Review for Geoscience Communication gc-2020-42 (https://doi.org/10.5194/gc-2020-42)

This is a well-written manuscript on an interesting topic. My main concern is that Sections 2.3 and 3 go in very much detail on the statistical analysis. This part is difficult for an audience that has a not so strong background in statistics and it distracts the reader from the main topic of the paper: how can we communicate uncertainty about spatial predictions effectively? I strongly recommend to move large parts of these sections, including quite a few of the tables, to the Supplementary Information. Instead, more attention could be paid to what we learn from the experiment conducted on communicating uncertainty (i.e., Table 12, Figures 7 and 8). Further, a more thorough comparison should be made with findings on spatial uncertainty communication and visualisation from the cartography and geo-information literature (I added a few entry citations at the end of this review).

I also think the experiment could have been conducted in a better way and that some basic mistakes were made in preparing the posters. These and some other points are worked out in the detailed comments below. I do not require that the experiments are redone but recommendations how to do better in future could be included in the Discussion and/or Conclusion.

I should add that I did not thoroughly check the statistical methodology part, but my impression is that this is solid work, as could be expected from the author list (in particular the last author has a strong reputation in high-level statistical analysis).

Detailed comments

(L36) Not in all kriging algorithms is the prediction a linear combination of the data.

(L39, L41, L130, etc.) Authors use the term ‘confidence interval’, but technically this should be ‘prediction interval’. There is a principal difference between the two, for example see https://en.wikipedia.org/wiki/Confidence_and_prediction_bands.

(L46, L110-114, L138) There is no need to use indicator kriging to compute exceedance probabilities. By invoking the normal distribution assumption for kriging prediction errors (which authors do, se L38), these exceedance probabilities can be easily derived from the kriging prediction and kriging variance. They will be more accurate than those obtained using indicator kriging.

(L61-65) I may be opening a box of Pandora, but authors will know that the uncertainty in the mapped concentrations of micronutrients in grain are heavily influenced by the support of the observations and predictions (i.e., the area or volume over which observations and predictions are made). Authors do not apply a change of support so the predictions and associated uncertainty refer to the support of the observations. Is this appropriate? What was it? This is not explained in L82-96: there is a lot of attention for the spatial sampling design but we learn nothing about how the field sampling was done. Were these point samples or bulk/composite samples? This is of key importance when addressing uncertainty.

(L89) was --> were.

(L103) This implies that predictions need to be back-transformed. How was this done (note that a naive back-transform returns the median, not the mean)? Information about the back-transform should be added.

(Eq. 1, L123) Here it should be upper case Z instead of lower case z, while in L132 and L134 it should be lower case z instead of upper case Z.

(L129, L340, Figure S3) Poster 3 should have shown the kriging standard deviation instead of the kriging variance. The kriging variance has different measurement units (the square of microgram per kilogram) and one cannot expect decision makers to account for this. Poster 3 also does not list the measurement units of the kriging variance. Moreover, the numbers are extremely small (around 1) and are almost certainly incorrect.

(Section 2.1.4, Figures S2 and S4) I doubt that computing the probability that the true value exceeds or lies below a threshold quantifies the uncertainty of predictions. For example, if the threshold is 38, the kriging prediction is 55 and the kriging standard deviation 8 then the probability of exceeding the threshold is extremely large (suggesting very small uncertainty, category "virtually certain"), while a kriging prediction of 36 with standard deviation 3 leads to large uncertainty (we end up in the category "about as likely as not"). But 8 is larger than 3, so can we maintain that the uncertainty of the predictions is quantified? These complications should have been addressed.

(L174) were --> where; where --> were.

(L186) Visiting posters in randomised order does not avoid carry-over effects, it only makes sure that the effects cancel out over a larger group. Perhaps rephrase this sentence to make this clear. Note also that instead of randomising it would have been better to have a deterministically determined sequence that guarantees that all posters occur in a completely balanced order.

(L207) Symbol $o_{i,j}$ not defined in the main text.

(L225) Two times "between the".

(L229, L230, L235, L296, L301) "was conducted", "participants are drawn", "This gives us", "there was", "There is". Please check entire manuscript on correct use of present and past tense.

(L277) less --> fewer.

(L324) a there is --> there is a.

(L334-335) Can and do you explain why the p-values were so different between Ethiopia and Malawi?

(Figure S1) Poster 1 has some important deficiencies. First, the mean has a continuous legend while the lower and upper limits have discrete units. This affects the map (discrete colour jumps in the limit maps). Second, all three maps should have had the same colour legend. For an example, see Figure 7 in https://onlinelibrary.wiley.com/doi/full/10.1111/ejss.12998.

Some publications on spatial uncertainty visualisation and communications that may be used as entry to a comprehensive literature search:

Aerts, J.C..J.H., K.C. Clarke and A.D. Keuper (2003), Testing popular visualization techniques for representing model uncertainty. Cartography and Geographic Information Science 30, 249-261.

Beven, K. R. Lamb, D. Leedal and N. Hunter (2015). Communicating uncertainty in flood inundation mapping: a case study. International Journal of River Basin Management 13, 285-295.

Johnson, C.R. and A.ER. Sanderson (2003), A next step: Visualizing errors and uncertainty. IEEE Computer Graphics and Applications 23, 6-10.

Kinkeldey, C., A.M. MacEachren and J. Schiewe (2014). How to Assess Visual Communication of Uncertainty? A Systematic Review of Geospatial Uncertainty Visualisation User Studies. Cartographic Journal 51, 372-386.

Kinkeldey, C., A.M. MacEachren, M. Riveiro and J. Schiewe (2017). Evaluating the effect of visually represented geodata uncertainty on decision-making: systematic review, lessons learned, and recommendations. Cartography and Geographic Information Science 44, 1-21.

Kunz, M., A. Gret-Regamey and L. Hurni (2011), Visualization of uncertainty in natural hazards assessments using an interactive cartographic information system. Natural Hazards 59, 1735-1751.