Response to Reviewer #4 comments

We thank Dr. Xin He for his valuable comments and suggestions. They will certainly improve the manuscript. In the following, we provide the response to the individual comments. The original review is quoted in *italics*, while our response is given in **bold** font.

The study aims to establish a workflow to carry out Multi-Point Statistics modeling for a testing area in Denmark using alternative 3D training images and various conditioning strategies. The research topic is of high relevance to those who work with hydrogeological modeling. The manuscript is well written with accurate language, rational methodology and convincing results. I recommend the manuscript is accepted for publication with minor revision. However, there are several details to be considered which are listed as follows:

As stated in the abstract, the introduction and the conclusion sections, one of the most important steps of the workflow is to develop 3D TIs in an iterative way. However, this part is only briefly mentioned in the method section and not at all mentioned in the result section. I am curious about how the TIs are evolved gradually with feedback information after each step of adjustment, namely from the initial TI to the final TI.
Additionally, is Fig 9 showing the initial or the finally TI?

We definitely see the reviewer's point and, in the new version of the manuscript, we will add further details regarding the development of the TI, from the initial guess in Fig. 9a, to the final result in Fig. 9b. In particular, we will stress the fact that: (i) the first TI simply consisted in a portion of the adjacent, pre-existing, geological model (the Tønder model); (ii) then, this initial attempt has been iteratively and manually adjusted based on the output of the associated unconstrained simulation; (iii) this iterative process ended when the final unconstrained realization was found satisfactory in terms of its ability to mimic the geological features we expect in the Miocene across the study area.

In the introduction section, I would suggest to add a few sentences indicating the main objectives of the study.

In the revised version of the manuscript, we will follow the reviewer's suggestion.

Lin 210-211. The moving window for calculating the borehole uncertainty in the vertical direction is 20 m. Meanwhile, in Fig 7(b), the thickness of the Miocene layer is about 150 m, which in principle corresponds to 7 to 8 intervals in each borehole at maximum. However, as far as I can count, there are usually more than 7 color blocks in each borehole. Am I mistaken for something?

20 m is simply the width of the moving window. This means that the information about the categories (the lithologies) is averaged across a 20 m wide interval. This does not necessarily imply that only 7 or 8 samples remain after the application of the moving window. Actually, the size of sampling interval is unchanged, and the only minor modification on this respect consists in a loss of a certain amount of samplings at the top and bottom of the boreholes due to the fact that, in our specific implementation, we considered a window with always the same width.

Fig 7, when interpolating the borehole uncertainty, are the borehole data outside the model domain being considered, both horizontally and vertically? If not, would there be extrapolation instead of interpolation towards the edges of the model domain?

We do not interpolate the uncertainty of the borehole. Instead, we krig the sand probability of the portions of the boreholes lying within the Miocene. So, if this was the question of reviewer, generally, towards the edges of the model domain, extrapolation does occur.

In the results section, L259-266, there are two TIs being tested, one clearly has more layers than the other. Do these two TIs have any relation to the iterative approach the authors try to present in the study? Or is it a separate issue here? Moreover, it says the second TI is chosen because it is closer to what has been presented in Kristensen et al., 2015. So maybe it is better to describe very briefly what is in the Kristensen's study, and why that one is used as benchmark.

The two TIs are respectively the first and final test along the iterative TI development process. In particular, the second TI (Fig. 9b) was selected to run all the subsequent conditioned simulations as the associated unconditioned realization (and not the TI itself) was fund able to mimic the geological structures characterizing the Miocene in the study area. The characteristics we expect for the Miocene structures are described in Kristensen et al., 2015 and discussed in the manuscript in the dedicated section "4 Establishing framework-model constraints" (and in the associated Fig.s 2, 3, 4, 5).