Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2018-315-AC2, 2019

Geoscientific Model Development Discussions
Open Access
EGU

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

Rohitash Chandra et al.

c.rohitash@gmail.com

Received and published: 25 November 2019

The authors sincerely thank the reviewer for these comments. In the revision, we have created an Appendix section that features the details of parallel tempering MCMC and Training algorithms for neural network surrogate model.

We are including python code with the paper along with data and sample results in order to ensure reproducibility. We also revised the Algorithm 1 to ensure that we make the method clearer and have amended the texts in these sections (highlighted in light brown) to ensure that all the details are presented clearly.

Comment: 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.

Printer-friendly version

Discussion paper

Correct? Please explain. Not clear the distinction needs to be made.

Response: This has been added: “In surrogate-assisted parallel tempering, global surrogate essentially refers to the main surrogate model that features training data combined from different replicas running in parallel cores. Local surrogate model refers to the surrogate model in the given replica that incorporate knowledge from the global surrogate in order to make a prediction given new input data (sample of proposal). Note that the training only takes place in the global surrogate and the prediction or estimation for pseudo-likelihood only takes place in the local surrogates. “

Comment: 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.

Response: The following has been added in the discussion section: “ In general, the proposed method achieves a lower prediction accuracy when compared to PT-Bayeslands. However, given the cross-section visualization, we find that the accuracy given in prediction by the surrogate based framework is not so poor. Moreover, application to a more computationally intensive problem (Tasmania) shows that a significant reduction in computational time is achieved. We demonstrated the method using small models that run in seconds or minutes, Computational costs of continental scale Badlands models is very large (5 kilometer resolution for Australian continent for 149 million years is about 72 hours) and hence, in case when thousands of samples need to be drawn, the use of surrogates can be very useful. However, we note that improved efficiency of the surrogate-assisted Bayeslands comes at the cost of lower accuracy and there is a trade-off between accuracy and computational time.”

Comment: 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.

Response: The results in Table 7 and 8 show computation time reduced and RMSE accuracy, with a fixed number of samples to provide a fair comparison. In our previous work, we have already shown the performance trend of PT-Bayeslands given different number of samples (Chandra et. a, 2019) https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GC008465

Furthermore, the following has been added in the discussion: “The results in terms of prediction accuracy given by the proposed method can be further improved in future work with the way the surrogate is trained. Rather than a global surrogate model, local surrogate model on its own can be used, where the training only takes place in the local surrogates by only relying on history of the likelihood and hence taking a univariate time series prediction approach using neural networks. Our major contribution is in terms of the parallel computing based open-source software and the proposed underlying framework for incorporating surrogates, taking into account complex issues such as inter-process communication. This opens the road to try different types of surrogate models while using the underlying framework and open source software. “

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