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*).

General comment: *Review of "Incorporating interpretation uncertainties from deterministic 3D hydrostratigraphic models in groundwater models" by Enemark et al.*

When regional groundwater models are developed, an important step is to build a conceptual hydrostratigraphic model based on the geological information at hand. Conceptual hydrostratigraphic models are based on mapping the 3D juxtaposition of geological layers and translating these to aquifers and aquitards, which are subsequently populated with hydraulic parameters (conductivities, transmissivities, storage coefficients) and used to schematize the 3D-makeup of a groundwater flow and/or transport model. The mapping of geological layers is preferably done by expert geologists that combine their conceptual knowledge of the depositional or structural geological environment with in-situ borehole descriptions, outcrop information and geophysical data (e.g. gamma logs, EM measurements etc). However, since there is much room for interpretation, no two geologists will provide the same conceptual hydrostratigraphic model.

In this paper, the authors use a recently developed method to assess this "interpretation uncertainty" in hydrostratigraphic models to assess how the uncertainty about the layer boundaries between hydrostratigraphic propagates to the uncertainty in groundwater model outcomes. They compare this degree of uncertainty with the uncertainty that accrues from unknown hydraulic parameters (a more common analysis). Apart from demonstrating the method in an uncertainty analysis (focused on capture zone size and median travel time), the authors also show that the schematization uncertainty is important of little in-situ data are available and if the layers to be identified and mapped are thin.

This is a valuable paper that presents a nice approach that is worth being be picked up by the groundwater modelling community in order to extent their toolbox of approaches in uncertainty assessment.

I think this paper deserves being published in HESS subject to resolving the following issues.

Reply general comment: Thank you for the overall positive assessment of our study. The main points of concern raised by the referee will be addressed below in the corresponding specific detail comments.

MAJOR CORRECTIONS

Comment 1: *The Low-Frequency model is insufficiently explained in the paper. It may well be based on work by Madsen et al, but this paper needs be readable on its own. Particularly:*

1.1. How is the manual interpretation model constructed? Are the smooth lines between the interpretation points in Figure 2 actual kriged values? Was this a 2D-kriging per surface? What semivariogram was used? How does one make sure that the boundaries between layers do not cross or do cross in case of a presumed erosive surface? And how is this resolved? Is there a manul postprocessing?

Reply 1: We will clarify this in the text. In short, the smooth lines shown on the figure is the layers that have been gridded from the interpretation points, using kriging. The semivariograms are adjusted for each layer surface during the geological modelling. The layers are adjusted using a layer adjustment tool in the modelling software (Geoscene3D), where the modeler defines the order of the layers.

1.2. The simulation of vertical perturbations at the interpretation points is based on categories and then the standard deviation per category. A table with standard deviations should be given per case and per category.

Reply 1.2: We like the idea of a table showing the used standard deviations. However, we would like to split it into three separate tables as 3 scenarios x 14 layers x 4 categories results in a very large table with 168 entries. These three tables will represent the three different certainty cases. These tables will still contain 56 entries each. Thus, we would like to add these tables as supplementary material.

1.3. Please be more specific about the nature of the LF model? How is it fitted? What are its equations? Are they smoothing splines? Or moving averages of a zero-error interpretation model interpolated with kriging between the randomly perturbed interpretation points? Or is it kriging with uncertain data? Perhaps a step-by-step procedure description would be helpful.

Reply 1.3: The LF-model is obtained by linear interpolation between interpretation points and then applying a smoothing kernel in a sliding window on the interpolated grid. The kernel width varies spatially to account for areas that contain steep transitions in the elevation. This is done by calculating the variance amongst the interpretation points within the window. If the variance is high, then the modeler has tried to convey large changes in the elevation and the smoothing becomes low to allow the LF-model to follow these structures. In the other case, a low variance of elevation amongst the points within the sliding window leads to more smoothing of the LF-model. The procedure is explained in-depth in Madsen et al. 2022, but all reviewers rightfully point out that this should be explained in more detail in the current manuscript. It was left out in the original draft to highlight the aspect of doing hydrological modelling on many hydrogeological model realizations rather than the generation of the hydrostratigraphy itself. Since this is clearly a weak point of the current manuscript as pointed out by all three reviewers, we will add some lines in the manuscript to summarize the steps taken to create the LF-model.

1.4. How does the degree of smoothing takes account of spatially varying degrees of uncertainty? Is the level of smoothing spatially varying as well?

Reply 1.4: As explained above, the procedure for obtaining the LF-model takes spatial variability into account in terms of the curvature of the surface boundary. However, in the current setup level of smoothing is not affected by the uncertainty of the points. The uncertainty is instead utilized in the small-scale variability.

1.5. It seems that with high uncertainty, the smoothing is more intense. But how does the LF model distinguish between actual locally higher spatial variability (while the data are certain and based on detailed borelogs) and spatial uncertainty?

Reply 1.5: In continuation of the previous reply, uncertainty is not related to smoothing in the LF-model. If so, this is purely accidental or that highly uncertain points maybe are interpreted more flat lying than in places where there is more certain information available. This would create more smoothing due to the low variance within the sliding window. This would probably make sense from an interpreter's point-of-view. In the revised version of the manuscript, we will address this apparent correlation between highly uncertain points and the variance in elevation between them.

Comment 2: *Section 3.2.3: here the 200 behavioural hydraulic parameter sets are selected with GLUE and assuming the manual interpretation model to be true. How is it guaranteed that these parameter sets are still behavioural for the other simulated realizations of layering? I understand that all combinations cannot be evaluated, but a few random hydrostratigraphies could be checked whether the 200 parameter sets are still close enough to be called behavioural?*

Reply 2: We agree that it cannot be guaranteed that the behavioral parameters in one model is behavioral in another model. The "behavioralness" of the realizations are illustrated in Figure 6. Realizations outside the red lines are non-behavioral. It is shown that almost 90 % of the realizations (between the 5th ad 95th percentile) are behavioral for the Low and Medium scenario while it is a bit less for the High uncertainty scenario. We will add a comment on this in the manuscript.

Comment 3: *The conclusion that with thick layers and lower uncertainty about the layer boundaries is small compared to the uncertainty from hydraulic parameters: To what extent is this conclusion dependent on model approach. How were the heads and fluxes simulated in these models: a quasi-3D aquifer-aquitard schematization or with a full 3D voxel model? In the first case, vertical fluxes in aquifers are ignored and this may underestimate the impacts of the thickness of a layer, particularly near wells. In this case it is also understandable that thickness (one order of magnitude variation) has much less impact than conductivity (with multiple orders of magnitude variation).*

Reply 3: The model we applied was a full 3D model in which the vertical discretization follows the hydrostratigraphic units. Vertical fluxes are considered. We realize that the statement in the conclusions may be a bit unclear and we will elaborate the argument in the revised version.

Comment 4: *I urge the authors to make the datasets (schematization, hard data, interpretation points and manual interpretation model) available.*

Reply 4: We agree that this is a good idea. Currently all geophysical data used for the manual modelling is available in the national Geophysical database of Denmark (GERDA) and all borehole information is available in the Danish borehole database (Jupiter). They can be accessed here: https://eng.geus.dk/products-services-facilities/data-and-maps/national-geophysical-database-gerda and https://eng.geus.dk/products-services-facilities/data-and-maps/national-well-database-jupiter. What is not currently available is the 3 sets of hydrostratigraphic models and the interpretation points used to generate them. It will however be possible to introduce both on the GEUS dataverse and be publicly available afterwards. We will do this for the forthcoming version of the manuscript and provide references.

MINOR CORRECTIONS

There are some small remarks and suggestions for improvements that I have put in the pdf attached.

Comment 5: *L28: something like a model cannot be uncertain; change to "subject to uncertainty".*
Reply 5: Thank you for the suggestion. We will fix it in the revision.

Comment 6: *L37: random is not the right word. unstructured would be better.*
Reply 6: Thank you for pointing out this mistake. We will fix it in the revision.

Comment 7: *L40: perhaps call this "interpretation"*
Reply 7: Thank you for the suggestion. We will fix it in the revision.

Comment 8: *L43: change to: propagate this to the uncertainty of the results of large-scale groundwater models.*
Reply 8: Thank you for the suggestion. We will fix it in the revision.

Comment 9: *L70: Delete "particles".*
Reply 9: Thank you for pointing out this mistake. We will fix it in the revision.

Comment 10: *L76: which parameters? Please provide some examples. Hydraulic conductivity? Storage coefficient?*
Reply 10: Thank you for the suggestion. We will provide examples in the revision.

Comment 11: *L93: First: Are the smooth lines between the interpretation points actual kriged values? And second: how does one make sure that boundaries do not cross or do cross in case of an errosive surface? And how is this resolved?*
Reply 11: The geologist provides a set of interpretation points to the geologic interpretation software GeoScene and the surfaces on the grid in between points are kriged or interpolated using some properties that can usually be set manually. Then for boundaries that crosscut there is functionalities that resolve these issues by the geologist setting a list of "erosive rules" to determine which layer's elevation should be adapted for both layers in the zone of overlap. The software then runs through all layers and post-processes each to make sure that there are no overlaps in the final 3D model. This practice was also used for each of the hydrostratigraphic realizations.

Comment 12: *L112: delete "s" in represents.*
Reply 12: Thank you for pointing out this mistake. We will fix it in the revision.

Comment 13: *L166: Can you be more sp[specific about the nature of the LF model? How is it fitted? What are its equations? And how does the simulation algorithm then work? 1)Simulate elevations first; 2) then apply the smoothing? But should the smoothing not mean that residuals are also more or less random? No spatial correlation?;3) how does the smoothing distinguish between uncertainty and large spatial variability of layering.*
Reply 13: See reply 1.3.

Comment 14: *L224: How is it guaranteed that these parameter sets are still behavioural for the other simulated realizations of layering?*
Reply 14: See reply 2.

Comment 15: *L247: How is the entropy calculated?*
Reply 15: Say $f(m_i)$ represents the probability of outcome i out om N possible outcomes, then the entropy H, calculated with a log-base of *N*, is:

$$H = \sum_{i=1}^{N} f(m_i)\, log_N(f(m_i))$$

H will then be a number between 0 and 1, where 0 implies perfect knowledge (one outcome is certain), while H=1 will imply maximum entropy (maximum uncertainty), i.e. that all outcomes are equally probable, $f(m_i) = N^{-1}$.

Comment 16: *L332: How were the heads and fluxes simulated: quasi-3D or 3D. In the first case, vertical fluxes in aquifers are ignored and this may underestimate the impacts of the thickness of a layer. In this case it is understandable that thickness (linear scale) has much less impact than conductivity (order of magnitude scale).*

Reply 16: See reply 3.

Comment 17: *L393: This is a very nice way of representing this.*
Reply 17: We appreciate the positive feedback.