The authors state unequivocally that geostatistical simulation can quantify uncertainty in the geological model. This capability is overstated. Geostatistics and geostatistical simulation rely on different statistical models that are, or should be, based on data and conceptual models of likely rock property distributions. While MPS can be more effective at simulating more-realistic sediment texture patterns than other geostatistical simulation techniques, these all are only approximations. And the statistical models they are based are only constrained estimates of the properties being modeled – to the degree that the selected models reflect aspects of the real system; and, that they sufficiently sample, with high reliability, the properties being simulated. I suggest the authors adopt this perspective, and add the word, 'estimate' or 'estimating', when talking about quantifying uncertainty. 'Quantitative estimate of uncertainty' is a good example.

We agree that the uncertainty quantified here is a representation of the uncertainty of the specific probabilistic system and should only be viewed in this context. This was misleading on our part. We have changed accordingly in the manuscript.

The authors are inconsistent in their representation of the impacts of bias, subjectivity, and independence. They often suggest that subjectivity induces biases that are always undesirable. At other times, they advocate clearly for choices or methods that induce a specific bias, often based on the subjective justification that the choice better represents a preferred conceptual understanding. This inconsistency, and more importantly the erroneous representation, are significant in that they lead to confusion about the authors' advocacy regarding these concepts, and even to how the authors understand them. On lines 84-87 they state "In fact, subjective biases are accepted as one of the weak points of cognitive geological modeling…" In this situation the authors cite Bond 2015 and Wycisk et al. 2007 for this statement. It appears, however, that the authors have misrepresented both of these papers. Bond does not make this statement in her paper, and frequently recognizes the importance of geologic knowledge for successful modeling. Wycisk et al. never use the words subjective or bias, and state that knowledge of the system is required to make any modeling successful. Biases are endemic to any form of modeling – cognitive or quantitative (numerical, probabilistic, geostatistical). Generally, you want to add a specific bias as a constraint to interpretation (e.g., texture distributions in different training images, or the proportion of specific sediment textures in a proximal proglacial sedimentary setting). The main down side of expert insight in cognitive modeling is the difficulty in elucidating the various biases that the experts use. However, expert biases can be evaluated for likely correctness, and these 'constrained' biases can be very helpful in ruling out improbable sediment distributions; both Bond (2015) and Wycisk et al. (2007) recognize this. The implication of the authors' phrasing suggests that quantitative modeling is always better and less biased than cognitive modeling. I suggest that the literature demonstrates that both can be very useful if done correctly and properly qualified – and both can be similarly erroneous and misleading if not done correctly or properly qualified. I recommend language like, 'Subjective biases are seen by some as one of the weak points of cognitive geological modeling (Bond, 2015; Wycisk et al., 2009). It can be difficult to identify and quantify uncertainties due to biases in subsurface predictions from cognitive modeling, and so these biases cannot be fully accounted for in subsequent analysis or process modeling efforts (i.e., hydrologic modeling).'

An important point raised here. We fully agree with the presented argument and that our phrasing was unfortunate. This is corrected to the suggestion made by the reviewer.

On line 1000 the authors also seem to misunderstand importance and need for independence, in a statistical sense, and how bias is intentionally inserted in MPS. The goal of a TI is to introduce a bias, subjectively generated based on a combination of geological knowledge of the area and depositional settings, and observational data of the location. The reason MPS is being recommended, rather than other categorical geostatistical methods, is that these other methods don't introduce the right biases. In fact, they impose biases that make the models more incorrect than is generally desired. MPS is a way to introduce biases of more spatial continuity and large- to medium-scale heterogeneity into simulations, in order to generate model realizations that are more biased towards patterns of heterogeneity that are observed in outcrop and modern depositional analogues. Assumptions (or requirements) of statistical independence are part of the theoretical underpinnings of the tests. However, these methods are being used as approximations of more complex systems, and these assumptions can be (and usually must be) relaxed without making the MPS simulations worthwhile.

We think that the word "bias" is used differently here by the reviewer than how it used in the manuscript, which could explain this confusion. The message we have tried to get across in this line is a general concept from geostatistics. The TIs in the current study are based to a certain degree on the conditioning data. When we generate realizations from such a model to represent the prior distribution it does carry some information that is also present in conditional data. In the combined posterior models, we thus lose some degrees of freedom, and therefore one could argue that it most likely would be preferable to use a more conceptual TI to ensure that no degrees of freedom are lost. Furthermore, MPS is a way to make geostatistical simulations based on a spatial variability quantified through a training image. That's it. It is not a way to introduce biases into simulations. The reviewer is right that the spatial variability described in the TI affects the outcome of the simulations. As they should. And also, that the TIs represent all our assumptions about the system within the investigation scale. We have been upfront about this in the manuscript as well.

Figure 2: Input data…Soil Map. Should be Surficial Geology Map

It doesn't say soil map in the figure, but instead Surface geology mapping. We have changed it to surficial geology mapping according to the suggestion of the reviewer.

Figure 8 revision is a great contribution. However, the Y axis label on 8a, 8b, c are wrong. Should be a frequency or count. I'm a bit confused about how this was generated. See other comments in this section.
This is correct. The Y axis was wrong in the figure and has been changed to counts instead.

Figure 12, the X axes are incorrectly labeled. Should be 'percent' or maybe 'percent occurrence' or change xaxis values to be 0-1 for probability.
Changed to "percent occurrence"

Figure 12b includes results of two realizations. This isn't sufficient estimator of the prior. You need to label it with some other label.

This is a valid point. The two realizations themselves are not sufficient to describe the prior distribution. We have now based this probability with depth upon more realizations of the prior.

In general the figure revisions are much better. Good job on these. The patterns are much more readable and easier to follow with the discussions. A few more fence diagrams might be helpful in showing more of the internal architecture.

Good point. Additional fences have been added to help convey some of the internal structure.

There seems to be a bit of misunderstanding about the relationships between the data used for modeling redox conditions and about what one of the data streams means. On lines 235-236 the authors state, "The benefit of using the two types of data is that they provide independent measurements of redox conditions." This is incorrect, these data are not independent. The redox colors of the sediments reflect long-term redox conditions which are consistent with some period of pore-water chemistry. Importantly, the authors do not mention that the redox sediment colors are indicators of long-term historic redox conditions and not necessarily reflective of the redox dynamics within the recent, human altered subsurface. This may be why they feel the two data streams are independent. This study is not structured to evaluate this independence, and prudence suggests it is better to assume some level of dependence. The benefits of using the redox colors are that there are more of those data values than there are of the water chemistry, and the authors may assume that the relatively smooth variations in lateral redox sediment colors allow more reliable interpolation of the colors to unsampled locations than can be assumed for the recent shallow water chemistry. It would be helpful if the authors could add information or interpretations on how the redox colors correlate to water chemistry. The results might show a good agreement, or not such a good agreement. Importantly, the water chemistry has a temporal dynamic that is not captured by the colors.

We agree on the reviewer's comment about the differences in the redox interpretations between the sediment colors and water chemistry. Therefore, we revised the manuscript as follows:

Line 217-222: The sediment colors may be the resultant of the cumulative effects of the redox structure evolution since the deglaciation while the water chemistry may display a snapshot of the short-term redox chemistry, which may be temporally variable. Therefore, we postulate that the redox conditions interpreted from the sediment colors may be more coherent with the geological structure than that of the water chemistry. In addition, the sediment colors provide 1D profile information of the redox conditions and more data points are available compared to water chemistry which provide point scale information.

Also regarding the redox discussion, on lines 406-411, the authors make statements about inferring groundwater flow conditions from the redox colors of the sediments in the lowland (buried valley). However, the authors haven't presented any data that lets them make inferences about flow in the groundwater system. They are obviously able to cite other work in the area that did study gw flow, but no conclusions about flow can come from this work. Instead, the authors could say their work is, or is not, compatible with hydrologic observations and conceptual models documented by others (e.g., Kim et al., 2019). As a case in point, the authors suggest no horizontal flow because of clay-dominating conditions. However, a plausible alternate interpretation could be that soil development has provided preferential flow paths through the upper 3 meters of material across the landscape, that this network is exploited for groundwater flow, and the extended flowpaths from infiltrating up-gradient positions might provide sufficient residence time of nitrate laden water for significantly more reduction. I am not advocating for either hypothesis, just noting that there are no data to prove or refute either of these ideas.

We have presented the sediment colors and water chemistry data in Figure 5, which demonstrate no secondary oxic layers in the buried valley unit. The reviewer has a valid point regarding the hypotheses to explain the buried valley's redox structure. Currently, we could not rule out either explanation. Therefore, we revised the manuscript as follows:

Line 380-383: A secondary oxic layer below the first oxic layer is not expected, due to the clay-dominant conditions of the surface geology (mainly clay-till; Figure 1c) and subsurface structure. We concluded that in the buried valley, oxidants are delivered either vertically via water infiltration or gas diffusion or the top oxic layer (4-6 meters below the land surface) from the Quaternary sequence, resulting in the planar type of redox architecture (Kim et al. 2019).

There is more clarity of explanation needed for the discussions about how soft data are used and processed. Importantly the authors made some very helpful changes from an earlier version. This is in lines 533-575 of the ATC1 version. I particularly like the clarification of the potential impacts of non-stationarity and how this was handled. I like the use of the three zones. That's all very good. However, I'm particularly concerned and confused about the processes surrounding figure 8. It looks like the histograms were generated from assumptions of resistivity/lithology and the resistivity grid. The authors need to clarify the uncertainties or potential for error in the assumptions made and how these relationships were applied. There doesn't seem to be any calibration or validation around these resistivity/lithology relationships. This seems very subjective.

Not necessarily wrong, but it seems to involve a several significant assumptions that warrant some recognition.

We disagree that these choices are very subjective. They are subjective but are mainly based on the relationship inferred from the TI. The TIs on which this relationship is mainly built is subjective. If subjectivity is a concern, the TI construction is more the area to focus on.

The most subjective part of our estimation of the relationship comes from the weighting with the background relationship. The reason we do this is because we want to be true to our general understanding of the lithology-resistivity relationship. To this we argue that changing the weights and number of bins would only significantly alter the relationship in zone 3 as zone 1 is dominated by clay till and there is relatively good separation of the different lithologies in zone 2. For zone 3, we think it is better to make sure that meltwater sand does not occur for very low resistivities.

Since only a small portion of the cells are affected by our subjective weighting it will only have a limited impact on the TI-established lithology-resistivity relationship. We have adjusted our wording to be more precise about what is meant in the text.

Table 4: I don't understand what the upper row signifies. I feel like I understand the text general discussion that supposedly addresses the table, but I don't understand what the values mean. The first row needs a better explanation.
We have tried to rephrase the sentence explaining table four in the main text. We have also changed the wording in the table to hopefully make it more readable.

On line 660 the authors cite Høyer et al. (2017) in their use of 'unconditional simulation'. In fact the simulations are conditioned by the Tis, but not on any data. However, Høyer et al. don't refer to this unconditional simulation as representing the prior distribution. They used this approach to evaluate how effectively the TI constrained the realizations. While the authors are using TI-conditioned simulation for this purpose as well, it is not correct to suggest that two realizations are able to provide sufficient representation of the whole prior distribution. This is important on Figure 12b, as you suggest it is the prior, but again, only two realizations (i.e., samples).

Regardless of whether Høyer et al. specifically refers to unconditional realizations as representing the prior distribution or not, the fact remains that it is exactly what they the authors are proposing. We prefer the wording of 'unconditional' as this is more in line with geostatistical literature. To name a few examples:

Gravey M, Mariethoz G (2020) QuickSampling v1.0: A robust and simplified pixel-based multiple-point simulation approach. Geosci Model Dev 13:2611–2630. https://doi.org/10.5194/gmd-13-2611-2020

Juda P, Renard P, Straubhaar J (2020) A Framework for the Cross-Validation of Categorical Geostatistical Simulations. Earth Sp Sci 7:1–17. https://doi.org/10.1029/2020EA001152

Laloy E, Hérault R, Lee J, et al (2017) Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network. Adv Water Resour 110:387–405. https://doi.org/10.1016/j.advwatres.2017.09.029

Remy N, Boucher A, Wu J (2009) Applied Geostatistics with SGeMS, 1st edn. Cambridge University Press

Tahmasebi P, Hezarkhani A, Sahimi M (2012) Multiple-point geostatistical modeling based on the cross-correlation functions. Comput Geosci 16:779–797. https://doi.org/10.1007/s10596-012-9287-1

The idea from Høyer et al. (2017) is to evaluate whether the realizations that arise from the chosen MPS algorithm (and parameterization) + TI provide the structural input that is intended.

It is true that the two realizations are too few to depict the full spatial variability of the prior distribution. In our analysis we therefore used more than these two realizations to assess the prior distribution. However, we only showed two realizations in the manuscript to provide an easy visual insight to our reasoning for the reader. This was not clear in the manuscript. We have corrected this.

On line 791 the authors state, '…and hard data increase the information content (lowers entropy)…'. I believe this is incorrect. This should be written as, 'and hard data **decreases** the information content (lowers entropy)…' Remember, the equation is for Shannon's information entropy, not thermodynamic entropy. (cf. https://stats.stackexchange.com/questions/101351/entropy-and-information-content). Alternatively, the authors could think of entropy in terms of bits – a la Shannon. It takes more bits to describe more complexity, and fewer bits to describe less complexity. More bits = more information content. Fewer bits = less information content.

Here we must object. Entropy is a measure of disorder. Both in thermodynamics and information theory. Hard data *do* increase the information content. It lowers the self-information (the level of surprise) though. In the past these two terms (information and self-information) were unfortunately used interchangeably which might still today provide some confusion. Consider the example provided in the attached link.

We can all agree that the maximum entropy for a bit happens when the underlying distribution is split 50/50 for each outcome. When a single bit is realized as either heads or tails the self-information is high because it is impossible for us to predict this event happening. I.e. we are surprised by its outcome. That conversely means that the information content in our underlying distribution is low.

Conversely, if we have hard data in a voxel the outcome is the same in each realization. There is then no disorder, low entropy, and low self-information because each bit does not provide any information or surprise. Instead, the underlying distribution is fully informed. Following the point from the link, it is far easier to compress information when the self-information is low. In the case with hard data, we basically can compress the information down to a single value (heads or tails). But this is because the underlying distribution is well-informed.

We refer the reviewer to this recent article of Hansen (2021) (https://link.springer.com/article/10.1007/s11004-020-09876-z) for a more in-depth description of entropy and information content of geostatistical models. We have also added this reference in the main text for the readers to find.

On lines 808-809 the authors state, 'This imply that simulation artifacts are not reoccurring in simulations…' I would strike this sentence. It's not phrased properly and is incorrect; analysis of the mode, as a statistic, doesn't prove or disprove either the frequency of artifacts or a bias. If the authors want to keep a reference to these figures, the text should be fixed so it doesn't erroneously say, 'no overall bias is found'. Maybe something like, 'The algorithmic biases that create these artifacts are generating a small number of artifacts that are not significantly affecting the posterior histograms.' However, this point was already made when the authors looked at the histograms earlier in the paper, so this sentence is not necessary or constructive. The statement, 'This is clearly an unwanted side-effect of the soft data conversion…' seems like an inference based on assumptions rather than an observation based on the algorithm. If that is the case, this statement is unsubstantiated and should be deleted. Consider a hypothetical example: any additions to the distribution would have no effect on the mode as long as they occurred less frequently than the modal value. These additions could be due to a systematic error (i.e., bias).

This is a valid point. We have deleted the sentence accordingly. Regarding the second point raised about the unwanted side-effect, we disagree. We observe no bias as such. We just observe that some of the converted soft data are placed as sole voxels in the mode modal. That means that even though these conditioning data provide some of the first hard data to condition on, there is no preference in the simulation to simulate oxic conditions around these. This is just an observation. We have removed the word "unwanted" from the sentence to remove the negative loaded interpretation on our behalf.

On lines 815-832, I think this passage is mistaken. Importantly, the redox scale only has three categories. The posterior distribution showed almost no occurrence of the middle, reducing, category. Don't forget the postglacial seds were assigned a highly reducing value. And even in other textures, there was a non-negligible proportion of reduced occurrences. These all probably contribute to generate a higher entropy at land surface. The prior and TI for redox were strongly biased towards reduced conditions at a certain point in the subsurface. This is reflected in the low entropy. I don't see this as counter intuitive.

Fair point. Although we would argue that it at first may seem counter-intuitive knowing that the entropy of aligns well with decreasing information sources with depth. We have deleted the 'counter-intuitive' and added a line saying that "The high entropy at the surface is likely also aided by the soft data."

On line 959, the authors state, 'Although the small size of the TI may pose a problem for reproducing the intended variability…' They have been a bit inconsistent with how they refer to the Tis in terms of their level of detail, the intended goals with respect to the level of detail the authors are trying to represent, and with the recommendations on how much detail to put into a TI. These decisions do involve a bit of professional judgement, more clarity is needed about the goals for the TIs, what the authors tried to do with the TIs, and how they are accommodating the limitations that the choices caused.

We have now added a line in the end of subsection 7.5 Training images and geological elements addressing our recommendations for TI creation in the future.

On line 1016 the authors state, '…allows the quantification of uncertainties in the input data…' No it does not do this. It is not correct to use statistical models that were created by a data set to evaluate the uncertainties of that same data set. There isn't that much real data to begin with, so there really isn't enough data to prove any distributional assumptions or models. The manuscript details how the uncertainty in the geologic sediment and redox condition distributions have been quantitatively estimated.

We agree. It allows a representation of the quantified uncertainties of the input data. It cannot quantify the uncertainty of the input directly. We have changed the wording accordingly.

Line Number: Comment
8: 'nitrogen (N) losses'…you are only dealing with one loss pathway, leaching. should use 'nitrogen (N) leaching' rather than 'losses'.
Corrected

13-15: 'and 2) information of lithology and redox conditions…' this list is presented incorrectly. it should be: '2) lithologies from borehole observations, 3) redox conditions from colors reported in borehole observations, and 4) chemistry analyses from water samples.'
Corrected

23: '...and conditional data'…should be: '…and conditioning data…' See Straubhaar, Renard, Mariethoz, 2016.
Corrected

26: maybe 'consistent' rather than 'coherent'?
Changed to "consistent"

30: 'which may be fundamental to better understanding of the retention of the subsurface…" might read better as: which may lead to a better understanding of nitrogen (N) fate in the subsurface…"
Corrected

34: maybe 'loss' instead of 'escape'…leaching is not ideal here because your citations also include runoff Losses
Corrected

50: questions: Do you mean nitrate concentrations, instead of conditions? If 'conditions' is correct, I'm not sure what that means. Can you clarify. Assuming concentrations is correct, shouldn't it also be 'leaching concentrations'?

This should be concentrations. 'Conditions' was revised to concentrations.

56: 'contaminants: 1)' should be something like, "contaminants. Modeling approaches have included: 1)…"
Corrected

66: I would suggest a better wording than, '...space thus requires…" might be, "…space would benefit from more-detailed…" Mostly, this seems to imply that your 25x25x2 is sufficient to capture all detailed heterogeneities in redox conditions, which isn't accurate (it's probably also not what you mean…it just seems implied in the current wording).
Corrected to proposal

80: 'is attributed to uncertainties such as' should be 'contains uncertainties from sources that include'
Corrected to proposal

84-87: "In fact, subjective biases are accepted as one of the weak points of cognitive geological modeling…" The authors cite Bond 2015 and Wycisk et al. 2007 for this statement. The authors have misrepresented both papers. Bond does not make this statemente and recognizes the importance of geologic knowledge. Wycisk et al. never use the words subjective or bias and state that knowledge of the system is required to make any modeling successful. See comments above about the authors' use of subjectivity, bias, and uncertainty concepts.
We have rewritten this part according to the suggestions of the reviewer to not misquote the references.

105: the geostatistical simulation doesn't "quantify uncertainty in…data". It does allow you to quantitatively estimate uncertainty in the unsampled locations, but not the sampled ones.
This has been rewritten to be more precise.

106: possible realizations of the subsurface…not a quantitative measurement of uncertainty?
We have rewritten this sentence to be more precise.

235-236: "The benefit of using the two types of data is that they provide independent measurements of redox conditions." No that is not correct. The colors clearly reflect long-term redox conditions which are consistent with the pore water chemistry. So they aren't independent. See comments above.

See our answer in the beginning

324-325: '…based on geological event chronology.' should it be '…based on sediment heterogeneity and geological event chronology.' ??

Corrected

406-411: You haven't presented data that lets you make any inferences about flow in the groundwater system. See comments above.
This is addressed in our earlier answer

458: '5.4 Conditional data' should be '5.4 Conditioning data' There are conditional simulations and conditional realizations that are conditioned to the data. This means the data become condition**ing** data.
Changed

533-575. I like the revisions to this portion of the manuscript. See comments above.
Thank you

586+ I'm a little concerned about the term, 'soft probability'. I realize Caers and Mariethoz may use it, but qualifying the word 'probabilities' with 'soft' seems inaccurate or at least unclear. 'Soft data probabilities' is more intuitive and seems more correct. I notice you use 'soft data probability' on page 10. Please check on the usage of this term and consider either using 'probabilities' or 'soft data probabilities'. I leave it to your discretion.

We agree that this was unclear. We have changed to 'soft data probabilities' throughout.

587: 'resistivity grid in Figure 7' probably makes more sense as, 'resistivity grid from Figure 7'
Corrected

614-616: While the mode shows layers in the buried valley, which might be bad, these same voxels show a lot of entropy, which implies a wide distribution over these cells. Which is good. You might make this point in the text. I think it's helpful analysis discussion for readers.
Agreed, we have now incorporated this point into our argumentation.

623: Table 4. I don't understand the upper row. I understand from the text, what general topic it's addressing, and I think it is probably helpful. However, I don't understand what the values mean. The first row needs a better explanation.
Addressed earlier.

633: 'In direct sampling, the nodes in the simulation grid…' You never discuss this term (direct sampling) nor contrast it with alternative MPS simulation modes. Maybe back up when you introduce MPS, put in a couple sentences about direct sampling and snesim and that your version uses direct sampling. Just for context here.

We stated in subsection 4.1 MPS modelling (second paragraph) that:

"In the current study we use direct sampling (Mariethoz et al., 2010) as implemented in the software package DeeSse (Straubhaar, 2019)."

Is this not sufficient? We are not sure that it will benefit the average reader to discuss differences between e.g. direct sampling and SNESIM, ENESIM or other types of MPS. We also stated our main reason for working with direct sampling:

"The main reason is its ability to utilize a bivariate training image that allows for joint simulation of geology and redox. "

We intend to keep it this way

635: can you add a citation or citations that you used to generate this list of parameters?
We have added a reference to the original paper from Mariethoz et al. 2010.

660: Høyer et al. (2017) do not refer to this unconditional simulation as the prior distribution. See comments Above
This is true, we have argued to keep the current formulation. See our response earlier.

667: 'only occur near the surface in accordance with the TIs, but more infrequent than portrayed.' …with some corrections… 'only occur near the surface in accordance with the TI, much more infrequently than portrayed '
Changed

686: '…incorporation of conditional data…' should be, '…incorporation of conditioning data…'
Corrected

709: 'conditioned data' should be 'conditioning data'
Corrected

718: '…the soft data is particularly dominating the realizations…' could be '…the soft data is the dominant constraint on the realizations…'
Changed to suggestion

726: '…This coherency explains…' could be '…This consistency explains…'
Changed to suggestion

740: question: why don't you more completely filter these oxic artifacts? even in post-simulation processing?
Because we do not want to focus on these artifacts specifically. We acknowledge their presence; discuss why they are there (due to inconsistencies between prior and conditioning data) and suggest how to handle such artifacts. By post-processing and filtering the simulation, one alters the posterior such that it does not represent the combination of the original input. This is a factor that needs consideration as well. Post-processing should be viewed as more of a necessary evil than the default for removing artifacts. This warrants a deeper analysis of how inconsistencies can be lowered in the first place, and hence alleviating the need for post-processing.

Furthermore, we argue that for further hydrological flow modeling these artifacts may be negligible. That does not alleviate the problem, but instead suggest that for a greater modeling workflow other issues needs to be prioritized.

778-779: '…which together represent the full posterior model.' I think it would be more correct to say something like, '…which together are used to represent the posterior model.'
Changed to suggestion

787: '…indicating that other categories are almost equally probable…' probably should be something like, '...indicating that these voxels have a near uniform distribution among several different categories.'
Changed to suggestion

791: '…and hard data increase the information content (lowers entropy)…' should be, 'and hard data decreases the information content (lowers entropy)…' See comments above.
We disagree and have kept the original wording. See our previous answer.

808-809: I would strike this sentence. See comments above.
We have done so. See earlier answer.

815-832: I think you're mistaken here. See comments above.
The section is slightly altered. See earlier answer.

836: 'examples of mapping redox conditions with multiple-point geostatistical simulation…' I don't think the use of mapping is appropriate here. It's the wrong audience for this use. A better wording would be, 'examples of simulating redox conditions with sediment-texture distributions using multiple-point geostatistical methods…'
Changed to suggestion

848: 'The spatial variability of the TIs is well represented in the prior realizations…" Not really. You only generated two realizations, which is insufficient to estimate the prior distribution, and the prior distribution that you list is not a good estimate of the redox TI - the green zone (reducing) is not well simulated at all.
We generated more than two realizations but showed only two in the manuscript. We agree that we overemphasized how well the spatial variability of the TI is represented in the prior distribution. We have removed the word 'well' form the sentence in question.

848: 'conditional data'…'conditioning data'
Corrected

854, 857: 'conditional data' …'conditioning data'
Corrected

847-865: Nice job on this section. It's a good contribution.

Thank you

868: 'The random path have a tendency to underestimate soft data…' ? You later suggest this is why the
877: 'randomly converted a fraction of the soft data…' can you specify the fraction? You are defining a novel method and this would probably be helpful to people.

The fraction is shown in table 4. Perhaps this was unclear before but should hopefully be clearer now.

939-961: 7.3 'The inclusion of geological mapping experts in the creation of ThIs introduces modeling subjectivity.' See comments above.
Yes, the inclusion of geological mapping experts introduces subjectivity modeling. This is not necessarily bad or good per default. We don't think there is any inconsistency in argumentation in this regard. We further proceed to argue that subjectivity is important to bring forward expert knowledge. We mention that others are working on data-driven approaches for generating TIs, but our reasoning for doing a manual interpreted TI is precisely in line with the reasoning of Curtis 2012 and Tarantola 2005.

In summary we do not try to avoid subjectivity in our study, but we highlight that it is present. We have tried to rephrase this section to emphasize the need for subjectivity in designing the training images

959: Although the small size of the TI may pose a problem for reproducing the intended variability…" earlier, you stated that you wanted a generalized model that didn't fully capture all of the possible geometries. can't then make poor fit a criteria.
We don't state that we want a generalized model. We explicitly say that we intend to make training images that are sufficiently large to capture the general patterns in each geological element. We do recognize that the small size might pose a problem when running the MPS algorithm as fewer matching configurations are possible. This is a trade-off between considering the amount of time it takes to construct a training image, the data availability, independence of information, how well the features can be reproduced (algorithm parameters) and the computation time. We recognize that there is a future study for investigating this trade-off, which is beyond the scope of the current study.

1000 "…would ensure independence of information." See comments above.
See our answer in the beginning

1003: conditional >> conditioning data
Corrected

1007: computationally feasible… 80 seconds per realization is trivial. Could have done many more than two realizations to estimate the prior.
We have also done so but chose to only show two realizations in manuscript as in our above-mentioned answers.

1016 7.7 …allows the quantification of uncertainties in the input data…" No it does not. See comments above.

We agree. It allows a representation of the quantified uncertainties of the input data. It cannot quantify the uncertainty of the input directly. We have changed the wording accordingly.