Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-17-RC2, 2019

Geoscientific Model Development Discussions

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.

Anonymous Referee #2

Received and published: 28 June 2019

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.

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?

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. krigging, neural networks etc., which may be better as models for data substitution.

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.

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.

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