Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2017-271-AC1, 2018

Geoscientific Model Development Discussions

Interactive comment on "Fast sensitivity analysis methods for computationally expensive models with multidimensional output" *by* Edmund Ryan et al.

Edmund Ryan et al.

edmund.ryan@lancaster.ac.uk

Received and published: 4 May 2018

Reviewer's comment #1: The authors present several estimators for the Sobol' indices. They may consider [1] where the "most efiňĄcient formulas available today..." for Sobol' index estimation is described.

Author's response #1: I'm happy to include [1] in the revised manuscript. However, I am having problems accessing it. I have asked for my university library to purchase the 'Handbook of Uncertainty Quantification' book, so assuming it has arrived before the deadline to submit the revised manuscript, I will include a reference to [1]. Changes to be made in manuscript: As stated above, assuming it has arrived before the deadline to

submit the revised manuscript, I will include a reference to [1] in the revised manuscript.

Reviewer's comment #2: What is being plotted in ïňĄgures 3 and 4? Based on the magnitudes, I assume this is the numerator of the Sobol' index, i.e. Var(E[f(X)|Xi]). Did you have to rescale anything to compare results from the different methods?

Author's response #2: These are the sensitivity indices (SIs) calculated using the five different methods (a-e in both figures). Using the variational methods (Sobol, eFAST and GAM) the SIs are computed using Si = Var(E[f(X)|Xi]) / Var(f(X)), so no rescaling needed. The formulae for computing the SIs are given in the methods section (eqn. 2 for the Sobol method, eqn. 1 and 3 for the eFAST method, eqn. 1and page 13 for the GAM method, and page 13/14 for the PLS method). Changes to be made in manuscript: I will amend the captions for figures 3 and 4 to make it clear what formulae or methods are used to compute the sensitivity indices which are graphically shown in these two figures.

Reviewer's comment #3: How did you reconstruct the spatially distributed sensitivity indices in ïňĄgures 3 and 4 from the PCA? Based on comparing the methods you clearly did it correctly; it would be nice to be a bit more explicit about this.

Author's response #3: Yes, this is a good point. In non statistical terms, using PCA for this purpose is a little bit like zipping a file to make it smaller in size and then unzipping it when you want to use the file again. Here, we go from an output dimension of 2000 (e.g. 2000 modelled ozone values at different latitudes / longitudes) to a dimension of say 5 (the first 5 principal components). When we compute the SIs using the eFAST method we need to run the emulator 5000 times for each of the 2000 model outputs. Using PCA, we run the emulator only 5000 times for each of the 5 transformed outputs. For each of the 5000 emulator runs, we can reconstruct the map of 2000 model outputs from the 5 transformed outputs. It's a little long to explain here, but in the methods I will add further detail to make this clear. Changes to be made in manuscript: In the PCA part of the methods section I will include extra detail about how the 2000 model outputs are recovered from the first 5 PCs.

Reviewer's comment #4: As mentioned in [2], it is frequently useful to have a scalar sensitivity instead of a spatially distributed one as in ïňĄgures 3 and 4. How can your results be “averaged” in space to provide one scalar sensitivity for each parameter?

Author's response #4: In a paper currently in preparation, Oliver Wild (my collaborator and line manager) will be presenting results to a sensitivity analysis where one of the outputs is global methane lifetime (i.e. the methane lifetimes presented here but just as one number). The sensitivity analysis in that paper has the same inputs and the same training runs as this study, so to avoid repetition of results I will not show these sensitivity analysis results in this paper. I will refer to his paper though in the revised manuscript. Changes to be made in manuscript: In the revised manuscript, I will talk more about the findings of the Wild et al. (in prep.) paper mentioned above.

Reviewer's comment #5: Two approaches are considered for constructing a meta-model for a spatial dependent output. One is based upon constructing a meta-model for each point in space, and the other is based upon constructing a meta model for each PCA mode. Would it be possible to construct a meta-model which is learned to predict all points in space simultaneously? In this case, it would be a function from Rn to Rm where n is the number of parameters and m is the number of model grid points. I could imagine training a neural network to learn this function. How would this approach compare with the methods of this article?

Author's response #5: Yes, this is possible and a neural network approach would work. The problem with neural networks is that they need a lot of training data, of the order of 1000s. The main reason for using a Gaussian process emulator is that it even with not many model runs to train it (80 in this study but in general < 200 for a high number of inputs), it has been shown in lots of settings how it can robustly and accurately approximate the input-output relationship of the computationally expensive model. Changes to be made in manuscript: In the revised manuscript, I will justify the reason for not

considering a neural network approach as part of this study.

References: [1] Clementine Prieur and Stefano Tarantola. Variance-based sensitivity analysis: Theory and estimation algorithms. In Roger Ghanem ,David Higdon ,and Houman Owhadi, editors, Handbook of Uncertainty QuantiïňĄcation. Springer, 2017. [2] Amandine Marrel, Nathalie Saint-Geours, and Matthias De Lozzo. Sensitivity analysis of spatial and/or temporal phenomena. In Roger Ghanem, David Higdon, and Houman Owhadi, editors, Handbook of Uncertainty QuantiïňĄcation. Springer, 2017.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271, 2017.