Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2018-315-RC2, 2019

Geoscientific Model Development Discussions

Interactive comment on "Surrogate-assisted Bayesian inversion for landscape and basin evolution models" *by* Rohitash Chandra et al.

Anonymous Referee #2

Received and published: 19 November 2019

The scope of this study is quite large, with many complex details included. For example, Bayesian inversion is described with on a complex model, an MCMC implementation is used and outlined which involves Parallel Tempering and also the parallel architecture of the authors code is also described in some detail. These are all relevant, if not original features of the author's computer code, and it seems there is an intent is to describe them all. While completeness is a good thing, and the intent is appreciated here, it can tend to obscure the main focus which is to evaluate the effectiveness of combining surrogate 'pseudo-likelihoods' which are trained 'in-situ' during the sampling process. The authors may improve the effectiveness of the manuscript by describing some, or most, of these algorithmic details in appendices, using the main text to give an overview and clearer focus on the primary point, i.e. the comparison between surrogate

and full forward modelling in the Bayesian sampling.

Having said that I think the authors should be commended for attempting to include full detail, which is appreciated. Since the primary focus here (should be) on the surrogate model, I found the section which described this, and both approaches for training, rather light. I didn't get a clear enough understanding to give me confidence that I could reproduce it. Given the central importance of this aspect I suggest an appendix devoted to describing the structure of the Neural network and its training be described carefully in an appendix. That is outside of the context it is being used, i.e. without reference to MCMC or PT or even Figure 4. The readers should have a clear picture of this as a 'stand alone' component of the overall algorithm.

All figures and tables need much better captions. They appear to be an afterthought. Variables, axes and details of the figure need to be explained or define. I suggest even including a hint at what you want the reader to notice in each figure/table. At present they are just titles.

Section 2.1 needs to be re-examined. E(x) is not defined, nor is W_L . Presumably eqn. 920 is the MCMC balance condition. If so where is the prior ratio, where is the proposal ratio. If it is assumed that these cancel this needs to be specified and explained.

The use of terms local and global need clarification, as far as I can tell it refers to things that happen on a parallel compute node compared to the master. Correct? Please explain. Not clear the distinction needs to be made.

There a reasonably large number of choices that need to be made for control parameters throughout, intervals of PT and surrogate, sizes of training sets, starting value sof Neural network etc. Several of which are listed on page 12 lines 25-30. We are told that 'All of these values are determined experimentally', but how? While I trust that the author has done a competent job, we still need to know what criteria were used to decide between desirable and less desirable values? Some explanation is required.

There are numerous typographical and grammatical errors throughout the manuscript which creates a poor impression. Below are a few I identified, but there are sufficiently regular to warrant a careful proofread by an independent person in any future version. I suspect this has not happened prior to submission. In some parts the text descends into obscure technical detail regarding data flow in the parallel structure, etc. Again independent feedback from a colleague might sort these issues out.

The repeated use of the term 'replica' is confusing. This appears to be describing unrelated models (sets of variables) at the time step of an McMC chain. In what sense are they replicas? My understanding is that the only thing in general any two 'replica's have in common is the same chain index.

I was unsure what the actual numbers of unknowns and what the typical compute cost of a Likelihood evaluation were in each experiment. It would be best to explicitly state this in each case, as it puts the calculations and MCMC sampling into perspective. I did not get a clear picture of this.

In page 10 line 20, it may be useful to mention that these are what is known as hierarchical MCMC models, as the variance of both data types are being treated as unknowns. This is not a good aspect I think but one that is glossed over.

The bottom line message from this manuscript as I understood it was that across several examples shown, both the computationally inexpensive 'Continental-Margin' and 'Synthetic-Mountain' cases as well as the more computationally demanding 'Tasmania' case there are time savings of between 7 and 65% when using the surrogate over the full forward model. Necessarily these numbers depend on details of tuning various control parameters and other choices made, and I assume a good job has been done. However whether this is of practical significance is not clear. If I had a computer that was three times as fast as the one used here then presumably I would achieve the same compute time as the surrogate with the more accurate full physics based model. Correct? While I think a saving has been demonstrated, the author should

really comment on the significance of the observed improvement in compute time.

As the author clearly points out well, the improved efficiency of the surrogate-assisted MCMC sampler comes at the cost of lower accuracy as measured ultimately in the Bayesian mean and standard deviations of the Elevation and Erosion-Deposition parameters. As I understood it the PT-Bayeslands results are considered the 'near truth' and the Surrogate-assisted, SAPT-Bayesland, as the approximate. So perhaps the more important question, is then is how to judge whether the trade-off of accuracy against compute time is significant. One way this might be done is ask whether the PT-Bayeslands could produce the same if not better accuracy than SAPT-Bayesland with the same computation budget, i.e. fewer samples. I assume it is possible to do such an experiment by rescaling the number of samples available to PT-Bayeslands by the relative compute times observed in the experiments. This question/experiment has not been addressed but it would be instructive to try it. Again the central question is one of significance of the results. It would be impressive for the reader to see some attempt along these lines.

Overall I think this is an encouraging piece of work which could be significantly improved by a restructured manuscript and more quantitative evaluation on the two points above.

Some typo and grammatical errors: P1 L10: 'has been with successfully' - ? P8 Figure 3b is missing? P11 Last sentence starting 'In our case,. . .' contains 'giving by the sampler' meaning? I did not actually understand this sentence at all. P12 L3 'paralle' P12 L12: 'the the true Likelihood' P12 L18 Is this the Gaussian Likelihood or the log-Likelihood? P13 L30 'they have imitations in training' ? P15 eqn (8) balance size of brackets. P15 L11 define 'J(W,b)' P16 L11 'for for' P16 L22 'we needs' P16 L29 where is footnote 3? P17 L23 'Howsoever' ? P18 L10 'We by notice that' P18 L13 what is surrogate probability? Do you mean accuracy in recovering marginal probability?

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-315,

2019.

C5