Author’s response

Date: 05/02-2021

Dear editor

I am writing this cover letter to present our resubmission of the manuscript: hess-2020-444. We have received some very constructive and helpful comments that have greatly improved on the old manuscript. Please do not hesitate to give our compliments in this regard.

Based on our answers in the initial discussion and the reviewer’s comments we have updated the manuscript. Since the initial submission more wells containing geochemical information (from Mapfield) have been analyzed and added to the pre-existing pool of data. This has meant that the TIs for redox have been remade and the MPS modelling consequently has been redone. This was also done based on the suggestions from reviewer 1. We have also addressed the underrepresentation of the soft data in the final realizations by converting a small fraction into hard data and the issue of inconsistencies between prior and conditional data. We hope that these major changes address the issues raised by the reviewers to a satisfactory degree.

First, we respond to the two reviewers point by point, following on from our initial response in the discussion phase. Then we present a summary list of the major changes to the manuscript, which is followed by the changes made to the figures and tables.

We enclose a version of the manuscript with track changes to help showcase where changes are made.

All authors have approved the submission of the enclosed manuscript.

Kind Regards
Ph.D. Rasmus Bødker Madsen
Geological Survey of Greenland and Denmark
C.F. Møllers allé 8
8000 Aarhus C, Denmark

Response to specific questions related to reviewers' comments

Reviewer 1:

The graphics are of good quality and well chosen. However, given the small size and high complexity of the models, the color differences in the geologic models are difficult to interpret. The captions often need to be improved; they need clearer descriptions of what is actually shown in each figure.

We have made a new colormap more in line with color choices usually used to represent geological models. We have tried to ensure that color differences are clear in the new color map. We have also worked on improving the captions and legends. For instance, for 3D plots the legend now shows voxels instead of only colors. Furthermore, we have added arrows in the figures pointing towards relevant parts of the simulation results that are mentioned in the text. The viewing angle is also changed as to give a better overview of the models.

This is more than just a conceptual problem. The MPS algorithms use the training image to guide the location of textures, but every texture has a non-zero probability of occurring anywhere in the model domain. This is not problematic when the facies are all generated within one, single depositional environment. It is problematic when the facies are from multiple depositional environments (and from multiple ice events).

MPS uses training images as analogs of the spatial variability to generate realizations of possible models of the subsurface. The conditional data (wells, geophysics) guide the location of the architecture. We have not been clear enough in delivering the modeling aim and the chosen method in the first version of the manuscript. In the revised version we have tried to emphasize this and our modeling aim/contribution to science in the introduction.

I noted early in my comments that the presentation quality of this manuscript is not ideal. It is possible that I have misunderstood the geologic setting and the approach taken to configure the MPS in this study. However, the description of the geologic deposits does not provide enough clarity to understand the geologic history of the model domain, or the consequent distribution of stratigraphic units and sediment textures. The authors also do not acknowledge this important geologic constraint to the successful application of MPS, nor do they provide discussions on their rationale for using MPS in this setting or on the design targets for texture proportion and zonation. This prevents a clear understanding of what the geologic history is known to be, and how this modeling approach is being used to reliably model the sediment while using that knowledge as a necessary constraint.

The geological elements are supposed to capture major depositional trends in the study area. Within each geological element we convey the spatial variability through a TI. These TIs in essence contain the geological choices (and rules) for the given domain. We have altered the section which describe the training images, which should hopefully clarify this point further in the revised version.

More importantly, the authors suggest that they could have made the solutions better fit the training images, but stopped with the simulations so they could present the method as a viable option. This needs to be fixed prior to submitting a methods manuscript for publication. Until an author can demonstrate clearly and objectively that they can reliably meet the stated modeling goals using the proposed method, the method is not ready for publication. Acknowledging these limitations within the manuscript, the authors are encouraged to fix these issues, remodel the area, and revise the manuscript. If these issues are corrected, this manuscript would be a worthwhile contribution to the literature.

We have remodeled the area, focusing on creating a better fit between training images and conditional data. We also included a section discussing our prior model selection and our approach for parameterization of the MPS algorithm. We have also solved the problem of soft data representation by converting a smaller fraction of the data into hard data and performed testing of different parameterization choices and whether rotation is effective in remedying inconsistencies between prior model and data, which is mentioned in the revised manuscript. We also quantify the effect of adding post-processing to the simulations through flagging percentages.

Reviewer 2:

RC2: "General comments: 1. I'm not very familiar with the MPS method and how it works exactly, so I can't really comment on the chosen parametrisation or even on whether the results produced are good/acceptable. However, I do wonder why the simulation artifacts occur? I think this should be elaborated on. It is stated that the geological artifacts are likely due to inconsistencies between TI and conditioning data (L565), but what does that mean exactly?"

The following is taken from our initial response:

"The main idea behind MPS is that spatial variability is described through a training image. For most MPS algorithms a realization is generated sequentially one voxel at a time. Consider standing at some location in the simulation grid and wanting to know what value to realize in this position. You then gather configuration of the surrounding already simulated nodes (conditional event). Specifically, in direct sampling, as used here, you then search the training image for a matching event. If such an event exist it is placed in the simulation grid and you can move to the next location and so forth until the entire simulation domain is filled.

Artifacts in the simulations might occur if it difficult to find a matching event in the TI. This is what is meant by line 565. If at some point it is impossible to find a suitable candidate, the next best thing must be selected. Sometimes this process leads to simulation artifacts. There are several ways this might be remedied but it all comes at a price. In all cases you are trying to broaden the number of possible matches in the training image (i.e., making sure the prior information is not in opposition to the conditioning data)

A simple way is to reduce the number of already simulated nodes to include in the conditional events. With smaller conditional events, more locations in the TI will be considered reasonable matches. This affects the ability to correctly represent the variability in the long-range structures from the TI. In general, it will lead to increased variability that is not necessarily related to geological variability as portrayed in the TI. Also, when simulating using fewer conditional data, the point being simulated might end up leading to inconsistencies for points to be simulated later.

These problems do not arise in 2-point based Gaussian simulation, because one can always compute the conditional distribution, while in MPS one must infer the conditional from a finite sized example (the TI) represent an outcome of the assumed underlying probability distribution.

An alternative could be to increase the size of TI and hence create additional possibilities for matching events in the TI. However, finding/generating a suitable TI is already the biggest challenge and strength of MPS simulation. The larger TI should still convey the same geology as before but with increased variability, which is not trivial.

We hope this sufficiently answers your question."

We have tried to rewrite the section on simulation artifacts, to a hopefully sufficient degree. We did not want to divert the paper too much from the core message by explaining in the level of detail as the answer above. We have included some lines discussing why these inconsistencies relates to the number of conditionals in the searching event.

RC2: "Why do the realisations deviate from the TI? "

The following is taken from our initial response:

"Following from the above we hope it is clear that realization will always deviate from the TI. It should not be the same. The only place we might expect to have a huge overlap between the TI and the realizations is the area from which the TI is derived…. The goal is to simulate new realizations that has the same multipoint statistics, but not the actual structures themselves. "

RC2: "The non-stationarity of the TI is mentioned but I don't understand what this means."

The following is taken from our initial response:

"The MPS simulation algorithm is expecting the same spatial variation in the whole training image. A training image that is stationary provides exactly that. For instance, postglacial sediments are only found at the top of TIs in the current study. When the algorithm searches for matching events containing postglacial sediment, matches can then only be found in some specific location. To be more technical, for a process to be stationary, all its statistical properties must be independent of spatial location. This is clearly not the case for a TI built on geology and redox for that matter. Both are extremely non-stationary by nature. Different processes shape the subsurface and it is not providing the same statistical properties at all locations. The issues of non-stationarity in MPS are discussed in detail in Mirowski *et al.* (2009). Another point is that issues with non-stationarity can be remedied with large amounts of conditional data (see e.g., Barford *et al.* (2018) or Vilhelmsen *et al.* (2019))."

We have added a few lines describing the concept of non-stationarity in the TI section. We refer to this during the discussion to help the reader get a fuller perspective on the issue. We hope this makes sense.

RC2: "It is suggested that artifacts could potentially be removed through better parametrization, but why would that be, and if so, shouldn't you have attempted to do some parameter 'tuning' as part of this work? "

The following is taken from our initial response:

"From our answers above we hope to lift the lid on why parameterization of MPS simulation algorithms is not straightforward. Parameterizing the algorithm depends on 1) the conditional data events available in the TI (its size, variability and stationarity), 2) computational demands and 3) ultimately if the realizations satisfy the variability that is expected within the training image. We focused more on later part from a qualitative perspective but can also include some quantitative measures also. By going too much into detail with parameter optimization we are afraid to divert the attention away from the core message of the paper. "

We ended up remodeling the area to create a better fit between training images and conditional data. In doing so we tested different parameterizations of the algorithm. We also discuss and justify our prior model through a set of realizations + we quantify the effect of post-processing the simulations through flagging percentages. We have rewritten the discussion to accommodate these new results.

RC2: "I understand that the artifacts are basically being averaged out over multiple realisations, and I think if the main aim was to produce best estimate and associated uncertainty (entropy) of the geology and redox (Figure 11), then the occurrence of artifacts would be less of an issue. However, if the aim is also to provide realisations as input for transport and fate modelling of nitrate, then I would think the artifacts could become much more of an issue, especially as flow and solute transport will depend non-linearly on those realisations. It is argued that for hydrological modelling at catchment scale,

the geological artifacts will have limited effect, which may be true, but I'm less convinced the same is the case for nitrate transport, especially given the redox artifacts."
The following is taken from our initial response:

"You are completely right here. This will remain an assumption on our part. We do not know to which extent these artifacts affect subsequent hydrological models. The logical next step and what we are currently working on is to explore how these MPS models of the subsurface architecture affect hydrological models. If these redox artifacts are significant, it is possible and trivial within the current workflow to e.g., enforce that below a certain depth only reduced redox conditions would occur. "

With the new modelling setup, we have reduced the number of artifacts. The remaining artifacts are hence less relevant than before. These artifacts are typically oxic conditions that appear at depth within an otherwise reduced environment and should hopefully not influence our hydrological modelling too much. But it remains an assumption.

RC2: "2. The joint simulation requires a bivariate training image (TI) for geology and redox to be developed. I understand how the geological TIs were developed, but I find it less clear how the redox TIs were produced, who produced them and how the bivariate/joint nature of the TIs are specified. I think this could be better explained."
We have fully rewritten this section to better inform the reader on the specific modeling choices of the redox TI.

RC2: "3. For the case study, two independent bivariate TIs are developed, one representing the Quaternary sequence and on e representing the buried valley. It is assumed that the delineation of these geological elements is known in the domain (i.e. no uncertainty as to whether a voxel is Quaternary or buried valley). But what is this delineation based on and with what certainty can this be done?"
The following is taken from our initial response:

"The delineation is based on geological interpretation. At the ground surface this interpretation is primarily guided by topographic information combined with the surface geology maps. At depth borehole information and geophysical data (tTEM, SkyTEM etc.) provides spatial information about the boundaries between the geological elements. Knowing the geological processes and the depositional setting from neighboring areas aids this delineation as well. You raise an interesting question regarding the lack of uncertainty for this boundary. We have discussed this internally within the author group. Clearly there is some uncertainty attributed to the location of these boundaries. The main question is how one can quantify this uncertainty. Also, one must find a way of incorporating these uncertainties in the simulation results. Finally, the effect on the realizations of having such an uncertain boundary must be assessed in comparison to a static boundary. To answer these questions in depth is beyond the scope of the current study. One idea that we would like to follow is to attribute the interpreted boundaries with a quantitative uncertainty based on the interpreter's qualitative uncertainty. By doing so we could simulate several boundary configurations prior to modeling. A set of simulation grids could then be produced that for which we generate one realization. The resulting set of realizations would then satisfy the independence between the geological elements and have a quantitative representation of the uncertainty of this boundary. Alternatively, there is also the possibility to, in direct sampling, simulate with an auxiliary probability grid covering the simulation grid telling which TI to use depending on spatial location. However, this would break the independence between the geological elements as conditional voxels from the buried valley would be included in conditional events for the Quaternary sequence. "

We put forth the idea of resimulating geological element boundaries in the revised manuscript, but have decided not to go into great detail as above within the manuscript.

RC2: "4. I can't really work out what the influence and significance of the TIs and the conditioning data are for the results. I think it would be relevant to include a discussion on the value of the information used. Given the large amount of conditioning data, does the TI become less important? Would the results be significantly different if you just used the same TI for both the geological elements? What is the effect of the soft data on the result? It would be interesting to see what the results look like if you excluded these. Except for the surface geology, the soft data looks quite uninformative."

We hope that the inclusion of the prior model now makes it more clear how much information comes from where. This also shows that if we were to use the same TI for both geological elements, we would get totally different results. Especially for redox, where the buried valley showcases planar type redox conditions whereas the Quaternaty sequence has the possibility of geological windows. Also, with the zonation of the lithology-resistivity relationship we have now made a much more informative transformation of the resistivity grid and hence more informative soft data. Furthermore, to make sure that soft data are not overruled in the realizations, some have been converted into hard data.

RC2: "5. It is not clear to me exactly how topography is used to inform soil moisture when deriving soft data for redox conditions. L365 just states how slope/topography affect moisture in general (it sounds like a TOPMODEL type of approach?). I think further explanation is required here. "

The whole section is deleted, because new data arrived since the last study and we did not need the topography to aid in deriving soft probabilities.

RC2: "6. Figures: The 3D figures look impressive, but there is a lot of them, and I think they are quite complex, and the colour choices make them not so easy to 'read' (especially the small ones). I wonder if some of the figures would perhaps be more illustrative in 2D rather than 3D. I don't think Fig 5 works very well as a 3D figure, and in Fig. 6 it is difficult to see that tTEM actually only covers part of the simulation grid. In Figure 1a it took me a while before I realised that the blue, black and red lines where actually the TEM and ERT data points. In Figure 7 the colours are difficult to distinguish in the histograms and it looks like there is a shade of blue in Fig 7a that is not in the legend."

We have made a new colormap more in line with color choices usually used to represent geological models. The figures have remained in 3D but following the reasonable criticism raised here we changed the viewing angle to better illustrate that tTEM does not cover the entire simulation grid. Furthermore, we have added arrows in the figures pointing towards relevant parts of the simulation results that are mentioned in the text. Figure 1 is changed along with the legend. Figure 7 is redone with the new color scheme as well as the new zonation.

RC2: "7. Overall, I think the paper is well-written, but there are places where I find the presentation of the material unclear and confusing and where more explanation would be helpful (particularly section 6 and 7). I appreciate that this is not helped by the fact that I'm not so familiar with the MPS method."

The following is taken from our initial response:

"Thank you very much. We will go through section 6 and 7 and try to rewrite them. As many readers of HESS would probably be unfamiliar with MPS we will try to elaborate on some of the concepts that might be confusing from an outsider. "

We hope this is now achieved.

Minor specific comments from reviewer 2:

Abstract: I struggled to understand L16-22 when reading the abstract the first time around (and the similar paragraph in the introduction, L106-110). It makes more sense to me after reading the paper, but I would encourage the authors to sharpen the text here.

We agree that this line is difficult to read without knowing the context of the paper. This is not ideal and has been changed to hopefully read better. In line 106-110 we deleted the part on how geological elements make the computations more feasible as this is covered in the discussion.

Figure 1c: Are the white areas unknown soils?

Yes, this is unmapped territory. The surface geology map is mapped over several field campaigns and some areas are still not mapped. We have updated the map and legend to convey this message better.

L167: Rephrase. It sounds like you are doing hydrological simulations as part of this work, which you are not.

Good point. This is not the case. Maybe this has happened because we are currently working on doing hydrological simulation on the realizations as a natural next step for assessing nitrate flowpaths. We have rephrased the sentence.

L169: There are 7 classes Table 1.

Yes, this is slightly confusing. The seven classes include Paleogene clay and unknown material, which is not part of the simulation. We have deleted the two classes that are not used in the geostatistical simulation. We hope this makes it less confusing.

L199: Not sure I understand why they provide independent measurements of redox.

This whole section has been rewritten.

L255: consists of. . .

Corrected

L306: From figure 1 It looks like the data availability is better in the southernmost part of the domain?

There are several criteria for selecting an appropriate site for TI construction. We fully acknowledge that the amount of geophysical data and boreholes tends to be slightly better in the southernmost part of the domain at surface level. However, the quality and depth of the boreholes in the chosen area is far superior to the southernmost part. There are only few wells that penetrate all the way down through the modelling domain in the area, and they are important to include in the TIs to make it more robust. We have tried to convey this message in the manuscript.

L310-311: I do not follow this. Why is sandy till not included in the TI? How does that affect simulations for the buried valley and can simulations here then meaningfully be conditioned to observations of sandy till?

No, they cannot. Sandy till is then only present at the surface where it is placed as hard data from the surface geology map. We have included some sandy till in the TI to accommodate this problem.

L321: I don't understand how its eases the construction of the TI to include parts outside the domain. Please clarify.

Because you get additional information about how the geology behaves that is independent information from whatever goes on in the simulation area. We have tried to convey this message in the revised manuscript.

P16 and Fig 7: I find it difficult to follow the text and observations that go with Fig 7. It is stated that sandy till is associated with some of the highest resistivities, but I can't see this from Fig 7. I don't follow how the general relationship /distribution of resistivities have been derived (Fig 7c). I think this whole section could be clearer.

To correct this, we have implemented a new colormap for resistivitity as well as for the lithology classes. This makes it easier to see the sandy till as it is now shown in black. More importantly we have accounted for some of the non-stationarity in the resistivity-lithology calculation by introducing three zones to obtain the relationship. We have tried to clarify the way to obtain resistivity-lithology relationship in the text with this comment in mind.

Equation 2: I'm not sure I follow the equation. Is the soil map (surface geology data) used to inform the lithology below the surface layer as well in a similar way as in the buffer zones as described in section 5.4.2? Or what is the probability of surface geology below the surface layer (intuitively I would think zero probability in which case the equation collapses)?

No, the surface geology map is only used to inform our model at the surface. The sum of all probability distribution over all categories must sum to 1. Hence the model will never collapse. You are right that in the case where all categories had zero probability this would happen. In an uninformed case the total probability of 1 is hence uniformly distributed amongst each category.

L435: . . .overlapping relationship. . .

Corrected

Soft data: do you derive soft probabilities for geology for all voxels in the simulation grid, even where hard data are present? It looks so from Fig 8 and 9.

Yes. But in practice hard data overrules soft data

457: To what extent did you experience "flagging" as part of the simulations you generated?

We have added a few lines describing the flagging process and the percentage of "flagging" and the impact of postprocessing for both prior and posterior model.

L485-500: I found some of the described observations here difficult to see in Figure 10. Maybe it would be easier to follow if there was a close-up on a relevant part of the transects instead?

Good point. Again, we want to convey the message that these models are in 3D but at the same time these 3D plots are not always satisfying for showing the results. We have now introduced arrows pointing towards phenomena that are mentioned in the text. This we feel provides a better overview for the reader.

L555: I'm not sure I understand why and when your simulations would not honour TI and data.

The TI is honored if the spatial variability and patterns shown in it are represented to some degree in the realizations. The conditional data should be honored by the algorithm. In practice, when using many conditional data with MPS, there will be inconsistency between the conditional data and the TI: The conditional events will not exist in the TI. This is due to the limited size of the TI (one of mulitiple sources).

This sentence is moved to the newly introduced paragraph on the prior model, which should hopefully clarify what is meant.

L574-585: I think the text here reads so well and it could be improved. This is also where terms like stationary TI and rotation in simulation are mentioned, which I find difficult to really follow not having a MPS background.

We have rewritten this section to hopefully read better. Also, we introduce the concept of stationarity in the TI earlier in the text of the current manuscript. This should hopefully aid in the understanding of this paragraph.

Summary of major changes to the manuscript

- Parts of the abstract is rewritten to make it more accessible to readers not familiar with the technical details of the paper
- The view of all figures of the simulation grid is made more oblique such that it is easier to see details of the model and aspects of the surface.
- We have added a more in-depth explanation on the creation of the TIs as well as a figure showing the profiles of redox conditions based geochemical observations used to make the redox TIs.
- Redox TIs have been re-modelled with new data. A small patch of sandy till is added to TI2 to allow the simulation of sandy till at the surface, which is more in accordance with the surface geology map.
- Added a section on prior model and how it fits with our expectation of the spatial variability of the system. This is now used as justification for using the current parameterization of the sampling algorithm
- Added a new colormap to geology more in line with usual color palette of geological mapping which is also makes it easier to separate colors visually in the realizations.
- The slope factor in the conversion of surface geology map into soft redox probabilities is discarded to avoid any confusion and new values are introduced based on new geochemical observations. The paragraph is therefore rewritten.
- We have implemented a zonal system for inferring the lithology-resistivity relationship. This means that we can account for some of non-stationarity that governs this relationship with depth. In addition, this new relationship helped the undermining of postglacial sediments in the former relationship as well as remedying the effect of very low resistivities for meltwater sand/gravel in the whole domain. This had also a positive effect on inconsistencies between the conditional data and the prior model, leading to fewer simulation artifacts. Consequently, the paragraph is rewritten at large as well as a new figure has been produced.
- The new lithology-resistivity relationship meant that soft probabilities have been recalculated and the section is altered slightly
- In terms of parameterization of the MPS algorithm we have expanded this section to include the prior model, which should hopefully reveal more of our thoughts and considerations regarding parameterization tuning.

- We have section in the discussion dealing with the role of soft data and expanded our thought on inconsistencies between prior and conditional data.

The following figures have been updated (old figure number in parathesis):

- **Figure 1 (Figure 1):** The legend for geophysical data is updated to lines instead of dots. The legend for geochemical data has been changed for "this study". The lakes are now shown in blue in (c), while the unmapped territory is kept in white.
- **Figure 2 (Figure 2):** Because the colormap of both resistivity and geology has been changed, the input data in Figure 2 is changed to create a more coherent manuscript
- **Figure 3 (Figure 3):** Changes to the new camera position of the simulation grid.
- **Figure 4 (Figure 4):** The colormap of geology is changed and a new legend is presented. a subfigure c) has also been added to show the zones used for inferring the resistivity lithology relationship. The TIs have also been slightly altered.
- **Figure 5 (NEW):** A new figure on the redox conditions of the study area used for a more in-depth analysis and justification on the creation of the training images.
- **Figure 6 (Figure 5):** Updated with new colormap and legend
- **Figure 7 (Figure 6):** A new colormap for resistivity is introduced to better visualize the logarithmical nature of resistivity measurements and thereby see the structures provided by the tTEM data
- **Figure 8 (Figure 7):** The new implemented calculation of the resistivity-lithology relationship for the three zones displayed in figure 4 is now shown instead of the combined calculation. The new colormap is also adopted for this figure.
- **Figure 9 (Figure 8):** Figures updated with new soft probabilities and added (f) to show the spatial distribution of the lithology-resistivity zones in the simulation grid.
- **Figure 10 (Figure 9):** Mode and Entropy recalculated from soft probabilities in figure 9 and new colormap adopted for the mode.
- **Figure 11 (NEW):** New figure showing two unconditional realizations from the prior model
- **Figure 12 (Figure 12):** Reworked figure of accumulative probability of redox. Now shown for TIs, Prior and Posterior instead of only posterior model as before.
- **Figure 13 (Figure 10):** Figure updated with new modeling results and made with same configuration as figure 11 for easier comparability
- **Figure 14 (Figure 11)::** Figure updated with new modeling results summary and new legend + colormap

The following tables have been updated:

- **Table 1:** We have removed the last two lithological group names as they are not present in geostatistical simulation and hence create more confusion than clarification. We have also changed "soil" into "sediment" to be more consistent with the manuscript.
- **Table 3:** The soft probabilities of the if redox conditions at the surface are changed and are now based on redox measurements from wells in and around the study area instead of the arbitrary "slope factor" that was originally introduced. We also do not discern between the two geological elements.
- **Table 4:** New table created to show the fraction of conversion from soft to hard data.