Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2018-327-RC2, 2019

Geoscientific Model Development Discussions

Interactive comment on "Efficient surrogate modeling methods for large-scale Earth system models based on machine learning techniques" *by* Dan Lu and Daniel Ricciuto

Tianfang Xu (Referee)

tianfang.xu@usu.edu

Received and published: 25 February 2019

The manuscript by Lu and Ricciuto proposed a computationally efficient strategy to construct fast-to-evaluate surrogate of a large-scale Earth system model using neural networks (NN). The strategy uses singular value decomposition (SVD) to reduce the dimension of outputs, which are spatiotemporally correlated, and Bayesian optimization techniques to select NN hyperparameters. In this way, a surrogate model with satisfactory accuracy can be constructed with as few as 20 runs of the Earth system model.

I found the manuscript overall well written, experiments are clearly described, and re-

sults are clearly presented. I suggest minor revision based on comments listed below, focusing on improving clarity of text and adding necessary citations. I also made a suggestion regarding publishing data from this study.

1. It is not clear from the manuscript the advantage of (1) using a single surrogate model for all model outputs (the strategy taken in this manuscript) versus (2) constructing one surrogate for each model outputs (e.g., Xu et al., 2017). As mentioned in Introduction, (1) leads to very large neural network and storage and operation of large matrices. Using (2) will lead to a very large number of NN models. However, each NN can be very simple, and its evaluation time will be negligible. Which method is more computationally efficient may depend on specific case. I suggest to add some discussion about the tradeoff.

2. For the case study of 8 parameters, it was found that 20 model runs suffice. How would the needed number of model runs scale with increasing number of parameters? This could be an advantage of the proposed methods over existing methods such as gPC.

3. Would the surrogate modeling performance depend on the location of the 20 parameter sets (training data)? It will be helpful to provide some guidance on how to select parameter sets to generate training data.

4. It is argued in several places that a slight change in hyperparameters can result in dramatically different NN performance. I suggest providing a citation or two to support this statement.

5. It is not entirely clear to me how the Bayesian optimization of the hyperparameters is implemented. I suggest to include more details to help interested readers, for example, what is the prior and posterior.

6. Line 44-49: from the context data-model integration refers to calibration/uncertainty quantification. However in other context the term can also refer to data assimilation

and other methods.

7. Line 261: what does “information” refer to? Do you mean variance?

8. I wonder whether it would be possible to publish (on a data repository such as HydroShare, https://www.hydroshare.org/) the training and testing data generated by the sELM model, given that sELM is computationally expensive. This will provide a very interesting and representative dataset for the uncertainty quantification community. This is only a suggestion, and I leave it entirely for the authors to decide.

References Xu, T., Valocchi, A. J., Ye, M., & Liang, F. (2017). Quantifying model structural error: Efficient Bayesian calibration of a regional groundwater flow model using surrogates and a data driven error model. Water Resources Research, 53(5), 4084-4105.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-327, 2019.