Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-17-AC2, 2019

Geoscientific Model Development
Discussions
Open Access

Interactive comment on "What do we do with model simulation crashes? Recommendations for global sensitivity analysis of earth and environmental systems models" *by* Razi Sheikholeslami et al.

Razi Sheikholeslami et al.

razi.sheikholeslami@usask.ca

Received and published: 2 August 2019

This document contains copies of all comments of the Reviewer 2 and our planned efforts to address them.

Reviewer 2: The authors argue to substitute data of failed simulation members in large ensemble simulations conducted for global parametric sensitivity analysis of dynamical earth system models. It is common for the models to crash for certain parameter value combinations that are randomly sampled from multidimensional parameter space using

standard automated techniques. Using case studies, the authors show that it may be better to fill in the data from the failed experiments with data substitution techniques rather than the general practice of ignoring those experiments completely. The paper is generally well written and motivated. I point out my concerns below.

Response: We greatly appreciate Reviewer 2 for reviewing the manuscript and providing positive evaluations.

1. The authors motivate the study well (Section 1.2). However, the authors state that the automated sampling method that they use - STAR-VARS breaks down if there are failed simulations for certain parameter combinations (Section 2.3). They do not provide a good reasoning for that, which I think is warranted. Are there other sampling methods that would not be sensitive to failed simulations? Why use STAR-VARS? Is the data substitution strategy only designed because of the limitation of STAR-VARS?

Response to 1: Thanks for this comment. As we mentioned in Section 1.2, those GSA techniques that use a sampling strategy with a specific structure will fail if the simulation model crashes at certain parameter configurations such as the widely-used variance-based method of Saltelli et al. (2010). To further explain, we have revised Section 1.2 as follows:

"Ignoring the crashed runs in GSA is only relevant when using purely random (and independent) samples (i.e., Monte Carlo method). In such cases, if the model crashes at a given parameter set, one can simply exclude that parameter set or repeat randomly generating a parameter set (at the expense of increased computational cost) that results in a successful simulation."

"Reducing the number of model runs and finding optimum locations for sample points in the parameter space are the main drivers for implementing improved sampling techniques in GSA. Typically, these sampling techniques are not only random but also follow specific spatial arrangements. However, when applying a sampling-based technique that uses an ad-hoc sampling strategy with particular spatial structure (e.g., the variance-based GSA proposed by Saltelli et al. (2010) or STAR-VARS of Razavi and Gupta (2016b)), we cannot ignore crashed simulations. In this case, excluding sample points associated with simulation crashes will distort the structure of the sample set, causing inaccurate estimation of sensitivity indices. As a result, the user may have to re-do a part or the entire experiment depending on the GSA implementation, by generating a new sample set (or a succession of sample sets), leading to a waste of previous model runs."

2. The impact of three data substitution techniques are compared. However, the first two methods are overly simplistic, and one can argue that they would yield poorer results a-priori - for example, the median is definitely not a good approximation for parameter combinations that are in the distribution tails, which may be more likely to crash. I do not see why the authors chose to present the results from those methods as one of their main results. It is fine to include them, but I think it would have been more useful to include results from different surrogate models, e.g. kriging, neural networks etc., which may be better as models for data substitution.

Response to 2: We certainly agree with reviewer that the median substitution is a very simple (perhaps naive) approach compared to other methods such as RBF. Nevertheless, we have adopted these methods considering our main goal in this study, which was finding simple and effective strategies for handling simulation failures. To improve the explanation, we have added the following statement to Section 2.2.1:

"In sampling-based optimization, assigning a very large objective function value for the parameter configurations violating the problem constraints is a common practice (known as the big M method). Inspired by the big M method we used the median substitution technique in this paper. However, since replacing crashes with a big value can magnify the effect of the crashed runs in GSA, it is reasonable to choose a central value such as median to minimize the impact of the implausible parameter configurations."

Regarding the application of other surrogate models (e.g., kriging, etc.) as we men-

tioned in Section 2.2.3 depending on the complexity and dimensionality of the response surface, other types of metamodels can be incorporated into the proposed framework. We did not intend to compare the performance of different metamodelling techniques in this study, so we only applied the well-known RBF technique. Of course, this could be a potential direction for future research. To address this comment, we have added the following sentence to the conclusion section of the revised manuscript:

"In addition, future work may include application and testing of other types of emulation techniques such as kriging and support vector machine to handle simulation failures."

3. The authors appear to consider simulation failure as numerical artefacts. It could well be that parameter combinations are unphysical resulting in genuine crashes. Substituting data for these model crashes would result in unrealistic sensitivity. Likewise, unrealistic parameter combinations could also result in successful runs without crashes distorting the sensitivity analysis. It will be good if the authors could discuss this. The authors discuss this partly in section 5.1 for MESH model while exploring the reasons of simulation failure, but do not seem to relate it to their substitution strategy which is their main point.

Response to 3: Thanks for this very good comment. The following paragraph has been added to Section 5.1 of the revised manuscript to improve the discussion:

"We conclude this section by highlighting a point that should receive careful attention when applying the substitution-based methods in handling model crashes. In addition to the numerical artifacts in simulation models, some combinations of parameter values, which are not physically valid, can also lead to simulation failures. As a result, substituting data for these crashed runs would result in unrealistic parameter sensitivity. Preventing this type of failure requires knowledge about the reasonable parameter ranges in DESMs. This type of crash can be significantly reduced by selecting plausible ranges of parameters based on physical knowledge or information of the problem (a process referred to as “parameter space refinement” (see e.g., Li et al., 2019;

Williamson et al., 2013)). However, DESMs often consist of many interacting, uncertain parameters, and therefore very little may be known a priori about the implausible regions of the parameter space. "

4. The title reads as if something useful can be done with simulations that crashed. But, the strategy of the paper is to actually substitute the failed simulations. The authors should think about revising the title so that its not too misleading.

Response to 4: We greatly appreciate the reviewer for this valuable suggestion. The new title now reads:

"What should we do when a model crashes? Recommendations for global sensitivity analysis of earth and environmental systems models"

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-17, 2019.