Review of the Manuscript number GMD-2018-4.
Title : Bayesian inference of earthquake rupture models using polynomial chaos expansion.
Authors : H. Cruz-Jiménez, G. Li, P.M. Mai, I. Hoteit, O.M. Knio.

General comments

This manuscript investigates an earthquake rupture model subject to 7 random fault plane properties. Polynomial chaos surrogates are built and validated to reproduce the uncertain Peak Ground Velocity (PGV), obtained from a discrete wavenumber/finite element method, at a set of 56 (virtual) stations. A sensitivity analysis is conduct to identify the main influent parameters : a partition of the uncertain input parameters into two groups highlights the strong impact of the hypocenter location. A Bayesian inference is then performed by using a Ground Motion Prediction Equation (GMPE) as observational measures. The results emphasize that additional physical constraints are valuable to increase the sampling efficiency.

The manuscript is clearly constructed and it would be suitable for the readership of the Geoscientific Model Development after the following revisions to clarify some aspects of the paper.

Specific comments

- page 6 : one sentence is missing between line 4 and 5 to provide the number of terms N_p in the PC series as a function of the stochastic space dimension n_d and the total polynomial order d, $N_p = (d + n_d)!/(d!n_d!)$.
- page 6, line 19 : the cross-validation process needs more details (leave-one-out or k-fold version, initial range of variation of the parameter γ with the discretization strategy to find the optimal value) with a citation (e.g. the book of Seber and Lee, Linear regression analysis, 2003).
- page 7, section 3.1 : the computation of the empirical error (8) with the training set $\mathcal{P}_{LHS}$ (blue dots) has only a minor interest because it simply shows that regression is a non-interpolating technique. A comparison between the empirical error estimated with the validation set (red dots) and a cross-validation error obtained with the training set is more relevant.
- page 8, line 12 (middle) : the sentence "The overall tendency of PC prediction uncertainty (...) seems to decrease with increasing R_{IJ} distance as well" relies on Fig. 6. This figure is hard to read and a new figure plotting only the (PC) standard deviations should be valuable (with a reminder in the text about the log-scale) to support the statement.
- page 8, line 16 (top) : two stations are selected for plotting the PGV. Their locations must be indicated (for instance with labels on Fig. 2).

- page 8, line 1 (middle) : The first sentence of the paragraph is incomplete since the complex dependency of PGVs to random inputs is not only due the mappings between the physical parameters and the standardized RVs $\{\xi_i\}_{1\leq i\leq 7}$. We can speculate that the complexity of the propagation model (discrete wavenumber/finite element method) plays a major role.
- page 11 : in Fig. 6, the GMPE standard deviation exhibits a higher level than the PC ones. A short discussion would be interesting to explain the causes/sources of this difference.
- page 13 : a prediction error, defined as the discrepancy between the GMPE and PC series is introduced. This is confusing in Bayesian inference framework where observations (or measured data) are used to infer the model parameters. As GMPE predicted PGVs serve as observational data (see page 11), it would be more clear to replace GMPE by observational data (and to replace prediction error by observational error) in section 4.1.

Technical corrections

- page 2, line 9 : replace is by are in "data is sufficient".
- page 2, line 16 : replace Mw 65 by magnitude 65.
- page 5, Table 2, line 3 : replace y_h by z_h.
- page 6, line 18 : "that" is missing, "note that [Ψ] is station invariant".
- page 8, line 6 (top) : the word "indeed" is useless.

Suggestions

- page 5, line 11 : "number of stochastic dimensions" sounds weird. "stochastic space dimension" or "number of uncertain input parameters" are more usual.
- page 5, line 16 : "instead of" seems to be inappropriate here and could be replaced by "which parameterize".
- page 6, line 13 : the set of LHS realizations could be written, "... $N_{LHS} =$ 8000 earthquake rupture model realizations through $\{\boldsymbol{\xi}_k\}_{1\leq k\leq N_{LHS}}$".
- page 8, line 16 : replace "with different PC truncation orders" by "with increasing odd PC truncation orders up to a degree nine".
- page 8, line 17 : replace "PC library is sufficient ..." by "PC expansions are sufficiently accurate ...".
- pages 9 and 10 : Fig. 4 and 5. represent distributions obtained with kernel density estimation. It should be mention in the captions or in the text.
- page 11, line 5 : Move the group of words "for the same magnitude and focal mechanism" in section 3.2 (page 8), line 10 after the reference Boore and Atkinson (2008).
- page 13 : explain a little bit more the partitioning of the data into four concentric groups (e.g. uniform discretization of the R_{JB} interval).
- There is a huge number of ground motion predictions equations (see for example the report `http://www.gmpe.org.uk/gmpereport2014.pdf`). A short description of the GMPE model (for instance in an appendix) could be worthwhile to have a self-contained paper.