Response to RC1 – Ahston Krajnovich

AR = Authors response

Thank you very much for your constructive review of our manuscript. We have incorporated many of your suggested improvements into our manuscript and believe this has significantly improved its quality and readability.

Most importantly, we have added our reasoning behind the choice of prior parameters and our reasoning behind choosing a known topology graph as our constraint. Additionally, we have revised the mathematical notation throughout the paper to make it consistent and in line with the cited literature and have edited so that definitions of technical terms appear before use of them. The placement of the figures will be subject to the final paper typesetting done by the journal and is thus not final in the current manuscript. We have addressed many of the suggested improvements to language, grammar and figure annotations to improve readability.

Please find all our detailed responses to your comments in the supplementary material, along with both the revised and change-tracked manuscript.

Specific Comments

Title: Consider rephrasing to avoid the repetitive use of the word "using". I would suggest: "Constraining stochastic 3-D structural geological models with topology information using Approximate Bayesian Computation in GemPy 2.1"

AR: We have changed the title.

Abstract: As the research is built in the GemPy environment, it would be beneficial to highlight it's usage in the abstract (perhaps at Line 13).

AR: We now mention GemPy in the abstract to improve clarity.

Line 129: Sentence requires revision to be accurate about what the likelihood function represents in Bayes' theorem. I suggest: "This updating process relies on the use of a likelihood function p(y|theta), representing the conditional probability of the observed data y given the prior probability of the underlying parameter theta and the theoretical connection to the occurring event."

AR: We have incorporated your suggestion into the manuscript to improve clarity for the reader.

Line 144: You have reversed the conditional probability described by the likelihood function, which is: the likelihood for observing the data y, given the model based on uncertain parameters theta.

AR: Thank you for pointing out this error, we have switched that around!

Line 147: This is unclear, as likelihood functions are inherently encoding information regarding not just the parameters theta, but also the observations y and the assumed theoretical relationship between theta and y. Consider removing or revising.

AR: We have removed the sentence to avoid confusing the reader.

Section 2.3.2: This section requires additional clarification between "observed data" and "simulated data". Refer to the treatment of ABC in Gelman et al., 2004 where y is the observed data (observed "summary statistic" in ABC) and y-rep is the simulated data (simulated "summary statistic" in ABC). The use of y-hat to represent the observed summary statistic and y to represent the simulated summary statistic creates additional confusion (as the observed data introduced in Bayes' theorem were defined as y, not y-hat).

AR: Thank you for pointing out this mistake. We have changed the notation to be in line with the literature and our description of Bayes' theorem.

Line 156-157: Perhaps add a reference to (Wood and Curtis, 2004)? (Geological prior information, and its applications to geoscientific problems)

AR: Added reference to provide the reader with additional literature to understand the issue of specifying likelihood functions in geology.

Line 160: Please add an additional clarifying sentence on what the summary statistic is in this work rather than the short parenthetical (to avoid confusion with typical summary statistics like mean, mode, median etc.). Also, a comment: In the proposed (approximate) inference scheme, the new evidence y (or data) is the "summary statistic". So, while the definition of the additional term "summary statistic" to describe "y" is useful for highlighting the approximate nature of ABC, the equivalency of these two terms should be clarified for the reader.

AR: We added additional clarification on what is usually used as summary statistics, and why we use topology graphs when comparing geomodels.

"While summary statistics are often measures such as the mean, mode or median of a model, they tend to be meaningless in summarizing geomodels. In this work we use the geomodel topology graph as a summary statistic of the geomodel to provide a meaningful comparison between geomodels."

Line 162-163: Clarify the 2nd part of the sentence to illustrate that the "observed summary statistic y-hat" is static for the entire geomodel ensemble (i.e., the known, observed topology graph), while "the summary statistic y" is tied to each individual geomodel realization (i.e., a simulated topology graph).

AR: We have fixed the mathematical notation from y to S(y) when referring to the summary statistic. We think this fixes the problem of clarity in this sentence. While we keep the observed topology graph static in this experiment, this is by no means necessary, as we discuss in the paper. We have also improved our discussion of this.

Line 165: Theta-prime has not been introduced. What does it refer to as opposed to theta? I assume you are referring to a single draw from the parameter distribution theta, but please clarify. When relying on mathematical notations from another work (the ones in question here seem to be borrowed from Sadegh and Vrugt, 2014), make sure notations are introduced properly. It also helps to also have a "sanity check" to make sure that the notation used is not confusing with respect to the broader statistical literature (e.g., where the observed data in Bayes theorem are typically represented without a ˆ or ')

AR: Theta has been introduced in the previous section as the model parameter distributions. We have added an explanation that theta prime is a sample from these distributions.

Section 2.5: Section could be made much more concise to avoid excessive overlap with existing works (seeing as the major contributions of the paper are not focused on novel applications of Shannon entropy).

AR: We have cut detailed explanations of the Shannon entropy and refer the reader to the relevant literature.

Line 227: How and why were the prior uncertainty ranges chosen? Were they considered to be broad, non-informative priors, derived emprically, based on background information or simply assumed by the modeler for the sake of simulation? Same question should be addressed more directly for the Gulfalks case study as well (Line 249-251), where the uncertainties appear to be derived from the referenced work though this is not stated definitively. Also, just a comment: I am quite interested to see how incorporating structural uncertainty (by way of the methods put forth by Pakyuz-Charrier et al., 2018a,b, Roberts et al., 2019 or Krajnovich et al., 2020) would influence the geomodel topology.. Intuitively, there is a high potential for confounding effects on the range of possible geomodel topologies when interface location and interface/fault orientation are varied together!

AR: We have added explanation on how we chose the prior parametrization. As this paper focusses on developing and showcasing a new methodology for constraining uncertain geomodels using topology graphs, prior parametrization has not been a focus of the experiments. We refer the reader to other works, as also mentioned by the reviewer.

Line 246: How was the interface uncertainty applied to the surface points? Independently at each node, or generally to the set of surface points (so as to retain surface shape). From reading into the supplemental codes, it appears that the uncertainty was applied to the group of surface points – but this information needs to also be included in the text for the typical reader. This also applies to the synthetic model, which appears (from the code provided) to have been modeled from similar groups of surface points, though this is not clarified in the text.

AR: We have added the missing description on how the interface point uncertainty has been applied in both the synthetic and the real-world examples.

Line 251: Tying back to the earlier comments on how prior uncertainty ranges were chosen, I believe that "ease of implementation" is somewhat of an inconclusive reasoning. The rest of the sentence provides more meaningful perspective but still could be expanded upon (e.g., what is "simplified uncertainty modeling" in this context?). Please add some more detail.

AR: We have addressed our inconclusive writing and now more clearly describe how the model is perturbed and why we chose this approach.

Line 268: A figure representing this most frequent topology graph from simulation (or other selected simulated topology graphs) would be quite insightful, especially if accompanied by a discussion of their geologic significance (e.g., tying back to points made during the introduction (Line 51), did any simulated topology graphs represent a compressional rather than extensional tectonic regime?). If length permits of course - perhaps if some figures are combined or suggested section lengths reduced, this could be added.

AR: This could be beneficial, but we think it would distract the reader from the main message of the manuscript: the method itself. Analysing ensembles of topologies for geological setting would require either painstaking manual evaluation of every model or require more research into defining how extension settings can be detected reliably from topology graphs. This would indeed be a very interesting topic, but out of the scope of this research.

Line 287: Since the Jaccard Index used could allow for multiple topologies to be present in the final model ensemble (depending on the rejection threshold used), it would be beneficial to see some exploration of what these possible model topologies looked like (how geologically unrealistic do they get? Are all 675 unused topologies absolutely unrealistic?). Including a discussion of this sort would help guide future works investigating uncertainty of the applied topology information itself (without requiring reproducing the results to show geomodel uncertainty when multiple simulated topologies were present in the final ensemble). See also Comment for Line 268.

AR: Please see our answer to the previous comment.

Line 295: In line with the missing clarification regarding the assumption of the observed topology graph being known without uncertainty, add some clarification behind the reasoning for setting the rejection threshold such that only the applied initial topology remains in the probabilistic geomodel ensemble. Was the goal of empirical testing of thresholds to find the largest threshold which resulted in only a single model topology remaining across the probabilistic geomodel ensemble?

AR: Our aim was indeed to show how to constrain a stochastic geomodel to a topological state --- to allow for the reliable simulation of uncertainty within a single topology state (kind-of like a single conceptual model). We think that more research into how to identify, and thus compare and constrain with, geologically similar geomodels from topology graphs is needed to allow the meaningful relaxation of the error threshold.

Line 297: How does simulation time for ABC-REJ compare to simulation time for the standard MC approach?

AR: We don't believe this comparison is very meaningful, as this is highly dependant on the choice of error, summary statistic etc. That's why we chose to only compare the ABC-REJ with the ABC-SMC – as they are aiming for the same constrained outcome, while the Monte Carlo forward

Line 298: This is a significant improvement in efficiency! Perhaps include a description of acceptance rates from each epoch of SMC, or at least a comparison of the final acceptance rate at the threshold value of 0.025 in SMC for comparison with the rate given for REJ. This information might fit naturally in Figure 12.

AR: We think that acceptance rates for the multiple epochs of the ABC-SMC are difficult to compare to the single acceptance rate of ABC-REJ. One could compare the average, or weighted average, but we believe that the comparison of overall simulation time is more meaningful in this case.

Line 324: If the information applied were non-meaningful (e.g., an incorrect topology graph), the geomodel ensemble would likely still exhibit a reduction in entropy due simply to the convergence of the model realizations towards the single model topology applied. That is, the reduction in uncertainty is arising from the reduction of possible model topologies, not necessarily the meaningfulness of the model topology used in the ABC algorithm.

AR: We fully agree, which is why we highlight the fact that only a meaningful constraint can lead to a meaningful reduction in model uncertainty.

Line 334: It appears that expanding the ABC approach proposed here to incorporate multiple observed topology graphs would not be a matter of "easily scaling". Revise to clarify that the general ABC framework would definitely allow for this, although it would require reparameterizing the current summary statistic and discrepancy measure (distance function), and also possibly changing the simulation method (as mentioned in Line 357-359).

AR: Incorporating multiple observed topology graphs for comparison in the ABC framework would indeed scale easily, as the computational complexity would scale linearly. Thus, a doubling in comparisons would double the Jaccard index computation, which by itself is trivial in terms of computation cost for such small networks in comparison with the computation cost of generating the geomodel in the first place. In the current implementation it would simply requiring looping over a set of topologies and computing the Jaccard index and accepting if one of them is below the allowed error threshold.

Line 335: This would be a good place to bring up again the implications of using the demonstrated ABC approach if there were uncertainty about the observed topology graph.
AR: The uncertainty in observed topology can be addressed via proposing several acceptable topologies or by scaling the error threshold.

Line 345: “: : :reducing the parameter dimensionality” – how so? The number of input parameter probability distributions is the same in standard MC or in ABC-REJ/SMC. The computation efficiency improvements arrive from reducing the number of input parameter draws that are run through uncertainty propagation to the 3D geologic model space, which in SMC also allows for reducing the size of the uncertainty space (note, not the parameter dimensionality) iteratively.

AR: We describe here a trade-off that could be made between uncertain geomodel parametrization and probability of subsequently simulated geomodel samples being valid geological models. Increasing geomodel complexity generally requires increased geomodel and statistical model parametrization – which thus scales the parameter space exponentially (see e.g. Betancourt, Michael. "A conceptual introduction to Hamiltonian Monte Carlo” (2017).). This increase in parameter space will drastically increase computational time. Thus, a balance generally needs to be made between model complexity and model parametrization. Our method could allow for lower levels of model parametrization while retaining model complexity by essentially filtering topologically wrong samples (which will inherently become more numerous when stochastically perturbing an under-parametrized complex geomodel).

Line 370: This was not discussed earlier in Section 3.2 when the acceptance rate was initially 0.0059 (0.59%). Does that low acceptance rate warrant reassessing the prior input uncertainties used in the probabilistic geomodeling? Should be discussed to better frame the current work and guide future work.

AR: It might! It definitely warrants a good look at the prior parametrization. But stochastic geomodels have such inherent topological complexity, as minor changes in the location of interfaces across a fault can have various effects on model topology.

Figure 4: Consider replacing X & Y with N & E to be more intuitive for geoscientists. Applies to all figures of geomodels with labeled axes.

AR: We have chosen non-descript axis labels for these figures as this model is entirely synthetic.

Figure 6: In my opinion, the XZ difference section (and possibly then also XY and YZ sections) from Figure 7 could be appended onto Figure 6 for ease of reference. Also, what do the overlain crosshairs show?

AR: The lines show the locations of the respective other sections. We have added an explanation to the figure description to make this clearer.

Figure 8: Figure does not show (a), (b), (c): : : tags. Also, as mentioned in the comment for Line 279, the figure does not show histograms.

AR: We have added subfigure labels a-f and we have removed the outdated reference to histograms to only KDEs.

Figure 10: The significant reduction in model entropy indicates the strong dependence on the initial topology used - this potential source of bias should be addressed. Please discuss the implications of using a rejection threshold which only allows one model topology across the entire final ensemble of geomodel realizations. Since the authors are operating under this (valid) assumption, it needs to be clearly stated earlier that the initial geomodel topology is "known" and treated without uncertainty. See also comments regarding Lines 324, 295, 287 and 268. Also, I believe this figure could be merged with Figure 11.

AR: We have added explanation of this bias in our elaboration on why we chose to constrain with a single topology state in our method description. We hope this clears things up for the reader.

Figure 12: Figure needs correction to show Y-axes labels. Perhaps acceptance rates per epoch would be useful to add as well, as they are tied to the processing efficiency improvement of 10.1x (see comment regarding Line 298).

AR: We have added Y-axes labels and removed then redundant titles. We are not sure how acceptance rate would improve would improve the bar chart. We have chosen not to add it to keep the figure concise.