Response to Referee Comments for gc-2020-42

We would like to thank the referee for the opportunity to revise our manuscript. We have revised the manuscript based these suggestions and the changes are shown in the tracked changed version of the manuscript. Our edits in the paper are in blue below, and the line numbers refer to the ATC version of the revision.

Referee 1

Referee Comment	Author Response
(L39, Eq. 1) I maintain that the lower case z should be an upper case Z. Here, the random function model is invoked and the kriging variance refers to the variance of the kriging prediction error Z(x_0)-Z~(x_0). I am surprised that authors (in particular the last author) maintain their earlier view that it should be E[{z(x_0)-Z~(x_0)}^2]. Authors refer to Webster and Oliver (2007) to support their claim, but please check Eq. 8.2 in their book, which clearly writes E[{Z(x_0)-Z~(x_0)}^2], as it should be. Note that in case of E[{z(x_0)-Z~(x_0)}^2] we have that z(x_0) is a deterministic constant and so we effectively get the variance of Z~(x_0), which is very different from the variance of Z(x_0)-Z~(x_0). For example, suppose x_0 is a measurement location. In that case we would get Z~(x_0)=Z(x_o) and hence E[{Z(x_0)-Z~(x_0)}^2]=0, while E[{z(x_0)-Z~(x_0)}^2]=Var(Z(x_0)). In line 155 it should be z~(x_0) instead of Z~(x_0), because here you refer to the cross-validation prediction errors, which are deterministic values, not random variables.	The suggest change has been made on L132 to L133, Equation 1.
I maintain that authors should in Poster 3 better have presented the kriging standard deviation instead of the kriging variance. From their response I now understand that the kriging variance is that of the	We do accept the severe limitations of the kriging variance as a means for communicating uncertainty, particularly here where a transformation is necessary. We are not

log-transformed selenium content, while the kriging prediction is the back-transformed (hence median) selenium content. This explains the extremely low values. Now I wonder, does it make sense to present a kriging prediction map of the back-transformed variable and a kriging variance map of the transformed variable? How can we ever expect users to make sense of that and grasp the uncertainty? Note that the spatial pattern of the variance map is strongly influenced by whether it is done for the log-transformed or back-transformed variable. Authors justify showing a kriging variance map of the log-transformed selenium because "we cannot back-transform" the variance. Well, why not use Eq. 8.40 in the Webster and Oliver book? Admittedly this refers to simple kriging, which is not the same as ordinary kriging, but the difference will be small. All in all quite unsatisfactory how uncertainty was communicated in Poster 3: not only authors show the kriging variance instead of the kriging standard deviation, but they also mix up log-transformed and back-transformed variables. As before, I am not requesting that the questionnaires are redone, but authors should mention the weaknesses of their approach in the Discussion. The text that they now include in the revision (L148-L150, L429-431), is a good step in this direction, but they might add that they could have done better."	convinced that kriging standard error would solve the problem in the case of log-normal variables, because the units are still on the log-scale, and the prediction interval is an obvious way to turn the kriging variance into an interpretable measure on the original scales of measurement. Neither are we happy with the idea of using the simple kriging variance as the referee suggests, because this would be anticonservative. We included kriging variance among the methods of communication because, as a standard raw output from the kriging equations which does indeed characterize local prediction uncertainty, it is commonly presented in map form as a measure of the reliability of predictions. We therefore add the following text from L141 to L157 to emphasize that these limitations were clear a priori.: Because the kriging variance is a direct output of kriging algorithms, it is common to see it mapped alongside kriging predictions and referred to as a measure of local prediction uncertainty (e.g. Holmes et al., 2007; Goovaerts, 2014; Hatvani et al., 2021). However, the interpretation of the kriging variance may be challenging, particularly for a non-specialist user of spatial information. One could take its square root, and present it as a kriging standard error with the same units as the target variable. However, the interpretation of the raw standard error can clearly be helped by rescaling it to a prediction interval, and we consider this option in the next section. The interpretation of the kriging variance is particularly difficult in the case of a variable which must be transformed prior to analysis. The kriging variance cannot be back-

	transformed to the original units (except for simple kriging). In this setting then, the kriging variance can serve as little more than a general uncertainty index, indicating in general where uncertainty is large, and where it is small. However, such generalized indices have been developed for 3-D geological information to serve the needs of engineering stakeholders (e.g Lelliott et al., 2009; Lark et al., 2014). For this reason, and because of the longstanding use of kriging variance as an uncertainty measure (see above), we included it as a measure of uncertainty in this experiment. One poster showed a map of the conditional median of Se concentration in grain (Section 2.1.1), with a map of kriging variance on the transformed units (see Table 1, Fig S3). The referee has an important point and in this context the kriging variance can serve as little more than a general "uncertainty index", and is therefore unlikely to be useful, and that the responses we received confirmed this. In cases where transformation is not an issue the kriging standard error, on the original units of measurement, may be more interpretable to the end user, but it remains an abstract quantity. Rescaling it to a confidence interval, presented either by the limits, or by its width, is likely to be more useful, although our results suggest that the communication of confidence intervals requires more attention. Although the kriging variance is a valid statistic it has very little value as a means for communicating uncertainty for a general audience. Therefore, we have Following on from this, we expanded the text on L388 to L401 to include the suggestions made by the referee.

	“Although the kriging variance is a valid statistic, in this context it has very little value as a means for communicating uncertainty for a general audience. That is particularly true in this case, where the kriging variance must remain on transformed units, and so serves as little more than a general ``uncertainty index". This was clear a priori, and is confirmed by the responses we received. Our findings here cannot, therefore, be regarded as definitive, and a similar experiment for variables which do not require transformation would be necessary in further research. In such cases one could also include the kriging standard error as an uncertainty measure, to assess (i) whether the fact it is presented in the units of the target variable makes it preferable to kriging variance and (ii) whether it is regarded as less-interpretable than its rescaled form as a prediction interval. That said, our results do show that the communication of prediction intervals requires more attention. These considerations aside, kriging variances, standard error and prediction intervals must be interpreted by the user along with other information (for example, is the predicted value close to the threshold or substantially different from it) in order to make a judgement at a particular location. Our results do show that probability measured, tied directly to the interpretative task, are clearer to the user than general measures of uncertainty.” Goovaerts, P. 2014. Geostatistics: a common link between medical geography, mathematical geology, and

medical geology. Journal of the Southern African Institute of Mining and Metallurgy, 114: 605–612

Hatvani, I.G., Szatmári, G., Kern, Z., Erdélyi, D., Vreča, P., Kanduč, T., Czuppon, G., Lojen, S., Kohán , B. 2021. Geostatistical evaluation of the design of the precipitation stable isotope monitoring network for Slovenia and Hungary. Environment International, 146, 106263.

Holmes, K.W., Van Niel, K.P., Kendrick, G.A. Radford, B. 2007. Probabilistic large-area mapping of seagrass species distributions. Aquatic Conservation: Marine and Freshwater Ecosystems, 17, 385 – 407.

Lark, R.M., Mathers, S.J., Marchant, A., Hulbert, A. 2014. An index to represent lateral variation of the confidence of experts in a 3-D geological model. Proceedings of the Geologists Association, 125, 267–278

Lelliott, M.R., Cave, M.R., Wealthall, G.P., 2009. A structured approach to the measurement of uncertainty in 3D geological models. Quarterly Journal of Engineering Geology and Hydrogeology. 42, 95–105.