Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2018-4-RC2, 2018

Geoscientific Model Development Discussions
Open Access
EGU

Interactive comment on "Bayesian inference of earthquake rupture models using polynomial chaos expansion" *by* Hugo Cruz-Jiménez et al.

Anonymous Referee #2

Received and published: 26 March 2018

General comments

The authors develop a polynomial chaos (PC) expansion representation to provide a surrogate model for a probability distribution of Mw 6.5 strike-slip earthquakes with a fixed fault geometry. Seven parameters are used to describe a particular realization, including the hypocenter location and parameters describing an elliptical asperity, a region of relatively high slip, defining a 7-dimensional stochastic space. The surrogate model allows the rapid estimation of the peak ground velocity (PGV) at each of 56 virtual observation points. The PC expansion is computed using synthetic seismogram observations at these points for a set of 8000 realizations. A second set of 8000 realizations is used for validation, to confirm that the surrogate model constructed from the first set agrees well with the direct simulation results for the second set of realizations.

Printer-friendly version

Discussion paper

The surrogate model is then used to rapidly compute the PGV for millions of additional realizations in order to gather statistics on the decay of PGV with respect to distance from the fault (measured using the Joyner-Boore distance R_{JB}, the minimal distance to the fault plane as projected to the surface), at the 56 observation points. The mean PGV and standard deviation at each observation point are plotted vs. the distance R_{JB}, and this data compared with the ground motion prediction equation (GMPE) of Boore and Atkinson (2008). The GMPE was derived based on observations of past earthquakes and so it is interesting to see that the statistics generated by the PC expansion generally follows this prediction and lie within one standard deviation of the GMPE as determined by Boore and Atkinson. This suggests that a simplified fault model consisting of a single asperity and a small set of parameters can perhaps predict PGV statistics well, and hence may be useful for predicting other GMPE curves, or for probabilistic seismic hazard analysis more generally. The first 3 sections of the paper give a nice development of these ideas.

I had more trouble understanding the goal of Section 4, which concerns the use of Bayesian inference to determine a probability distribution on the space of PC parameters that yield an event to best match the GMPE. It seems to me that the GMPE is only intended to predict the average and standard deviation of the PGV over a large set of potential earthquakes, and so I do not understand the point of this statistical inversion to try to determine the characteristics of one particular earthquake that best matches the average. The authors conclude that the best match is more likely to have the hypocenter located in the lower right quadrant of the fault plane, and the elliptical patch centered in the lower left quadrant. Why is this useful to know? Is this meant to have geophysical significance, e.g. that real strike-slip earthquakes of this magnitude tend to have their hypocenter and asperities located in this way? How does this relate to the actual slip patterns of the real events that went into the Boore and Atkinson GMPE model, to the extent those are known? There is no discussion in the paper of these topics. I also wonder about the way this inversion is used in Section 4.5, as discussed in one of my specific comments below. I think the paper would be stronger if the motivation for doing this inversion was better explained, since I found it hard to assess the usefulness of this part of the paper.

Specific comments

1. Page 3, line 2: The fault plane geometry is fixed with width 10 km and length 27 km. It is stated that this is obtained from 100 realizations following the scaling relation in Wells and Coppersmith (1994). How are 100 realizations used to determine these dimensions?
2. Page 3, lines 5–7: Why is the slip set to S_{max}/e outside the asperity? How is the slip in the asperity set? Since the area of the asperity varies with the input parameters, the slip must also vary to keep the magnitude fixed. It is stated that S_{max} varies with the ellipse size but it is not clear how.
3. Page 3, line 15–17: For completeness it would be good to state the grid resolution used in the COMPSYN simulation of the seismic signals, and the domain size, boundary conditions imposed, etc.
4. Page 3, Figure 2: The 56 observation stations surround the fault plane on all sides. Since the fault plane is vertical and the velocity model is vertically layered, shouldn't the observations be symmetric about $Y = 0$? If so, it would seem clearer to simply use points in the upper half plane, for example, rather than asymmetric points scattered on both sides.
5. Page 11, Figure 6: The points here are presumably the mean PGV observed at each of the 56 observation points, plotted vs. the distance R_{JB}. These points are calculated by evaluating the PC expansion at 1,000,000 sample points and are presumably quite accurate estimates of the mean at each observation point. But this figure shows that two points that have very similar R_{JB} can have quite

different PGV, presumably because the two points have quite different azimuthal orientation relative to the fault, even though they are the same distance away. This is interesting to observe, but since the GMPE curve ignores orientation it seems like it might also be interesting to try to average over different orientations for each distance. This could be facilitated if a number of observation points were placed at each distance, for a discrete set of distances, i.e., place the observation points on concentric rings with fixed R_{JB}. It also seems like a much larger set of observation points could be used than 56, since the PC model is so quick to evaluate. If many points were placed on many different concentric rings, then one could average over all points at a given distance to get points that might be expected to agree better with the GMPE curve in Figure 6. It would then also be possible to explore in more detail how the PGV varies with orientation along each ring.

6. In Figure 2 there are sets of points that have different colors/symbols that are arranged somewhat in rings, but the distance for each color do not seem to be constant. The use of colors/symbols is not explained anywhere I could find, and should be.

7. Page 5, Table 2: The caption says that “(*) denotes dependent parameters”. It is not clear what this means. Does this refer to the comment in line 5 of this page, where it is noted that “These restrictions lead to nonlinear dependency between feasible ranges of different physical parameters”?

8. The fact that some of these parameters are constrained based on the choice of other parameters means that the probability distribution of parameters is not really given by (1) on page 5 as is stated. Some choices from this 7-dimensional box have probability zero due to the constraints, while others have greater probability due to several non-allowed choices mapping to the same set of modified parameters when the asperity falls near the edge of the fault plane. Does this

affect the validity of the PC expansion and/or results? At any rate, this should be discussed.

9. Page 17, Section 4.5: In this section it is stated that a uniform distribution of parameters over the 7-dimensional space ignores various geophysical constraints suggested by previous work. This is discussed in the context of choosing a prior for the Bayesian inference, but it seems like it would be even more important to incorporate these constraints into the analysis of Section 3, where the PC expansion is used to generate statistics on the PGV for comparison with the GMPE. Why should the statistics obtained with the uniform distribution be expected to match the GMPE well if it is known that this is the wrong distribution? This is addressed to some extent in Section 4.5 where the inversion that incorporates these constraints is then used to generate statistics that are compared to the GMPE curve in Figure 15. But at this point the inversion process has been used to to further constrain the posterior distribution based on trying to match the GMPE curve, so comparing the result to the GMPE curve does not seem to provide any validation that the PC expansion could predict the GMPE curve for other scenarios, for example. I may be missing the point here, but I think it needs more explanation.

Technical corrections

1. Page 3, line 2: Presumably the rake is fixed at 0 degrees for a strike-slip event, but this should perhaps be stated?
2. Page 7, line 27: What are the index sets $\mathcal{S}_i$ and $\mathcal{T}_i$? The sets are used in the summations of (7a) and (7b) respectively, but not really defined.
3. Proper latex fonts for trig functions should be used in expressions such as (A1), e.g. $a \cos \beta$ rather than $acos\beta$.

Printer-friendly version

Discussion paper

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-4, 2018.