Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-136-RC2, 2020

Geoscientific Model Development Discussions
Open Access
EGU

Interactive comment on "Constraining stochastic 3-D structural geological models with topology information using Approximate Bayesian Computation using GemPy 2.1" *by* Alexander Schaaf et al.

Anonymous Referee #2

Received and published: 21 October 2020

General Comments:

The article proposes a method for applying constraints derived from topological geological knowledge to ensembles of geomodels generated by Monte Carlo algorithms. The concept is simple, its application fits well with common Bayesian methodology and I can see its practical value and relevance to this journal. Experiment setup and results are well presented and the software used is thoroughly described so as to be reproducible. While some readers will have questions about how exactly various prior parameters were chosen, I interpret this article as being about the method of incorpo-

Printer-friendly version

Discussion paper

rating topological data and not the practice of deciding on priors which is covered in many other works. I am asking for minor revision for the following reasons:

1) Many sentences are poorly worded.

2) The term 'likelihood-free' and its implications are presented inconsistently and in a very misleading way.

3) Terms like observation, prior and likelihood are applied in an unclear manner often contradicting convention.

4) The method description is much longer and more elaborate than it needed to be. This stems from the choice to phrase the approach in terms of Approximate Bayesian Computation (ABC).

To be clear, I think the choice to frame this in terms of ABC was unnecessary and only makes the article much longer and more tedious than was needed. I am not asking for this to be changed, as that would be more effort than it is worth. I make this point here simply for the record and in the hope that it will help make future articles less convoluted.

ABC (both rejection and SMC) approximates likelihood in the following ways: Firstly, exact conformity to the the predictions of the model (theta) is relaxed using a distance measure and threshold. Secondly, if the simulation linking cause (theta) to observation (y) is stochastic then it uses a finite set of MC realisations instead of an integral over all outcomes of the random variables not of interest.

Neither of these properties are used in the work presented here. Nothing of what separates ABC from the traditional Bayesian approaches is used here. That is not to say that it cannot be cast into ABC terms but rather that it is ABC only in the most superficial sense. The proposed approach could simply have been presented as applying a probabilistic constraint to an ensemble using rejection sampling and an SMC variant of it. This would have saved readers from a lot of irrelevant reading.

The topological constraint is computed from an 'initial' adjacency graph which is not observed but derived subjectively from interpreted seismic data (for the real data case). In common Bayesian practice this constraint would therefore be called a prior or empirical prior (if you want to emphasise that some data did inform it), not a likelihood. The use of distance measures to defines such priors is common in traditional Bayesian methodology and needs no appeal to more recent trends to be explained or justified. I do not have any issue with the subjective nature of deriving this as it is an unavoidable part of all inference, Bayesian or frequentest. However, the author claims to circumvent specification of a 'likelihood function'. In this case there was no need, the implicit 'likelihood' (empirical prior constraint on topology) is simply a uniform distribution on all geomodels not conforming to a binary topological constraint centred on the initial graph. It could have been specified simply as a typical prior constraint used in any traditional Bayesian application without the need to force an ABC interpretation. This would also have made the implicit assumptions more transparent. In short, the method could simply be described as using rejection sampling (much older than ABC) to apply a uniform empirical prior (also much older than ABC) on topology graphs of geomodel ensembles.

Whether the volumes (L,F & T), or connectivity graph, or adjacency graph, or its Jaccard index are considered y or S(y) is completely arbitrary here as none of these are directly obeserve in the experiments. The ABC distinction between y and S(y) doesn't aid anything in this particular application.

A more concerning problem is the author's misrepresentation of what the term 'likelihood-free' entails. The claim that specification of a 'likelihood' is circumvented is not true. Nor does this particular application simplify the process of specifying this probabilistic constraint any more than would be the case for any typical definition of a prior over model parameters in geoscience. As mentioned before, the implicit 'likelihood' used here is trivial and easy to formulate. This might not be the case for many other applications of ABC but it is here. I don't believe that the author intended for this

to be interpreted as ABC reducing the need for subjective assumptions but due to how things are phrased, many readers will take away exactly that message.

Specific Comments:

1) Remove claims all of circumventing or simplifying specification of topological knowledge due to ABC, these are untrue for the constraint presented here. I have highlighted these in the attached annotated pdf along with more detailed comments for each. Lines 60, 150, 155, 165.

2) Remove all uses of the term 'likelihood-free'. There is no place in this article where its use helps clarify how the proposed approach works. Its only effect is as a potential source of misinterpretation. To avoid unneeded additional review rounds I am asking for complete removal and not fixing its use.

3) Do not refer to the initial connectivity graph as an observation. It is a semi-subjectice semi-empirically derived parameter to a subjectively chosen constraint family. Describe briefly how it is obtained and what the reasoning behind using the 'initial' graph was.

4) I am not asking that you replace 'likelihood' with 'prior' when referring to your topological constraint. Instead, please state somewhere that you choose to go with this label but that it could also be considered a prior or empirical prior and that the application of these terms is not always clear cut. State that you are simply treating the adjacency graph (y) as an observation.

5) Two topology distance measures are defined in section 2.4 which are never used. Since they are never discussed, analysed or compared, they serve no purpose. I suggest you remove them to simplify and shorten the already long paper, but feel free to ignore this suggestion.

6) Remove the mention of fuzzy sets in section 2.5, it is not relevant. Your posterior is probabilistic not fuzzy. It represents degrees of certainty concerning a single underlying truth, not degrees of membership to a category.

7) Several sentences need to be reworded for clarity or readability. I have highlighted these in the annotated pdf. Please try to address most of them.

Please also note the supplement to this comment:
https://gmd.copernicus.org/preprints/gmd-2020-136/gmd-2020-136-RC2-supplement.pdf

---

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-136, 2020.

Printer-friendly version

Discussion paper