Author Responses to Referee Comments:

We would like to thank the referees for their thorough review of our manuscript. We have revised the manuscript based on their suggestions and comments. We reply to each of the comments below. Our changes in the paper are in blue below, and the line numbers and sections refer to the revised manuscript:

Referee Comments 1 (RC1):

Referee Comment	Author Comment
1. I appreciate the effort that has gone into improving the contents of this manuscript, especially regarding the extensive structural and content changes based in part on my concerns. Unfortunately, while the basis of the study is interesting, most of my concerns remain unresolved or only partially addressed in the author response. In my previous evaluation, I suggested focusing on the human decision process in planning spatially explicit surveys, which is purportedly the focus of this paper. This required a thorough restructuring to understand the perspective of respondents, the scenarios they were presented with, and why only 26 people were selected or participated. However, the revised manuscript only mentions that respondents "self-identified" as "x-y-z" professionals, without independent verification. There was no attempt to explain the statistical power of the 26 people, its limitations, and how (un)representative the 26 people are for their representative professions that they seem to be speaking for. This is especially concerning given the small sample size and the overrepresentation of agronomists (18 out of 26). Considering that these respondents are practitioners in their various fields, it would have been useful to know their background information such as age, sex, and years of experience, and most importantly if they have any involvement in planning geospatial surveys amongst others. This missing context is crucial for	We appreciate your suggestions regarding the participant composition and representativeness. In this study, we aimed to capture perspectives from professionals with direct, self-reported experience relevant to geospatial survey planning, particularly focusing on individuals actively involved in agricultural and nutrition fields. The intention was not to generalize findings beyond this group, but rather to explore practical insights on decision-making processes from those currently involved in survey design. Regarding background and sample size (see responses 9 and 8 below) We appreciate this thorough reading of our paper. However, we do not agree that the primary focus is on the decision, rather it is on the means to communicate the information required for that decision. We agree that the actual decision process is complex and specific (and have addressed this elsewhere, e.g. Chagumaira et al 2022). We have tried to clarify this at various points in the abstract (L1—L21), introduction (L91—L100) and discussion (L427—L450), and have also made an edit to the title of the paper. We have addressed the issue of the size of this experiment in our response to RC3 (see responses 8 and 9 below). On the issue of `self-identification' this was not to define a population of stakeholders in terms of expertise. The population was defined as professionals in the listed fields who were engaged with Universities in the Gates-funded GeoNutrition

readers to understand why respondents would prefer one method over another.	project and the allied UK Research Council-funded Zim-GRTA project. They were identified by the country lead on the project. By virtue of contributing to this project (e.g. as extension officers in national extension services or national public health researchers or nutritionists) the participants were in the target population. The self-identification was simply to show how the different specialisms were distributed- we clarify this on L119—L123. We regarded the mathematical background and experience as essential additional information, but could not justify collecting personal information, such as gender, to the Research Ethics Committee. It must be emphasized again that the objective was not to elicit a sample density, but rather to examine the comprehensibility and usability of alternative measures of how uncertainty and sample effort are related.
2. It is particularly worrying that authors consider only studies involving “indistinguishable pills” could be blinded or randomized in an experiment, and that randomization or independence of samples is irrelevant in sociopsychological studies such as this one. This kind of social experiment has strong theoretical basis in economics of “choice experiments”, and I expected the authors to ground their work in such an exemplary well-established theory. Indeed, this is quite surprising that such a theoretical foundation is missing for a study that is so heavy on frequentist statistics such as this one. From my assessment, most changes in the revised manuscript are largely editorial and do not address the core content I previously suggested.	Our comment on " indistinguishable pills " was specifically in response to the original suggestion that double blinding was an appropriate approach. It is surely self-evident that an experiment in which a researcher explains a set of methods to communicate information to a participant, who then applies these methods, and records their impression, cannot be "double-blinded". The researcher cannot explain a method without knowing what it is! Even the most experienced of agronomists could not apply and rank a set of methods, without knowing what they are. Regarding randomization, we gave the explanations of the methods in an order which we believed would allow participants to build their understanding of geostatistical predictions and their uncertainty, starting with the simplest concepts, then introducing the more complex. Randomizing this order would only cause confusion. We accept that randomizing the order of tasks would be ideal. However, we felt that it was appropriate to let the participants approach the task of applying the different methods as they found comfortable. They could approach the tasks in the order they preferred. For an exercise necessarily undertaken online, we felt that this was the most appropriate way to collect useful

	responses. We discuss this as a matter for further work (see L426—L440).
3. Moreover, it seems arbitrary to treat respondents from four different countries (Ethiopia, Malawi, Zambia, and the UK) on two different continents (Europe and Africa) as having comparable experiences and backgrounds. One wonders what soil scientists in the UK have to do with a study designed to understand the planning of a geospatial survey using the context of a province in Malawi. Under what circumstance are soil scientist in the UK grouped together with agronomist and public health professionals from four African countries? It is insufficient to claim that respondents are motivated by some MND project goals, as if they are a homogenous group of people. The paper inaccurately assumes that by eliciting information from these 26 self-styled professionals provides broader insights into the planning process of geospatial surveys. This is clearly not the case. Not only are the thoroughly explained geostatistical methods and the survey approach lacking novelty, but the study also wrongly assumes a unidimensional simplification of multidimensional complex human decision-making process.	The participants were all engaged in a common project, bringing their expertise and experience to bear. We agree that there could be differences in the *decision process* regarding an intervention in UK and African contexts in so far as the losses arising from decisions, which are suboptimal because of the uncertainty in the information on which they are based, would differ. However, our focus is not on the decision process as a whole, but on the accessibility of the different forms of information which we have presented see introduction (L91—L100), methods (L119—L123) and discussion (L426—L450). We think that this will reflect the focus of education, training and experience, but see no reason to expect a systematic difference between UK and Africa-based professionals. The reviewer makes some rather sweeping statements about the novelty of the statistical measures of uncertainty that we trialled. The kriging variance was proposed as a statistic for sample design in the early 1980s, there are relatively few hard examples of its application in the way proposed. The related confidence interval has also been proposed, but the particular way we have expressed it for interpretation is new and based on our previous experience (Chagumaira et al 2021, https://gc.copernicus.org/articles/4/245/2021/) using prediction intervals as measures for the assessment of uncertainty in spatial information. The offset correlation was first proposed by one of us (RML), but this is the first attempt to use it with a stakeholder group. The joint probability we use in this study has not been proposed for this purpose and is the first attempt to use a generalized uncertainty measure which accounts for the location parameter of the variable as well as its spatial dependence (see L95—L100). For this reason, we think that the comparative assessment of how these different methods is received and applied is a useful contribution. With regards to location of respondents please see response 1 (above).

4. What is most disturbing is the superficial confirmatory arguments throughout the paper. For instance, we are told that respondents chose Off-set correlation because, in an unrelated study by Hsee (1998), people prefer bounded attributes over absolute ones. Yet, this same explanation does not apply to the joint (conditional) probabilities with similar boundaries, due to their "probabilities". Such superficial empirical studies directly contradict the rigorous 'Popperian' falsification advocated in modern scientific inquiry. In my previous review, I advocated looking into the reason behind one person (representing 5% of the 26 people) who selected 100 km grid space as the optimal, and the authors' response is "… the analysis tells us that it is potentially misleading to find explanation", as if the "analysis" is absolute and final. This rather bizarre answer is another indication of how the study made no attempt to find alternate explanations to their confirmatory responses.	Our aim was to find out to what extent are these methods usable by stakeholders from different discipline, and to highlight general trends in survey planning preferences among respondents, while recognizing individual variability. The reference to Hsee (1998) was intended as a supplementary point rather than a definitive explanation. It started from the observation that the offset correlation was generally preferred. The offset correlation takes values in [0,1] (it cannot be negative) which corresponds to a range from "zero information" to "perfect information". This clearly relates to Hsee's conclusion. The conditional probability does not have such a simple interpretation. We noted (see line 406 of the original paper) that is also bounded on the interval [0,1], which might explain why it was ranked highly even though it was generally misinterpreted. We think that the fact this "relation to reference" effect appears to explain the high ranking of offset correlation and conditional probability, even though the latter was clearly widely misunderstood, is an important finding, related to our key objective. We have added the following sentence from L374: A method might be regarded as easy to interpret, because of its form, even when it is not (in this case a large value of the probability indicated that there was no spatial information in the map to make its predictions better than the overall mean). The way in which the conditional probability is specified depends on the problem for which the spatial information is used (a threshold), and whether a large or small value is preferable depends on its exact formulation. This was explained to participants, but clearly it was less accessible than the offset correlation. We make this point at L95—L100 in the revised paper.

	We do not think that the question of falsifiability is relevant here. We are not testing hypotheses about decision making. We are making a practical evaluation of the extent to which each method succeeds in communicating how uncertainty depends on the method used.
5. After reading the entire paper, I still feel that the four proposed methods are not directly and necessarily comparable as presented in this study. The underlining explanations of the various geostatistical methods do not indicate the need to choose one over the other. Thus, study participants may be subjected to a hypothetical situation that is neither necessary nor realistic. The revised manuscript explains: (a) Prediction interval is based on the kriging variance (which is estimated from the empirical variance taken from an earlier survey data?); (b) Joint (conditional) probability assumes a location requires some intervention (what sort of intervention?) based on the kriging variance, indicating that a prediction does not correspond to a particular threshold set a priori (obviously more information is needed to understand this); (c) Implicit loss function, which is not based on the kriging variance but on a hypothetical loss (but it is unclear whether this loss is economic or information loss or both) for making a spatial decision that is correct or an error (where it is undefined what constitutes correct and erroneous decision); (d) Offset correlation is based on the consistency of spatial information at two hypothetical grid spacing, not on the kriging variance. Yet, it remains unexplained why these four methods, requiring different data inputs and different underlying assumptions, need to be compared as this study did. None of the data and resource constraints used to build scenarios for participants make it reasonable to choose one method over the other. The methods may be useful under different conditions and applicable situations, which may not necessarily overlap requiring	The methods are not all directly comparable, and this is clear. However, they are all derived from a common statistical model of the variable, and so are mutually consistent. Our view is that all the methods could be used in the proposed scenario but provide different information. The information provided in each case was explained to the participants. What we learn from their responses is how far they regard themselves as being informed by the information about how uncertainty and sample effort are related. We clarify this point in the revised paper at L95—L100.

users to use one instead of the other, for planning a geostatistical survey. Given the lack of basic information on the comparability of the methods, the missing key information of the respondents, and the lack of theoretical grounding, along with the fact that almost all my concerns are adequately unresolved, I cannot endorse the publication of this manuscript. Additionally, the paper needs a careful editing for grammar and sentence clarity as many sentences were impossible to parse.	

Reply on RC3

Referee Comment	Author Comment
6. This paper had a number of strengths, including a well-written and robust introduction section and literature review. It was a pleasure to read an introduction that so eloquently explained the wider research project, implications of that research, and why you are trying to find the most effective means of communicating uncertain data. Additionally, you have highlighted an area of study not yet discussed in the scientific literature. I have a few general recommendations in addition to some specific edits:	Thank you for the opportunity to revise our manuscript. We would like to thank the referee for their time and for the constructive comments they have provided
7. This paper is very long, I recommend consolidating information and monitoring the paper for conciseness. I realise it is hard to balance thoroughly explaining the background and considering length, but as a reader, I became lost and fatigued trying to understand all the communication methods you tested. It seems this could be shortened considerably. Perhaps one way of doing so is to show the examples in section 1.3 with a very brief explanation of what they mean, for example, a kriging method to determine probabilities of different pH.	We thank the referee for their feedback. We have revised the paper and made some statements more concise.
8. There is little to no discussion about the the impact of sample size on your results. There are very few people involved in your study, which impacts the robustness of your results. It's likely that there are statistically significant findings in your method, but there are too few people in the sample to suss that out. I recommend extending this study to additional people, including entire departments from the universities involved in the study, governmental bodies involved in the study, etc. For such a small sample size, the focus of the results should be on more qualitative data, for example the feedback sessions rather than statistical analysis of results. I'm particularly concerned about the results from the tests you ran comparing the groups divided by	We thank the referee for this comment, and we make the following change from L413: All the information users recruited in this study were employed in public sector institutes (e.g., universities, civil organisations, research, and extension) and had experience in their respective fields in an SSA setting. We had no basis for a power analysis to identify a sample size for this activity. Given the exploratory nature of this research, our primary aim was to capture insights from as many relevant participants as possible within each institution. As a result, our major consideration was recruiting individuals willing to participate and with experience in their respective institutions. We therefore attempted to recruit the entire set of suitable

specialism - 8 people in one group doesn't make a robust statistical analysis in a survey.	respondents in each country. We recognize that the small sample size limits the generalizability of statistical findings. While this study provides insights into participant perspectives by specialism, the lack of demographic information—such as gender, age, location, and years of experience—limits the depth of analysis. These characteristics may impact responses; for example, different age groups or experience levels might prioritize certain issues differently. Future studies should consider including these demographic details to explore how such factors influence perspectives, thus enhancing the robustness of the findings and allowing for subgroup analysis. For this reason, we have interpreted results cautiously and have also incorporated qualitative insights from participants to provide a richer context for understanding these early findings. Moving forward, we plan to include an initial power analysis and possibly extend the study through broader collaborations to enhance robustness.
9. The only information you give on participants is the number of participants in each specialism. Please include information such as gender, age, location, years of experience in their fields, etc. Add to the discussion how these characteristics may impact your survey results	We have added information on the composition of the participants in the appendix and the following text from L285: There was reasonably even spread in terms of the location of our participants, see Figure B1 (Appendix B). About 54% of the participants were constantly using statistics/mathematics within their job role. Only a few participants were educated to the level of certificate/diploma (8%). We also have edited the following in the discussion from L419: While this study provides insights into participant perspectives by specialism, the lack of demographic information—such as gender, age, location, and years of experience—limits the depth of analysis. These characteristics may impact responses; for example, different age groups or experience levels might prioritize certain issues differently. Future studies should consider including these demographic details to explore how such factors influence perspectives, thus enhancing the robustness of the

	findings and allowing for subgroup analysis. For this reason, we have interpreted results cautiously and have also incorporated qualitative insights from participants to provide a richer context for understanding these early findings. Moving forward, we plan to include an initial power analysis and possibly extend the study through broader collaborations to enhance robustness.
Line 176: I'd like a sentence or two on what Se is and what a lack of Se in the diet does to a person. You mention the average requirement of Se in adult women, but that leads the uninitiated reader unclear on why they should care about Se concentration	We thank the reviewer for pointing this out, and we have added the following text from L184: Selenium is an essential micronutrient with critical roles in human health, and lack of it can cause thyroid disfunction, and suppressed immune response (Fairweather-Tait et al 2011).
Lines 381-384: From my memory, this is the first time the feedback session was mentioned. This should be discussed at the beginning of the methodology section and more focus should be given to this part of the study as qualitative feedback is more robust with a small sample size.	We thank the referee noticing this and we have added the following from L160: We had a feedback session to allow the participants to seek clarification on the presented methods.
Lines 398-401: The statement of "it is not clear how to select an appropriate effect size" is unwarranted. There are a lot of peer reviewed articles involving surveys and expert elicitation,	Thank you for this observation and we have addressed this comment, see response 8