We would like to thank the reviewer for the careful and thorough reading of this manuscript and for the helpful suggestions, which certainly will improve the quality of this manuscript. Our response follows (the reviewer's comments are in *blue italics*).

The paper "Incorporating interpretation uncertainties from deterministic 3D hydrostratigraphic models in groundwater models" addresses the importance of characterizing uncertainties of both the hydrostratigraphic model and the model parameters of the groundwater model. The topic is well presented and of high importance. Therefore, I recommend publication after the following points are addressed.

MAJOR CORRECTIONS

General Remarks:

Comment 1: *At first, I found it challenging to understand the procedure of the performed uncertainty quantification. For example that both the hydrostratigraphic and model parameter uncertainties are investigated. This should be clarified in several parts of the paper. For instance on p. 3 in l.71-76, three scenarios are mentioned. However, it should be clarified that not a single realization but multiple realizations are run per scenario.*
Reply 1: To clarify the methodology, we will expand Section 3 and explain the full workflow in more details.

Comment 2: *There exists extensive literature about how to deal with uncertainties, especially in the field of geological modeling (e.g., Wellmann and Caumon, 2018), which needs to be added to the introduction to provide a broader perspective on this topic.*
Reply 2: We will broaden the introduction and relate our methodology to other geological modelling approaches. The suggested reference is highly relevant and a very enjoyable read. We purposedly avoided this topic as it was already covered in Madsen et al. 2022, but we acknowledge, as also pointed out by the other two reviewers, that we need to rewrite the current manuscript to be more stand-alone in terms of generating the hydrostratigraphic realizations.

Comment 3: *I agree with the previous reviewers that the information provided about the low-frequency and manual interpretation model is insufficient and needs to be extended.*
Reply 3: See reply to Reviewer #1.

MINOR CORRECTIONS

Further Remarks:

Comment 4: *p.1 l.14: The authors talk about "the qualitative and subjective nature of uncertainty". In general, one distinguishes between epistemic and aleatoric uncertainties. While the statement is true for epistemic uncertainties it is not true for aleatoric uncertainties. So, the statement needs to be specified by explaining which type of uncertainties are addressed in the paper.*
Reply 4: In this study, we aim to characterize the aleatory part of the uncertainty associated with interpreting a hydrostratigraphic model. To avoid confusion, we will delete "subjective" from the sentence in the revision.

Comment 5: *p.2 l.42: A manuscript that is in preparation is cited. Please either publish that manuscript as a preprint and cite this preprint or use a different reference since the current reference is not available to the reader.*

Reply 5: The mentioned manuscript has since been published as a preprint and the reference will therefore be updated in the revised manuscript.

Comment 6: *p.2 l.42-43: Clarify which type of uncertainties the paper addresses (see also the first comment under further remarks)*
Reply 6: This paper addresses uncertainties related to the perceived uncertainty of the geologist' while producing the deterministic interpretation model. We will try to make this clearer in the text.

Comment 7: *p. 5 Figure 1: It would be helpful to denote the profile lines with a,b, and c according to Figure 2. Such that these two figures can be better set in relation to each other.*
Reply 7: Thanks for suggestion, will do.

Comment 8: *p. 5 l. 111: "The synthetic well field does exist in the real world … ". Should the formulation not be "The synthetic well field does not exist in the real world"?*
Reply 8: Thank you for pointing out the mistake. We will fix it in the revision.

Comment 9: *p. 10 l. 183-184: Why were 50 realizations chosen? Has a convergence test been performed?*
Reply 9: For computational reasons. The current setup scales badly because the number of model runs is a multiplication of both hydrological parameter realizations (200) and hydrostratigraphic models (50). According to the results obtained in Madsen et al. 2022, the minimum number of realizations to fairly represents the entropy of the hydrostratigraphy is 50. Thus, 50 realizations were chosen. However, more would have been better as we do acknowledge that in fairness it would probably be reasonable to have the same number of realizations of hydrostratigraphy as the number of hydrological parameter realizations. How to computationally optimize such a system is ongoing research in other related research projects.

Comment 10: *p. 10 l. 189: Where are the uncertainties of the medium scenario listed?*
Reply 10: We will make a comprehensive table with all the standard deviations used in all three scenarios in the revised version of the manuscript.

Comment 11: *p. 11 Section 3.2.2: Provide the exact description and definitions of the boundary conditions and governing equations. It is not sufficient to list only the used packages.*
We agree that the description of the boundary conditions should be improved, see Reply 12 to Reviewer #2.

Comment 12: *p. 11 l. 215: Why was a random sampling strategy chosen and not a quasi-random strategy such as the Latin Hypercube Sampling (LHS) method? The LHS would have the advantage of better sampling the parameter space with few samples and avoiding the clustering of sample points as it often occurs for the random sampling method.*
Reply 12: In the manuscript, we have stated that random sampling has been applied, but we have actually applied a Latin Hypecube Sampling. The parameter values for the realizations are sampled from the prior parameter distributions using the Latin hypercube approach implemented using LHS class from the open-source Python framework pyDOE. The Latin hypercube designs obtained from pyDOE are transformed to uniform and log-uniform distributions using the values in Table 1 by applying the classes uniform and log-uniform, respectively, from the open-source Python. Thank you for pointing out this mistake. We will fix the mistake in the revision.

Comment 13: *p. 11 l. 217: Why was a uniform prior chosen while all other considerations so far targeted normal distributions?*

Reply 13: For the parameter values in the groundwater model, all values within the range are considered equally likely and a uniform distribution have therefore been applied. For the interpretation points, the interpreted value is considered most likely, while values above or below the interpreted value with equal distance are considered equally likely and a normal distribution have therefore been applied.

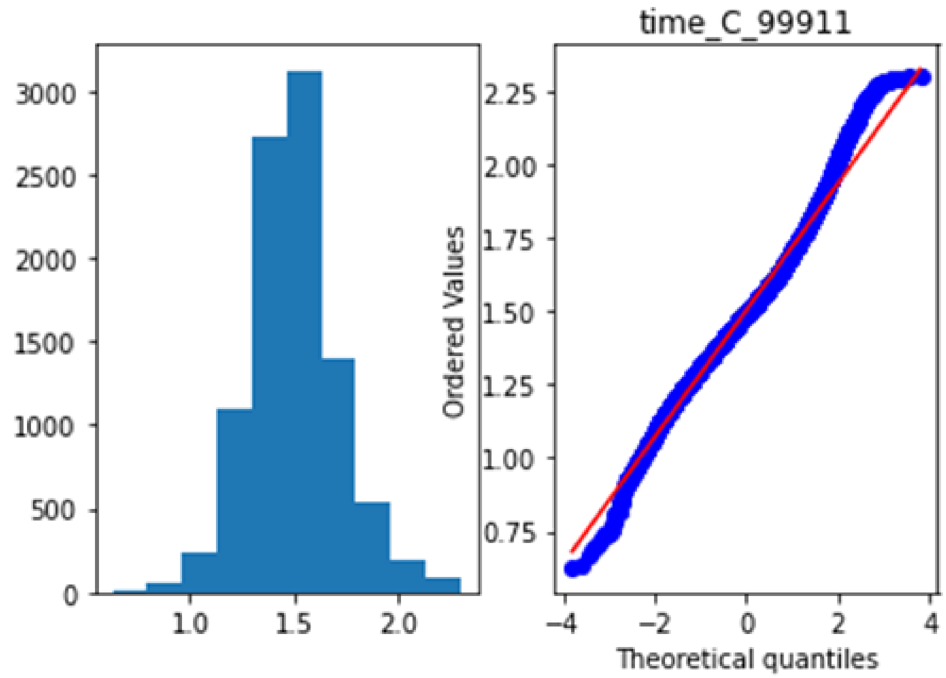
Comment 14: *p. 11 l. 220: Provide details on how it was determined that the parameters are insensitive. Which type of analysis was used to get to this conclusion?*
Reply 14: A local sensitivity analysis was carried out in the Manual Interpretation model to evaluate the sensitivities of the parameters.

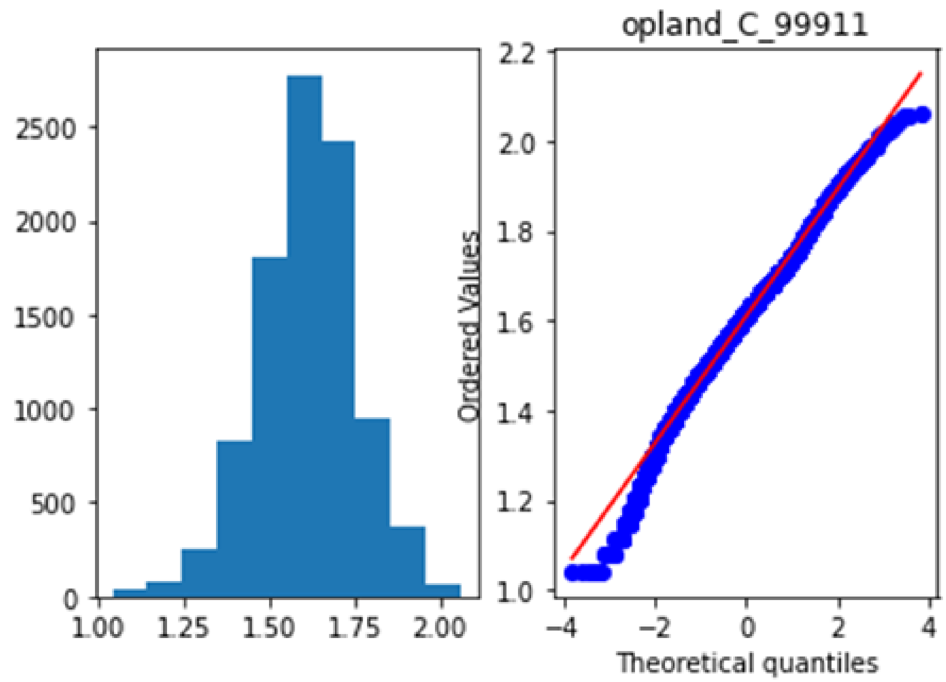
Comment 15: *p. 15 l. 294-296: The reasons why the solutions are not converging should be listed. Especially if the non-convergence is related to specific parameter ranges this might have a significant impact on the interpretation. Why is a trend of decreasing convergence problems observed with an increase in uncertainties? Would one not expect it to be the other way around?*
Reply 15: The non-convergence is not related to specific parameter ranges but rather to the individual model grids of the realizations. To improve the discussion on convergence, we will add the following sentences to section 4.2 in the revision:

'The convergence rate is likely influenced by the model grids that are unique for each realization as it follows the layer elevations. The model grid is thereby influenced by the smoothing factor of the hydrostratigraphic model (Figure 3). The low smoothing factor of the Low uncertainty scenario allows larger changes in layer elevations than the high smoothing factor in the High uncertainty scenario. In areas where the layers are thin this may result in lack of lateral continuity between adjacent cells, which causes an inability to simulate flow between cells in the same layer'.

Comment 16: *p. 15 l. 298-299: It should be explained if and when why the analysis is still representative if in one case 46 % of the realizations are discarded and in the other two scenarios only 6 % or 1 %.*
Reply 16: We agree that the results are not representative when 46 % of the realizations have been discarded. We have therefore indicated the results of this scenario in a parenthesis and with a dashed line in Figure 7. We will add a sentence to the caption of Figure 7 and line 403 to emphasize this in the revision.

Comment 17: *p. 16 l. 320: The standard deviation is only a valid measure for normal distributions. Have, for instance, q-q plots been generated to show that the data follows a normal distribution?*
Reply 17: Thank you for pointing this out. Q-q plots have been generated and a few examples are shown below. As illustrated, deviation is seen around the edges of the distribution. However, considering that these areas have fewer data points, we acknowledge and tolerate this deviation.


time_C_99911
Ordered Values
Theoretical quantiles


opland_C_99911
Ordered Values
Theoretical quantiles