation, what is robustness of the final map, what is the final map to start with, why do you need a variogram and why would the offset correlation (or is it the survey effort?) be sensitive to arbitrary variation of the origin of the survey grid. This is not understandable without proper explanation. If it is made understandable, I wonder how relevant such a very specific case is to the general problem of accounting for uncertainty in decision making.

(L162) We get a case study but the methodology has not adequately been explained.

(Figure 2) What is the difference between this figure and Figure 1? What does "Trail established" mean? It is a real pity that this manuscript has so many errors. Why didn't the authors carefully check it before submission, if only out of respect of the reviewers??

(Sections 4 and 5) Now we get the Materials and Methods section, which makes me wonder: what did we get the previous 245 lines? The manuscript is getting very long and I am afraid I am losing it. Perhaps it could have worked if authors had explained the structure of the work at the end of the Introduction, but they didn't. Sections 4.2 and 4.4 are mainly about advanced geostatistical modelling, Section 4.3 on random forest for spatial prediction, so where is the handling of uncertainty in decision making? Where is the link with Section 2? This is only very marginally addressed. The same holds for Section 5, this presents the outcomes of the geostatistical and machine learning analysis, but where is the connection with Section 2? Maybe Section 5.3 aims to do that, but again we get an advanced statistical text with a lot of jargon that has little connection with the decision process descriptions addressed in Section 2. Stakeholders will have a hard time understanding this all. I also did not understand why different mapping models have been applied and were compared (OK, REML E-BLUP, RF), Why detailed information about one of these methods (e.g. see Figure 4, but no details about another mapping model (i.e., random forests)? And then in Figure 5 we find that also indicator kriging is included, why not in Figure 3? Section 5.2 then explains the application of a random forest model, but results of this model were already presented in Section 5.1, does this make sense?

(L483-485) That is what this paper is about and what should have been demonstrated, but that did not happen.

(L485-487) I do not understand why these statements are made here and what they exactly mean. It is not connected with what was presented before in this section.

(L561-570) The Conclusion is poorly written, and many of the claims made here were not realised in this paper. There are quite a few platitudes, and a mishmash of messages. This paper lacks focus and has too many diverse approaches.

