Efficient Calculation of PSHA with Epistemic Uncertainty in the Ground-Motion Model

Work Package 5

AUTHORS		REVIEW		APPROVAL	
Name	Date	Name	Date	Name	Date
M. Lacour and N. Abrahamson	*2019/04/22*	[illegible] Ian Main [illegible]	*YYYY/MM/DD* 2020/10/19	Albert Kottke	*YYYY/MM/DD* 2020/12/21

__DISSEMINATION__: This document must not be distributed to any person, institution or company other than members of SIGMA-2 steering and scientific committees, except under written formal permission of SIGMA-2 steering committee.

Document history

DATE	VERSION	COMMENTS
2019/04/22	*0*	*Initial document. This has been submitted as a paper to the BSSA*
2020/10/14	*1*	*Revised to address reviewer comments and added section on partially correlated GMMs*

Executive Summary

In PSHA, the epistemic uncertainty of the ground-motion model (GMM) is traditionally included using logic trees with a small set of alternative published ground-motion models (GMMs) on the branches. The PSHA is then conducted for each of the GMMs on the logic tree. Recently, the published GMMs have begun to be treated as samples from a continuous distribution of possible GMMs. In the SWUS project, Sammon maps were used to discritize the continuous distribution of GMMs into a suite of about 30 representative GMMs, which were then used as the branches on a logic tree. The weights on the GMM branches were statistical weights based on the sampling of the continuous distribution.

PSHA practice is moving toward using non-ergodic GMMs in which the median ground motion for a given magnitude and distance varies as a function of the location of the earthquake and the location of the site. For the non-ergodic approach, use of a small number of alternative GMMs in the logic tree will not adequately capture the effects of the epistemic uncertainty in the location-specific source and path effects in the non-ergodic GMMs. Instead, hundreds of alternative GMMs will be needed to capture the epistemic uncertainty of the non-ergodic path terms from different ray paths through the crust. Using a large number of banches on the logic tree leads to a large increase in the computational time required for the PSHA.

To make the fully non-ergodic GMM method practical for PSHA applications, an efficient method to propagate the epistemic uncertainty in median ground motion is needed to capture the range of source and path effects for each source and site location. Rather than requiring faster and faster computers, we can dramatically speed up the calculations using the Polynomial Chaos (PC) approximation for propagating the effect of the epistemic uncertainty in the hazard due to the epistemic uncertainty in the median ground motion. Using the PC method, more accurate fractiles can be estimated in less time as compared to the traditional approach using logic trees with discrete sampling of the distribution for the median GMM. For example, using the PC method in the HAZ45 PSHA program, the computational savings using the PC method as compared to the traditional logic tree approach with 100 branches for the GMMs is about a factor of 40. In addition, the PC method leads to significant reductions in the memory requirements compared to a logic tree with a large number of branches for the median GMM: the memory reduction is the ratio of the number of GMM branches to the order of the PC expansion.

A limitation of the methodology presented in the first part is that the epistemic uncertainty in the median ground motion is assumed to be fully correlated across all scenarios: as in the scaled-backbone approach, if the median ground motion is above average for one scenario (M and R) it is also above average for all other scenarios. This assumption of full correlation does not affect the mean hazard, but it leads to broader fraciles. The assumption is removed in the second part of the report. For the NGA-W2 GMMs, the effect of using partial correleation rather than full correction is a narrowing of the fractiles, with the difference mainly for the lower fractiles. This effect could be different for other sets of GMMs. There is an additional computational cost of about a factor of 5 for including partial correlation.

INTRODUCTION

An important part of a probabilistic seismic hazard analysis (PSHA) is the incorporation of the epistemic uncertainty of the seismic source characterization (SSC) and ground-motion characterization (GMC) models in the hazard calculation. The epistemic uncertainty affects the mean hazard and also is used to show the limitations of the current earthquake science information for constraining the seismic hazard at a site. In current PSHA practice, the epistemic uncertainty in the GMC and SSC models are treated using logic trees. Each node of the logic tree has a small number of discrete branches with alternative models or parameter values.

Traditionally, the epistemic uncertainty in the median ground-motion model is captured using a small set of alternative ground-motion models (GMMs). Typically, the available GMMs are reviewed and those models judged to be applicable to the specific region and site of interest are selected as candidate GMMs. Bommer et al. (2010) describe an approach for selecting the alternative GMMs to include on the logic tree. A drawback of this approach is that it is limited to the set of published GMMs which represent a small sample of the full range of credible GMMs.

Recently, the published GMMs have begun to be treated as samples from a continuous distribution of GMMs. The scaled-backbone approach (Atkinson et al., 2014) is the simplest example of treating the GMM uncertainty range as a continuous distribution by defining a range of scale factors which are applied to a a single reference GMM. A limitation of the scaled-backbone approach with a single representative GMM is that it does not account for different scaling with magnitude and distance. To address this limitation, a set of alternative representative GMMs can be used, each with a range of scale factors.

To capture the epistemic uncertainty range of magnitude and distance scaling in GMMs as a continuous distribution, the Southwestern United States (SWUS) ground-motion model project, described in GeoPentech (2015), reparameterized each selected GMM using a single common functional form. While such common-form models cannot duplicate the original GMM for all scenarios, they can be optimized to be similar to the original GMMs for the scenarios that are important to the hazard at the site. Because the common-form GMMs have the same functional form, the covariance matrix of the common-form coefficients can be computed. Using Monte-Carlo sampling of the multivariate normal distributions for the coefficients, thousands of alternative sets of correlated coefficients can be generated that define the continuous distribution of the epistemic uncertainty in the median GMM. While this method is straight-forward, computing the hazard using a logic tree with hundreds of GMMs is computationally expensive and not practical for most applications. In the SWUS Project, Sammon maps were used to discretize the continuous distribution of GMMs into a suite of about 30 representative GMMs, which is near the current practical limit for a reasonable computational effort for large seismic hazard models.

PSHA practice is moving toward using non-ergodic GMMs (Anderson and Brune, 1999) in which the median ground motion for a given magnitude and distance varies as a function of the location of the earthquake and the location of the site rather than being the same for all site/source combinations in a region (e.g. Villani and Abrahamson, 2015). For the non-ergodic approach, use of a few GMMs in a logic tree will not adequately capture the effects of the epistemic uncertainty in the location-specific source and path effects in the non-ergodic GMMs. Instead, hundreds of alternative GMMs will be needed to capture the epistemic uncertainty of the non-ergodic path terms from different ray paths through the crust for all the site/source location combinations. As an alternative to a brute force approach of using large number of branches on the GMM logic tree, we develop a computationally efficient method for computing the epistemic uncertainty in the seismic hazard due to the epistemic uncertainty in the median ground motion. The method is applicable to both ergodic and non-ergodic GMMs, but the computational savings will be greater for the non-ergodic GMMs due to the larger number of logic tree branches required.

A complete GMM includes both the median and the aleatory variability (sigma) for a given earthquake scenario. This paper only addresses the effect of the epistemic uncertainty in the median on the epistemic uncertainty of the hazard for the fully correlated case. Refinements to the approach to use less than full correlation of the epistemic uncertainty of the median ground motion between scenarios and to include the effect of the epistemic uncertainty in value of the aleatory standard deviation will be addressed in subsequent studies.

PSHA NOTATION

The probabilistic seismic hazard is commonly written as:

$$Haz(z) = \sum_{k=1}^{Nsource} N_k(M_{min}) \int_M \int_R f_{m_k}(M) f_{r_k}(R|M) P(Y > z|M, R)\, dR\, dM \tag{1}$$

in which $Haz_(z)$ is the annual rate of the ground motion exceeding z at the site of interest, $Nsource$ is the number of seismic sources in the seismic source characterization model, $N_k(M_{min})$ is the annual rate of earthquakes above the minimum magnitude of interest for the k^{th} source, $f_{m_k}(M)$ and $f_{r_k}(R|M)$ are the probability density functions for magnitude and for distance conditional on magnitude, for the k^{th} source, respectively, and $P(Y > z|M, R)$ is the conditional probability of exceeding ground-motion level z, given an earthquake with magnitude M at distance R has occurred.

The focus of this paper is on the GMM, so we use a more compact notation for the source terms in the hazard integral based on the rate of the scenarios:

$$rate_{s,k} = N_k(M_{min}) f_{m_k}(M_s) f_{r_k}(R_s|M_s)\, dR\, dM \tag{2}$$

in which s is an index of the scenario. With this notation, the marginal hazard for a single earthquake scenario, (M_s, R_s), from the k^{th} source is given by:

$$Haz_{s,k}(z) = rate_{s,k} P(Y > z | M_s, R_s) \tag{3}$$

Most GMMs assume that, for a given earthquake scenario, the variability of the ground motion is lognormally distributed. The GMM gives the median, μ, and aleatory standard deviation, σ, of the ground motion for the given scenario. If μ and σ are both given in natural log units and the aleatory variability of the ground-motion parameter is assumed to be lognormally distributed, the conditional probability of exceeding the ground motion is given by

$$P(Y > z | M_s, R_s) = \left[1 - \Phi\left(\frac{\ln(z) - \mu_s}{\sigma_s}\right)\right] \tag{4}$$

in which $\Phi(\epsilon)$ is the cumulative distribution function of the standard normal variable given by:

$$\Phi(\epsilon) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\epsilon} e^{-0.5t^2} dt \tag{5}$$

The ϵ value is the number of standard deviations above the median to reach z:

$$\epsilon = \frac{\ln(z) - \mu_s}{\sigma_s} \tag{6}$$

UNCERTAIN MEDIAN GROUND MOTION

In PSHA, the traditional approach to addressing the epistemic uncertainty in the median ground motion is to use logic trees with multiple GMMs with weights on the branches of the GMM logic tree. In our approach, we replace the discrete distribution of the median ground motion for a scenario given with a continuous distribution. Specifically, we assume that the epistemic uncertainty in μ_s is normally distributed with a scenario-specific mean, $m_\mu(M_s, R_s)$, and epistemic standard deviation, $\sigma_\mu(M_s, R_s)$.

The epistemic uncertainty in the marginal hazard shown in equation (3) can be approximated using analytical methods such as Polynomial Chaos (PC). For a single scenario, the epistemic uncertainty in the median, μ, leads to fixed horizontal shifts in the hazard curve along the log ground-motion axis (Figure 1). The PC expansion is an expansion of the output of the hazard curves as polynomials that are functions of a standard normal random variable, ξ. In terms of probabilistic seismic hazard, the expansion is for the $P(Y > z|M_s, R_s)$ term. Conceptually, we want to find a function (sum of polynomials) that will approximate each of five alternative curves shown in

(Figure 1) given the number of standard deviations of the sampled epistemic uncertainty (here, -2, 1, 0, 1, 2). That is, we want to be able to predict the change in the y-axis values given the horizontal shift along the x-axis. If we input $\xi = -2$ into the polynomials, the function should follow the minus two standard deviation curve, and if we input $\xi = 1$ into the polynomials, the function should follow the plus one standard deviation curve. The coefficients of the polynomials will depend on z, m_μ, and σ_μ.

Mathematically, the polynomials used in the PC expansion can be thought as a vector basis of the space of epistemic uncertainty in the hazard. The key to the PC method is that uncertain hazard curves for the different scenarios are expressed as a linear combination of functions of the same random variable. This allows us to efficiently obtain the total hazard by simply summing the coefficients for each polynomial in the PC expansion over all scenarios. From the PC expansion of total hazard, we can approximate the full epistemic distribution of the total hazard (i.e., epistemic fractiles) by simply generating samples of the random polynomials used in the expansion.

This method is different from sampling many discrete branches of a logic tree, in which the entire hazard calculation is repeated for each alternative GMM. With the PC method, realizations of total hazard are generated only during post-processing. The uncertainty in the hazard calculation is propagated analytically through the PC expansion of each hazard curve. For logic trees with many branches, quantifying the hazard uncertainty using the PC expansion is much more computationally efficient than direct sampling of the distribution of the median ground motion used in current PSHA practice.

POLYNOMIAL CHAOS EXPANSION

The first definition of the PC method was introduced by Wiener (1938). Here, we give a simplified overview of the main definitions and properties of the PC method and refer the reader to the cited references for more details. There are two main approaches used: intrusive approaches that modify the computer program used in the analysis and non-intrusive approaches that leave the analysis program unchanged and instead use the results of the output of the analysis for a suite of cases. The intrusive approach can be more efficient, but it is limited in terms of the complexity of the problem that can be addressed as the number of multivariate polynomials and their associated coefficients explodes as the complexity increases. The non-intrusive approaches have an advantage that they can be extended to much more complex problems. Here, we use an intrusive approach rather than a non-intrusive approach because the the problem is not overly complex, and as will be shown later, the required modifications to the PSHA program are very simple. The intrusive approach also allows uses of the PSHA programs to run the programs in the normal way without having to understand the PC method.

If we have a arbitrary function of a random variable, $f(\xi)$, the distribution of the function can be estimated using Monte Carlo sampling of the random variable (ξ), evaluating $f(\xi)$ for each realization, and computing the resulting distribution of $f(\xi)$. As an alternative to this type of brute force Monte Carlo sampling, the PC method uses a linear combination of a set of polynomials to approximate the distribution of $f(\xi)$ without having to evaluate $f(\xi)$ directly.

Using PC, an arbitrary function of a standard normal random variable, ξ, can be approximated by:

$$f(\xi) = \sum_{i=0}^{P} C_i \Psi_i[\{\xi\}] \tag{7}$$

in which Ψ is a family of polynomials and C_i are constants.

For our PSHA application, the PC approach uses a set of polynomials to analytically estimate the distribution of the hazard due to a distribution of the median ground motion. Assuming that the epistemic uncertainty in the median ground motion is lognormally distributed, the resulting distribution on the hazard can be estimated. By using the same polynomials for each earthquake scenario, the net effect of the uncertainty of the median ground motions on the total hazard can be computed efficiently, by summing the coefficients for each polynomial from all of the scenarios. The distribution of the hazard due to the uncertain median ground motion is a skewed distribution whose shape varies depending on the value of z, m_μ, σ, and σ_μ. Using the PC method, the skewed distribution on the hazard can be approximated at each z level.

When using the PC method, the appropriate family of polynomials should be used so that the approximation in equation (7) has good convergence as additional terms are included. In the PSHA application, we assume that epistemic uncertainty in μ_s is normally distributed. We use Hermite polynomials for $\Psi_i[\{\xi\}$ because they lead to optimal convergence to any function of a Gaussian random variable (Xiu, 2010). The Hermite polynomials of a Gaussian random variable form an orthogonal set. The Hermite polynomials of ξ up to order 6 are given in Table 1 and the shapes of the distributions of the first four terms for random samples of ξ are shown in Figure 2. Using this set of basis functions with appropriate scale factors (polynomial coefficients), we can accurately approximate the skewed distribution of the hazard.

To approximate a function of a random variable, $f(\xi)$, with the Hermite polynomials using the PC method, the coefficients of the polynomials, C_i, are computed by projecting the function f onto each Hermite polynomial. Specifically, the coefficients of each Hermite polynomial are given by the expectation (mean) of the dot product of the function $f(\xi)$ and the Hermite polynomial, normalized by the variance of the Hermite polynomial. The coefficient, C_i, for the i^{th} polynomial is given by

$$C_i = \frac{1}{\text{Var}[\Psi_i[\{\xi\}]]} \langle f(\xi) \cdot \Psi_i[\{\xi\}] \rangle \tag{8}$$

in which $\{\Psi_i\}_{i=1}^{P}$ represent the Hermite polynomials and $\langle\rangle$ is the expectation operator.

The coefficients C_i in the expansion in equation (7) can be obtained by different approaches (Xiu, 2010). The most efficient approach is to use analytical formulations of the expectation of the product involved in equation (8), but, this is not always possible. As described in the following section, for seismic hazard, while there are analytical solutions to equation (8) for the the Hermite polynomials for $i \geq 1$, there is no analytical solution for the first term, i=0; however, this first term has the same form as $P(Y > z|M_s, R_s)$, so numerical approximations for this term are already available in PSHA programs.

PC Expansion of Hazard for Uncertain Median Ground Motion

To model the epistemic uncertainty in the median ground motion, we define ξ as:

$$\xi = \frac{\mu(M,R) - m_\mu(M,R)}{\sigma_\mu(M,R)} \tag{9}$$

which is a standard normal random variable. With the inclusion of epistemic uncertainty in the central estimate of the median ground motion, m_μ, the marginal hazard given in equation (3) becomes:

$$Haz_{s,k}(z,\xi) = rate_{s,k} \times \left[1 - \Phi\left(\frac{\ln(z) - (m_\mu(M_s,R_s) + \sigma_\mu(M_s,R_s)\xi)}{\sigma(M_s,R_s)}\right)\right] \tag{10}$$

Note that ξ in equation 10 is independent of the scenario index, s, and the source index, k. That is, ξ is the same for all scenarios and all sources (for all s and k) for each realization of the epistemic uncertainty in the median ground motion due to the assumption of full correlation of the ξ. This means that the alternative GMMs are given by shifts of the central GMM as shown in equation

$$GMM_i(M,R) = GMM_0(M,R) + \xi\sigma_\mu(M,R) \tag{11}$$

The shifts are constant in terms of the number of standard deviations (ξ), but the amplitude of the shift for a given scenario depends on the size of the epistemic uncertainty ($\sigma_\mu(M,R)$ for the scenario. Also, note that the assumption of full correlation of the epistemic uncertainty is not related to the correlation of the aleatory variability.

The conditional probability of exceeding the ground-motion is given by:

$$P(Y > z|M_s, R_s, \xi) = \left[1 - \Phi\left(\frac{\ln(z) - (m_\mu(M_s, R_s) + \sigma_\mu(M_s, R_s)\xi)}{\sigma(M_s, R_s)}\right)\right] \tag{12}$$

We expand the conditional probability from equation (12) with uncertain median along the Hermite polynomials as shown below:

$$P(Y > z|M_s, R_s, \xi) \approx \sum_{i=1}^{P} C_{si}(z)\Psi_i[\{\xi\}] \tag{13}$$

Using equation (8), the coefficients, $C_{si}(x)$, are obtained by

$$C_{si}(z) = \frac{1}{\mathrm{Var}[\Psi_i[\{\xi\}]]}\langle P(Y > z|M_s, R_s, \xi) \ \cdot \ \Psi_i[\{\xi\}]\rangle \tag{14}$$

in which

$$\begin{aligned}\langle P(Y > z|M_s, R_s, \xi) \cdot \Psi_i[\{\xi\}]\rangle = &\int_{\xi=-\infty}^{+\infty} \left[1 - \Phi\left(\frac{\ln(z) - (m_\mu(M_s, R_s) + \sigma_\mu(M_s, R_s)\xi)}{\sigma(M_s, R_s)}\right)\right] \\ &\times \Psi_i[\{\xi\}] \times \frac{1}{\sqrt{(2\pi)}} e^{-\xi^2/2} d\xi\end{aligned} \tag{15}$$

To solve equation (15), we note that the Hermite polynomials satisfy the property:

$$\frac{d}{d\xi}\{\Psi_i e^{-\xi^2/2}\} = -\Psi_{i+1} e^{-\xi^2/2} \tag{16}$$

Using this property, the integral in equation (15) can be solved using integration by parts. The projection of the hazard curve on the Hermite polynomials has the following form:

$$\langle P(Y > z|M_s, R_s, \xi) \cdot \Psi_i[\{\xi\}]\rangle = \frac{\sigma_\mu}{2\pi\sigma} \int_{\xi=-\infty}^{+\infty} e^{a(z)\xi^2 + b(z)\xi + c(z)} \ \Psi_{i-1}[\{\xi\}] d\xi \tag{17}$$

in which

$$\begin{aligned} a(z) &= -\frac{\sigma_\mu^2}{2\sigma^2} - \frac{1}{2} \\ b(z) &= (\ln(z) - m_\mu)\frac{\sigma_\mu}{\sigma^2} \\ c(z) &= -(\ln(z) - m_\mu)^2 \frac{1}{2\sigma^2} \end{aligned} \tag{18}$$

Solving the integrals in equation (17), we obtain the coefficient of the Hermite PC expansion of a hazard curve as in equation (14).

The coefficients up to order 6 of the PC expansion of $P(Y > z|M, R, \xi)$ are given by:

$$\begin{aligned}
C_1(z) =& \alpha \times \frac{1}{(-a(z))^{1/2}} \\
C_2(z) =& \frac{\alpha}{2} \times \frac{b(z)}{2(-a(z))^{3/2}} \\
C_3(z) =& \frac{\alpha}{6} \times \frac{-2a(z)(1+2a(z)) + b(z)^2}{4(-a(z))^{5/2}} \\
C_4(z) =& \frac{\alpha}{24} \times \frac{-b(z)(6a(z)(1+2a(z)) - b(z)^2)}{8(-a(z))^{7/2}} \\
C_5(z) =& \frac{\alpha}{120} \times \frac{12a^2(z)(1+2a(z))^2 - 12a(z)(1+2a(z))b(z)^2 + b(z)^4}{16(-a(z))^{9/2}} \\
C_6(z) =& \frac{\alpha}{720} \times \frac{b(z)(60a^2(z)(1+2a(z))^2 - 20a(z)(1+2a(z))b(z)^2 + b(z)^4}{32(-a(z))^{11/2}}
\end{aligned} \tag{19}$$

in which

$$\alpha = \frac{\sigma_\mu}{2\sigma\sqrt{\pi}} \times e^{c(z) - b(z)^2/(4a(z))} \tag{20}$$

For the mean term ($i = 0$), there is no analytical solution for the integral; however, C_0 is simply the mean of the conditional probability of exceeding z.

$$C_0 = \langle P(Y > z|M_s, R_s, \xi)\rangle = \int_{\xi=-\infty}^{+\infty} \left[1 - \Phi\left(\frac{\ln(z) - (m_\mu + \sigma_\mu\xi)}{\sigma}\right)\right] \frac{1}{\sqrt{(2\pi)}} e^{-\xi^2/2} d\xi \tag{21}$$

In GMMs, the aleatory variability and the epistemic uncertainty are not correlated (i.e., ϵ and ξ are independent), so the total standard deviation can be computed from the combination of the aleatory, σ, and epistemic, σ_μ, standard deviation terms:

$$C_0 = \left[1 - \Phi\left(\frac{\ln(z) - m_\mu}{\sqrt{\sigma^2 + \sigma_\mu^2}}\right)\right] \tag{22}$$

As a check of the analytical integrals, the values of the PC coefficients for i=1, 2, 3, and 4 using equation (19) are compared to the values computed using numerical integration. Figure 3 shows the comparison for one scenario, with $m_\mu = 0$, $\sigma_\mu = 0.2$, and $\sigma = 0.5$.

The total hazard, $Haz(z)$, is the sum of the marginal hazards from all of the scenarios for all sources:

$$Haz(z) = \sum_{k}^{Nsource} \sum_{s=1}^{N_{scen_k}} Haz_{s,k}(z) \tag{23}$$

The marginal hazard for each scenario can be replaced with the PC approximation:

$$Haz_{s,k}(z) \approx rate_{s,k} \sum_{i=0}^{P} C_{s,k,i}(z)\Psi_i[\{\xi\}] \tag{24}$$

Combining equations (23) and (24), the total hazard is written in terms of the PC coefficients:

$$Haz(z) \approx \sum_{k}^{Nsource} \sum_{s=1}^{N_{scen_k}} rate_{s,k} \sum_{i=0}^{P} C_{s,k,i}(z)\Psi_i[\{\xi\}] \tag{25}$$

Reordering the summations:

$$Haz(z) \approx \sum_{k}^{Nsource} \sum_{i=0}^{P} \Psi_i[\{\xi\} \sum_{s=1}^{N_{scen_k}} rate_{s,k}\, C_{s,k,i}(z) \tag{26}$$

The hazard in equation (26) is a linear combination of the hazard curves from the different sources projected along the Hermite polynomials. This allows us to efficiently compute the total hazard from all scenarios by simply summing the PC coefficients, one by one, for all scenarios. Defining $D_{k,i}(z)$ as the sum of the weighted PC coefficients for all scenarios for the the k^{th} source:

$$D_{k,i}(z) = \sum_{s=1}^{N_{scen_k}} rate_{s,k} C_{s,k,i}(z) \tag{27}$$

The hazard is then given by

$$Haz(z) \approx \sum_{k}^{Nsource} \sum_{i=0}^{P} \Psi_i[\{\xi\} D_{s,k}(z) \tag{28}$$

Rather than computing the hazard for each GMM directly, we sum up the weighted PC coefficients inside the hazard program for each source and then use post-processing to compute the epistemic uncertainty fractiles of the hazard. In the post-processing, samples of the standard normal variable ξ are generated, the Hermite polynomials are evaluated for the given value of ξ, and the hazard is computed using equation (28).

Implementation in PSHA Programs

For PSHA, we are interested in the mean hazard for each source as well as the mean and the uncertainty fractiles of the total hazard. Using the PC expansion, the mean hazard for the k^{th} source is given by the $D_{k,i}$ term for i=0.

$$Mean\,[Haz_k(z)] = D_{k,i=0}(z) \tag{29}$$

This is the same approach used in current PSHA programs, but with the change that the mean hazard is computed for just the central GMM using the combined epistemic and aleatory standard deviations equation (22) rather than computing the weighted average of the hazard from multiple discrete GMM medians.

The epistemic fractiles for the total hazard need to consider the epistemic uncertainty in the SSC model as well as the epistemic uncertainty in the GMM. One approach currently used to compute the epistemic fractiles is to have the PSHA program write out the hazard curves for each combination of branches from the SSC and GMC logic trees to a file. Then, in a post-processing program, Monte Carlo sampling of the branches of the logic trees for the SSC model and GMC model is used to generate a large suite of realizations of the total hazard. The fractiles of the total hazard can be computed from this large suite of realizations.

We can modify this approach to use PC expansion for the epistemic uncertainty in the median ground motion. To implement the PC method for the epistemic fractiles, the PC coefficents also need to be written to a file during the hazard calculation. Specifically, the values of $D_{k,i}(z)$ are written to a file for $i = 0, p$. In the post-processing using the PC expansion method, we still sample the SSC model logic tree using Monte Carlo sampling, but we replace the sampling of the GMC logic tree for the median ground motion with a random sample of the ξ. Given an ξ value, the PC polynomials can be computed and the total hazard computed using equation (28). From multiple realizations of total hazard, the epistemic fractiles of hazard can be computed at a given ground motion level, z.

In some PSHA applications, the aleatory variability of the log ground motion is modeled using a mixture model (sum of two normal distributions) rather than a single normal distribution (GeoPentech, 2015). In this case, the PC coefficients are computed for each of the two normal distributions and the weighted average of the coefficients is computed, in which the weights are the same weights used in the mixture model.

If the epistemic uncertainty in μ is not normally distributed, then a mixture of multiple normal distributions can be developed to approximate the epistemic distribution of μ. Again, the PC approach can be applied to each normal distribution in the mixture model and the weighted sum of the PC coefficients is then used, with the weights corresponding to the mixture-model weights.

EXAMPLE APPLICATION

To simplify the application of the PC method to PSHA, we assumed that there is full correlation of the epistemic uncertainty of the median ground motion between scenarios (i.e., there is full correlation in the ξ values between scenarios). As an example of full correlation, consider the GMM with epistemic uncertainty shown in Figure 4. If the median ground motion for the scenario with $M = 7$ and $R_{rup} = 10$ km is 1 standard deviations above the mean (shown by the upper dotted line in Figure 4), then the median ground motion for all of the other scenarios (all other magnitude and distance combinations) will also be 1 standard deviations above the average median. In this case, the GMM shown in Figure 4 will be on the upper curves for all distances for both the M7 and M5 scenarios. Similarly, if the median ground motion for a scenario is -1 standard deviation below the mean, then the GMM shown in Figure 4 will be on the lower curves for all distances for both the M7 and M5 scenarios. Assuming full correlation will lead to broader fractiles than assuming partial correlation (i.e. allowing for different magnitude and distance slopes for the alternative GMMs); however, the mean hazard is not affected by the full correlation assumption.

Note that full correlation does not require that the epistemic uncertainty in the median, σ_μ, be constant: it can vary as a function of magnitude and distance as is common when extrapolating GMMs outside the data range. In Figure 4, the standard deviation of median ground motion is smallest for scenarios well constrained by data and is largest for scenarios not well constrained by data (e.g., large magnitudes and short distances). Similarly, the aleatory standard deviation, σ, does not need to be constant. A heteroskedastic model for the aleatory variability can be used with the PC method. In equation (19), the σ and σ_μ values for the individual scenarios are used.

We assumed full correlation to keep this initial application of the PC method to PSHA relatively simple, but the full correlation assumption will be applicable to some studies. In particular, full correlation is applicable to the scaled-backbone approach in which the GMM is shifted by a constant for all scenarios (Atkinson et al., 2014). It is also applicable to the additional epistemic uncertainty model of Al Atik and Youngs (2014), which gives a scenario-dependent epistemic uncertainty which are typically assumed to be fully correlated in PSHA applications.

The full correlation assumption has been used in major hazard studies in the past: for example, in the Yucca Mountain PSHA (Stepp et al., 2001) and the PEGASOS PSHA for Switzerland (NAGRA, 2004), individual ground-motion experts assumed full correlation, but the aggregation of the results from the different experts did not assume full correlation. The decision to simplify the epistemic uncertainty by assuming full correlation for the ξ values between scenarios is up to the judgment of the group developing the ground-motion characterization. We are not making a recommendation about the use of either full correlation or partial correlation of ξ in this paper. Instead, we are providing an efficient method to propagate the epistemic uncertainty due to the median GMM if full correlation is assumed. Partial correlation can be incorporated in the PC method. Including partial correlation leads to a more complex PC expansion with more terms required. The application of the PC method for the partial correlation case is being addressed in a separate study.

To implement the PC method into a traditional PSHA computer program, the following changes need to be made. First, include the epistemic uncertainty in the median ground motion as a function of the earthquake scenario parameters as part of the input. This can be included in the GMM subroutine so that it returns the epistemic uncertainty in the median as well as the median and aleatory variability for a given scenario. Second, compute the mean marginal hazard for each scenario using the combined aleatory and epistemic standard deviations (eq. 22), which gives the $C_{s,0}(z)$ coefficient. Third, compute the additional PC coefficients for each scenario and each z value (eq. 19). Fourth, compute the sum of each PC coefficient weighted by the scenario rate (eq. 27). To be able to compute the mean and fractiles of the hazard by source, save the $D_i(z)$ terms for each source separately.

As a check of the method, we computed the fractile of the conditional probability of exceeding the ground motion from a single scenario using Monte Carlo sampleing of the GMM distribution and using PC. Figure 5 compares the fractiles for a scenario with $m_\mu = 0$, $\sigma_\mu = 0.2$, and $\sigma = 0.5$. This figure shows that the PC expansion provides a good approximation of the 10th, 50th, and 90th fractiles of the $P(Y > z|M, R)$ term for a single scenario. The 10th fractile from the PC method begins to deviate from the 10th fractile for Monte Carlo sampling at z values greater than 6 g, corresponding to $\epsilon > 3.5$ relative to the central median value, m_μ.

Figure 6 shows how the PC expansion works by plotting the cumulative distribution function (CDF) of the $P(Y > z|M, R)$ term and the CDF for different orders of the PC expansion. For order 0, the PC expansion just gives the mean value, so the CDF is a step function at the mean value. For order 1, the PC expansion starts to capture some of the shape of the CDF. For order 2, the PC expansion is getting close to the shape of the CDF. As the order is increased to order 3 and order 4, the PC expansion fits the lower tails of the distribution better.

Figure 7 compares the CDF of the epistemic uncertainty of the $P(Y > z|M, R)$ term for two values of z for the same scenario with $m_\mu = 0$, $\sigma_\mu = 0.2$, and $\sigma = 0.5$. Using the same set of random variables, ξ, the fourth-order PC approximation can reproduce the different distributions at the two z values for $\sigma_\mu = 0.2$. For comparison, the step function shows how the fractiles of the median GMM are typically modeled in current PSHA for a logic tree with five branches to sample the normal distribution of the median.

Figure 8 shows that the fourth-order PC approximation looses accuracy at the lower third of the epistemic range for large values of σ_μ. For $\sigma_\mu = 0.4$, the fourth-order PC approximation is in good agreement with Monte Carlo sampling for the upper half of the CDF, but it begins to deviate for $P(Y > z|M, R) < 0.03$. The accuracy of the PC approximation can be improved by including additional terms. Using a sixth-order PC expansion, the PC approximation for the lower part of the CDF is improved with good agreement for $P(Y > z|M, R) > 0.008$.

Next, we show an example of a hazard calculation using a simple source model that consists of one fault and a background zone with uniformly distributed rates. The seismic source parameters used in this example are listed in Table 2. To focus on the ability of the PC approach to capture the effects of the epistemic uncertainty in the median ground motion, no epistemic uncertainty in the source model or in the aleatory variability of the GMM is included in this example. As a check, we use 10,000 samples of the median GMM distribution, generated using Monte Carlo sampling. The hazard is computed for each of the 10,000 realizations and the epistemic fractiles are computed (Fig. 9). The mean and 10^{th}, 50^{th}, and 90^{th} fractiles computed using the PC method are shown by the symbols in Figure 9. There is good agreement between the brute force Monte Carlo approach with 10,000 realizations and the fourth-order PC approximation.

The full cumulative distributions of the epistemic fractiles for z=3 g for the logic tree approach and the fourth-order PC approach are compared in Figure 10. This z value corresponds to $\epsilon = 2.2$ for the selected m_μ. This figure shows that the PC approach provides a good agreement with the Monte Carlo sampling of the epistemic uncertainty over the full range of the epistemic uncertainty for $\sigma_\mu = 0.2$

The savings in computational time is shown in Figure 11. The calculation time for the logic tree approach is shown using two methods for computing the median ground motion for each scenario: (1) evaluating the GMM for for each branch and (2) computing m_μ once and then computing $m_\mu + \xi\sigma_\mu$ for each branch. There is some computational savings by computing m_μ just once per scenario but it is much less than the savings using the PC method. Using the PC method, the run time to compute and save the fourth-order PC coefficients is only 15 percent longer than the run time to compute the hazard for a single GMM. Using sixth-order expansion, the run time increases to 25 percent over the run time to compute the hazard for a single GMM. The computational savings may differ depending on how the calculations in the PSHA program are structured.

Using PC can also lead to a significant reduction in terms of memory requirements. For the PSHA programs that store the marginal hazard from each branch of the GMM logic tree for computational speed, the memory requirements are approximately proportional to the number of epistemic branches for the alternative GMMs. Using PC, the memory requirements scale with the number of PC terms rather than the number of GMM branches.

Including Correlation

A limitation of the fully correlated case is that it does not capture changes in the distance or magnitude scaling between GMMs. An example of the distance scaling of the T=0.75 sec spectral acceleration for the five NGA-West2 GMMs is shown in Figure 12 for a magnitude 7 strike-slip earthquake. Some of the individual GMMs have different slopes with distance, indicating that there is partial rather than full correlation of the epistemic uncertainty in the median ground motion between the different M,R scenarios. If there was full correlation, the alternative GMMs would just be shifted up or down without crossing.

In this section, we extend the PC method described earlier to include partial correlation of the epistemic uncertainty between different GMMs, allowing the alternative GMMs to have different magnitude and distance scaling. To include the correlation structure into hazard calculations, the correlation structure of median GMMs is discretized using the Karhunen-Loeve (K-L) expansion, and the K-L expansion is substituted back into the original expansion given in equation (24). The result is a multivariate PC expansion of hazard curves that can represent both their marginal distribution at one given scenario and the correlation between different scenarios. This improvement in the PC expansion comes at a computational cost, but, as shown later, the additional computational cost is still small compared to direct sampling of a large number of logic tree branches for the median GMMs.

Correlation Structure of Median GMMs Across Scenarios

To capture the correlation structure between the alternative GMMs, we compute the semi-variogram for the median ground motions for a suite of earthquake scenarios. Here, we only consider the semi-variogram for the magnitude and distance scaling. In this case, the semi-variogram is a measure of the average dissimilarity between the alternative GMMs for a given difference in the magnitudes (ΔM) and a given difference in the log distance (ΔlnR). The semi-variogram is one half of the variance between pairs of alternative GMMs:

$$V(\Delta M; \Delta \ln R) = \frac{1}{2}[E\{GMM_{M,R} - GMM_{M+\Delta M, \ln R + \Delta \ln R}\}^2] \tag{30}$$

The correlation is given by:

$$C(\Delta M; \Delta \ln R) = \frac{1 - V(\Delta M; \Delta \ln R)}{V(\Delta M; \Delta \ln R)} \tag{31}$$

A parametric model for the correlation can be fit to the estimated correlation values.

Consider the epistemic uncertainty in the median ground motion as a Gaussian random process, $\{\gamma(M,R)\}$, that has a correlation function $C(\Delta M, \Delta lnR)$ between two scenarios (M,R) and (M',R'). The Karhunen—Loève (KL) expansion is an efficient method for discretizing a second-order stochastic processes with a known correlation function. The KL expansion comes from Mercer's theorem (Mercer (1909)) which states that there exists a set of eigenvalues λ_k and orthonormal eigenfunctions f_k for the correlation function C, such that:

$$C(\Delta M, \Delta lnR) = \sum_{k=1}^{+\infty} \lambda_k f_k(M,R) f_k(M',R') \tag{32}$$

The eigenvalues λ_k and eigenfunctions f_k satisfy the Fredholm integral equation defined over the 2-D domain $\mathcal{D} = [M_{min}, M_{max}] \times [R_{min}, R_{max}]$:

$$\iint_{\mathcal{D}} C(\Delta M, \Delta lnR)\ f(M',R') dM' dR' = \lambda\ f(M,R) \tag{33}$$

This integral equation is a homogeneous Fredholm equation of the second type. The symmetry and positive definiteness of the correlation kernel ensure that the eigenvalues are positive and that the eigenfunctions have real values.

Using the K-L expansion, the random process γ can be expressed in terms of these eigenvalues and eigenfunctions:

$$\gamma(M,R) = \sum_{k=1}^{\infty} \sqrt{\lambda_k}\ f_k(M,R)\ \xi_k \tag{34}$$

where $\{\xi_1, \xi_2, ..., \xi_k\}$ form a set of uncorrelated random variables.

To solve the integral eigenvalue problem from equation (33), we use the finite-element method, as described in Ghanem (1991). The methodology remains the same as for a regular engineering problem: (1) discretize the 2-D domain of magnitudes $\mathcal{D} = [M_{min}, M_{max}] \times [R_{min}, R_{max}]$ using a finite-element mesh, (2) approximate the unknown field (here, the eigenfunctions f_k) using shape functions (linear, quadratic, ...), and (3) proceed to a Galerkin projection over equation (33). This leads to a system of equations to solve for the values of the eigenfunctions at each degree of freedom.

The expansion used in equation 34 is truncated to the order N_{KL}. The truncated approximation of γ from the KL expansion is denoted γ_{KL}, and is given by

$$\gamma_{KL} = \sum_{k=1}^{N_{KL}} \sqrt{\lambda_k}\ f_k(M, R)\ \xi_k \tag{35}$$

The value of N_{KL} controls the accuracy of the representation of the correlation structure by γ_{KL}. For correlation functions with large correlation lengths, only a few terms are required in the K-L expansion to accurately represent the target correlation structure. For small correlation lengths, more terms in the K-L expansion are needed (Sakamoto and Ghanem (2002)).

Because the random variables, ξ_k, of the truncated KL expansion are uncorrelated and the eigenfunctions are orthonormal, the variance of γ_{KL} is simply the sum of the eigenvalues:

$$Var[\gamma_{KL}] = \sum_{k=1}^{N_{KL}} \lambda_k \tag{36}$$

Example Correlation

As an example, we use the five NGA-West2 GMMs (Gregor et al., 2014) and include the additional epistemic uncertainty given by Al Atik and Youngs (2014) as is common practice. Using the five NGA-West2 GMMs each with three alternative values of the additional epistemic uncertainty (total of 15 alternative GMMs), the median PGA is computed for a suite of magnitudes and distances for strike-slip earthquakes. The correlation of the differences between the 15 GMMs across magnitude and distance is shown in Figure 13 as a function of the differences in magnitude (ΔM) and the difference in the $ln(R_{Rup})$ (ΔlnR).

Based on the shape of the point estimates of the correlation, We used the following parametric form for the correlation function

$$C(\Delta M, \Delta \ln R) = c_1 * \Delta M + c_2 * \Delta \ln R + \exp[c_3 * \{\Delta M\}^2 + c_4 * \Delta \ln R + c_5 * \Delta M * \Delta \ln R] \tag{37}$$

We obtain the fitting coefficients $\{c_1, c_2, c_3, c_4, c_5\}$ using the non-linear least-squared method. The estimated values of the coefficients are listed in Table (3).

The number of terms in the truncated K-L expansion is selected so that we capture 99% of the variance:

$$\frac{\sum_{k=1}^{N_{KL}} \lambda_k}{\sum_{k=1}^{N_{DOF}} \lambda_k} \geq 0.99 \tag{38}$$

In this example, using $N_{KL} = 6$ is large enough to satisfy the last criteria. The first six eigenvalues and their cumulative sum are plotted in figures 14 and 15, to show that higher terms contribute less and less to the expansion. The first six eigenfunctions are shown in figure (16).

Discretization of Correlated Hazard Curves: Multivariate Polynomial Chaos Expansion

As explained earlier, the epistemic uncertainty in a hazard curve for a given (M, R) scenario at a given z-value, can be approximated using a Hermite Polynomial Chaos expansion, such that:

$$Haz(z, \gamma) = \sum_{i=0}^{P} C_i(z)\Psi_i[\{\gamma\}] \tag{39}$$

where $C_i(z)$ are the deterministic Polynomial Chaos coefficients of the expansion, γ denotes a standard normal random variable representing the epistemic uncertainty in median GMMs across all scenarios, and Ψ_i represents the Hermite polynomials in γ.

We now consider the epistemic uncertainty in GMMs to be partially correlated across different scenarios. The standard normal variable γ is replaced by the Gaussian process, $\gamma(M, R_{RUP})$, that has the correlation function given by equation (37). Theγ term is approximated by its Kahrunen-Loeve expansion γ_k:

$$\gamma_{KL} = \sum_{i=1}^{N_{KL}} \sqrt{\lambda_k}\ f_k(M, R_{RUP})\ \xi_k \tag{40}$$

where λ_k and f_k are the solution of equation (33), and $\{\xi_1, \xi_2, ..., \xi_{N_{KL}}\}$ represent a set of independent Gaussian random variables. Substituting this expression of γ_{KL} inside the Polynomial Chaos expansion of hazard from equation (39) leads to its multidimensional Hermite Polynomial Chaos expansion, as explained in (Sakamoto and Ghanem (2002)). For example, with a Polynomial Chaos order of $P = 3$:

$$
\begin{aligned}
Haz(z,\gamma) &= C_0(z) + C_1(z)\gamma_{KL} + C_2(z)[\gamma_{KL}^2 - 1] + C_3(z)[\gamma_{KL}^3 - 3\gamma_{KL}] + C_4(z)[\gamma_{KL}^4 - 6\gamma_{KL}^2 + 3] \\
&= C_0(z) + \\
&\quad C_1(z)\sum_{k=1}^{N_{KL}} \sqrt{\lambda_k}\ f_k(M, R_{RUP})\ \xi_k\ + \\
&\quad C_2(z)[\{\sum_{k=1}^{N_{KL}} \sqrt{\lambda_k}\ f_k(M, R_{RUP})\ \xi_k\}^2 - 1]\ + \\
&\quad C_3(z)[\{\sum_{k=1}^{N_{KL}} \sqrt{\lambda_k}\ f_k(M, R_{RUP})\ \xi_k\}^3 - 3\sum_{k=1}^{N_{KL}} \sqrt{\lambda_k}\ f_k(M, R_{RUP})] + \\
&\quad C_4(z)[\{\sum_{k=1}^{N_{KL}} \sqrt{\lambda_k}\ f_k(M, R_{RUP})\ \xi_k\}^4 - 6\sum_{k=1}^{N_{KL}} \sqrt{\lambda_k}\ f_k(M, R_{RUP}) + 3]
\end{aligned}
$$

By expanding and rearranging the terms in the previous equation, the uncertain hazard, $Haz(z,\gamma)$, can be rewritten as a linear combination of the multivariate Hermite Polynomial Chaos:

$$Haz(z) = \sum_{i=0}^{N_{PC}} c_i(z)\Psi_i[\{\xi_k\}_{k=1}^{N_{KL}}] \tag{41}$$

This multivariate Polynomial Chaos expansion of hazard, contrary to the univariate Polynomial Chaos expansion shown in the fully correlated case, can represent the partial correlation of GMMs across scenarios, while still being able to represent the marginal distribution of hazard for a given scenario, at different ground motion values. This correlation will not affect the mean and standard deviation of hazard; however, it will affect the tails of the probability density and cumulative distribution functions. The trade-off for representing the correlation of GMMs between scenarios is the number of terms involved in the PC expansion. For a fixed order of the PC expansion, the total number of basis functions in the Hermite PC increases tremendously with the number of random variables involved, which come from the K-L expansion.

The total number of terms in the expansion (41) is given by the binomial coefficient:

$$N_{PC} = \binom{P + N_{KL}}{P} = \frac{P!}{N_{KL}!(P - N_{KL})!} \tag{42}$$

where N_{PC} is the total number of terms involved in the expansion, P is the order of the Polynomial Chaos/maximum degree of the polynomials in the expansion, and N_{KL} is the number of terms used in the Kahrunen—Loeve expansion.

In this example, we used a degree of $P = 4$ for the case with only one Gaussian random variable (full correlation in GMMs across scenarios). When including the first (mean) term, the number of terms in the total expansion for the full correlation case corresponds to $N_{PC} = 5$; however, with $N_{KL} = 6$ terms from the Kahrunen—Loeve expansion, there is a total of $N_{PC} = 210$ terms in the multivariate Polynomial Chaos expansion. This is a large increase in the computational burden, but it is still more efficient than discrete sampling of a large number of logic tree branches for the median GMM. Using the HAZ45 PSHA program, the run time using PC with partial correlation is about a factor of 6 longer than for a single GMM using the traditional approach. With a large number of PC terms for the partial correlation case, there is no savings in terms of the memory requirements.

Example Application

We designed a specific hazard calculation to try to show maximum effect of partial correlation between GMMs on the resulting total hazard using the NGA-West2 GMMs. The source model includes two sources: the first source generates large magnitude earthquakes (M7-M8) earthquakes at large distances (100 km), and the second source generates moderate magnitude earthquakes (M5.5-6.5) at short distances (0-10 km) as shown by the disaggregation in Figure 18. The dominant scenarios from each source present a difference of 1.5 magnitude units and a difference in 2.3 log distance units, which leads to a correlation between the medians for the GMMS of near 0 when reading from the semi-variogram (13). That is, for these two scenarios, the correlation is near zero compared to the assumption of full correlation used in the earlier sections.

In order to maximize the effect of the reduced correlation between GMMs across all the scenarios, we adjusted the activity rate of each hazard curve such that each hazard curve has an equal contribution on the total hazard (Figure 17). Attributing equal contribution of each hazard curve in this manner will prevent the hazard from being dominated by just one of the two scenarios.

To start, Figure 20 shows the effect of the partial correlation on the epistemic fractiles using Monte Carlo sampling. For this set of GMMs, the difference between the hazard for partial correlation and full correlation is small.

Figure 21 compares the epistemic fractiles for the partially correlated case using Monte Carlo sampling and using the PC approximation with 210 terms. The PC approximation provides a good estimate of the epistemic uncertainty with partial correlation over most of the range. For the lowest fractile (10^{th} fractile), the PC expansion leads to slightly wider fractiles at high ground-motion values (i.e., high ϵ values). This small error can be reduced by using additional orders in the PC expansion, for example PC order 5 or 6, as shown earlier. This would result in even more accurate hazard fractiles but at the cost of more computational expenses. Again, the mean hazard is the same from Monte Carlo sampling and PC expansions and is obtained analytically from the derived equations. Including partial correlation in the GMMs does not change the mean hazard compared to the case of full correlation.

In this example, the uncertainty in the total hazard including partial correlation between the median GMMs is similar to the uncertainty from the fully correlated case. That is because the GMMs from California are very similar to each other. For other regions where the GMMs differ more, the effect of the correlation is expected to be more significant, resulting in larger differences in the tails of the distribution of the total hazard compared with the fully correlated case; however, the mean hazard will be the same for the fully correlated case and the partial correlated case for any region.

CONCLUSIONS

Using the PC method, more accurate fractiles can be estimated in less time as compared to the traditional approach using logic trees with discrete sampling of the distribution for the median GMM. In addition to the significant computational speed improvements, the PC method also has reduced memory memory requirements compared to a traditional logic tree approach with a large number of branches for the median GMM: the ratio of the memory requirements is given by the ratio of the number of GMM branches to the order of the PC expansion. For example, using a PC expansion of order 6 will require 1/5 of the memory required for a logic tree with 30 branches for the median GMM.

As PSHA moves to using fully non-ergodic GMMs, an efficient method to propagate the epistemic uncertainty in median ground motion is needed to capture the range of source and path effects for each source and site location. Rather than requiring faster and faster computers, we can dramatically speed up the calculations using the PC approximation. The PC method is well suited to work with the Gaussian process (GP) approach for developing non-ergodic GMMs (e.g., Landwehr et al., 2016). The PC method requires an estimate of the median ground motion and the standard error of the median for each scenario. The GP approach directly estimates of the median ground motion (m_μ) and the epistemic uncertainty in the median (σ_μ) for a given earthquake scenario and a given source location and site location, without requiring estimates of the spatially-varying coefficients (see eq. 10 and 11 in Landwehr et al., 2016). This makes the PC method an ideal approach for implementing non-ergodic GMPEs into PSHA.

For the partially correlated case, Monte-Carlo simulations show that the distribution of the total hazard will differ at the tails compared to the fully correlated case, but similar mean hazard will be obtained in both cases. The distribution of the total hazard with partial correlation can be accurately approximated with multivariate PC expansions. Multivariate PC presents significantly more terms than for the case of full correlation but is still more efficient than Monte-Carlo simulations by several orders of magnitude. For the region of California, the median GMMs are so similar to each other that the effect of partial correlation on the hazard distribution is negligible compared to the fully correlated case. For computational efficiency, we therefore recommend the assumption of full correlation between the GMMs in California and the use of univariate PC expansions to efficiently propagate the epistemic uncertainty in the median ground-motion models. For other regions and other sets of GMMs, the effect of partial correlation between the GMMs still needs to be investigated and is the scope of future studies.

Acknowledgments

Ian Main and Paolo Bazzurro provided useful comments on the manuscript as part of the review by the SIGMA2 scientific committee. This work was supported by the PG&E Geosciences Department Long-Term Seismic Program and is part of PG&E's contribution to the SIGMA2 project.

References

Al Atik, L. and R. R. Youngs (2014). Epistemic uncertainty for NGA-West2 models. *Earthq. Spectra* **30**(3), 1301–1318.

Anderson, J. G. and J. N. Brune (1999). Probabilistic seismic hazard analysis without the ergodic assumption. *Seismol. Res. Lett.* **70**(1), 19–28.

Atkinson, G., J. Bommer, and N. Abrahamson (2014). Alternative approaches to modeling epistemic uncertainty in ground motions in probabilistic seismic hazard analysis. *Seismol. Res. Lett.* **85**, 1141–1144.

Bommer, J. J., J. Douglas, F. Scherbaum, F. Cotton, H. Bungum, and D. Fah (2010). On the selection of ground-motion prediction equations for seismic hazard analysis. *Seismol. Res. Lett.* **81**(5), 783–793.

GeoPentech (2015). Southwestern United States Ground-Motion Characterization SSHAC Level 3 - Technical Report Rev. 2. March 2015.

Ghanem, R. and P. Spanos (1991). Stochastic Finite Elements: A Spectral Approach, Springer-Verlag.. 1991.

Gregor, N., N. A. Abrahamson, G. M. Atkinson, D. M. Boore, Y. Bozorgnia, K. W. Campbell, B. S.-J. Chiou, I. M. Idriss, R. Kamai, E. Seyhan, W. Silva, J. P. Stewart, and R. Youngs (2014). Comparison of NGA-West2 GMPEs. *Earthquake Spectra* **30** (3), 1179–1197.

Landwehr, N., N. M. Kuehn, T. Scheffer, and N. Abrahamson (2016). A nonergodic ground-motion model for California with spatially varying coefficients. *Bull. Seismol. Soc. Am.* **106**(6), 2574–2583.

NAGRA (2004). Probabilistic Seismic Hazard Analysis for Swiss Nuclear Power-Plant Sites (PEGASOS Project). Technical report.

Stepp, C., I. Wong, J. Whitney, R. Quitmeyer, N. Abrahamson, G. Toro, R. Youngs, K. Coppersmith, J. Savy, and T. Sullivan (2001). Probabilistic seismic hazard analyses for ground motions and fault displacement at Yucca Mountain, Nevada. *Earthq. Spectra* **17**(1), 113–151.

Mercer (1909). functions of positive and negative type, and their connection the theory of integral equations.. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **209**, 441–458.

Sakamoto, S. and R. Ghanem (2002). Polynomial chaos decomposition for the simulation of non-gaussian nonstationary stochastic processes. *Journal of Engineering Mechanics* **128**(2), 190–201.

Villani, M. and N. A. Abrahamson (2015). Repeatable site and path effects on the ground-motion sigma based on empirical data from southern California and simulated waveforms from the CyberShake Platform. *Bull. Seismol. Soc. Am.* **105**(5), 2681–2695.

Wiener, N. (1938). The homogeneous chaos. *Am. J. M* **60**(4), 897–936.

Xiu, D. (2010). *Numerical Methods for Stochastic Computations: A Spectral Method Approach.* Princeton, NJ, USA: Princeton University Press.

Table 1: Hermite Polynomials up to Order 6

Degree	**Hermite Polynomial**
$i = 0$	1
$i = 1$	ξ
$i = 2$	$\xi^2 - 1$
$i = 3$	$\xi^3 - 3\xi$
$i = 4$	$\xi^4 - 6\xi^2 + 3$
$i = 5$	$\xi^5 - 10\xi^3 + 15\xi$
$i = 6$	$\xi^6 - 15\xi^4 + 45\xi^2 - 15$

Table 2: Source Parameters for Example Calculation

	Fault	**Background**
Length (km)	50 km	
Crustal Thickness	15 km	15 km
Dip	90	90
Slip Rate	0.5 mm/yr	
Activity Rate ($M > 5$)		0.0007 eqk/yr
Magnitude pdf	Youngs and Coppersmith	Truncated Exponential
b-value	0.9	0.9

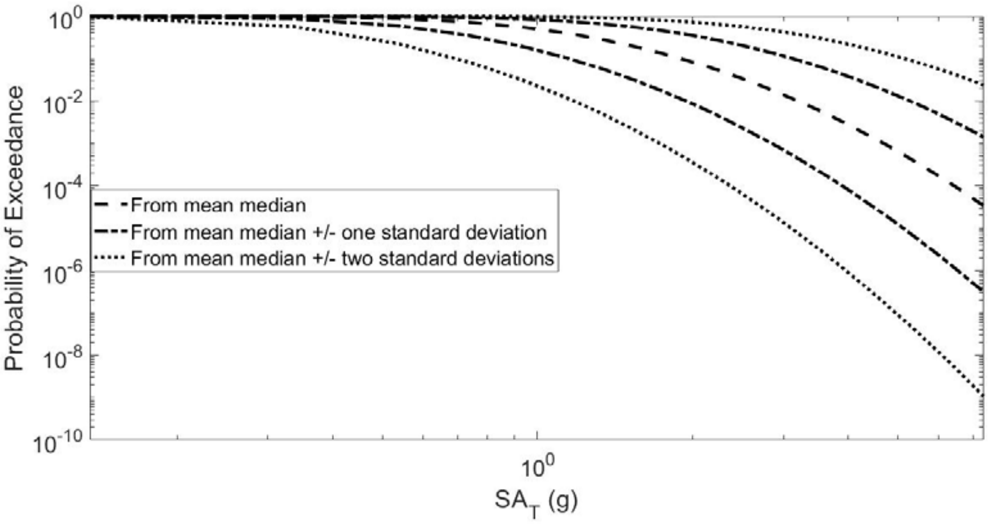

Figure 1: A hazard curve with epistemic uncertainty in the median ($m_\mu = 0$, aleatory $\sigma = 0.5$, and epistemic $\sigma = 0.5$ natural log units)

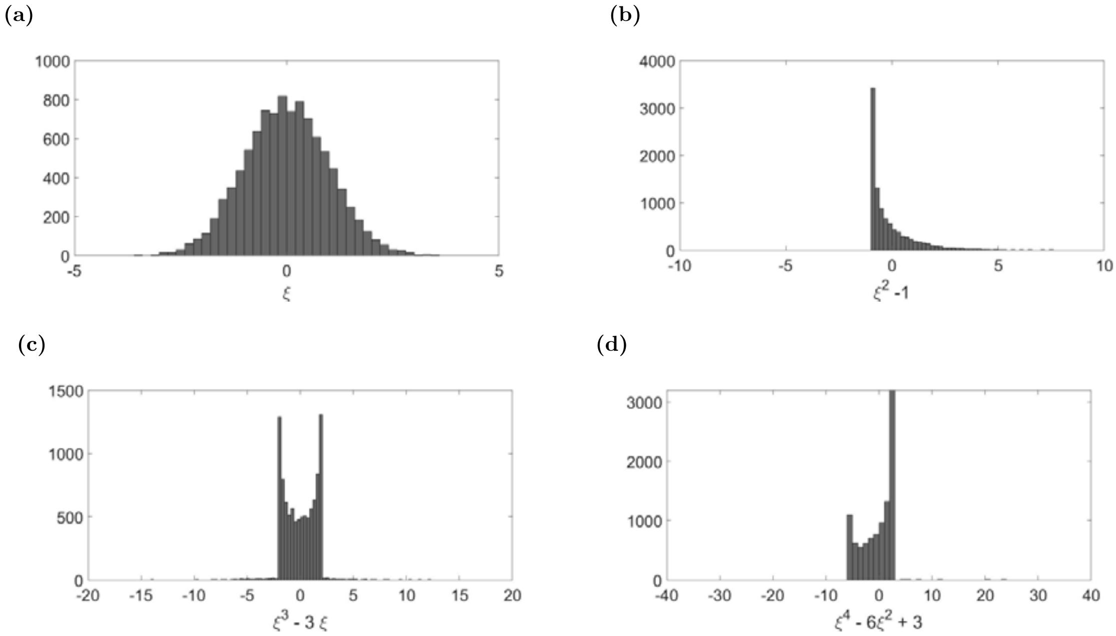

Figure 2: Realizations of first (a), second (b), third (c), and fourth (d) Hermite polynomials (N = 10 000)

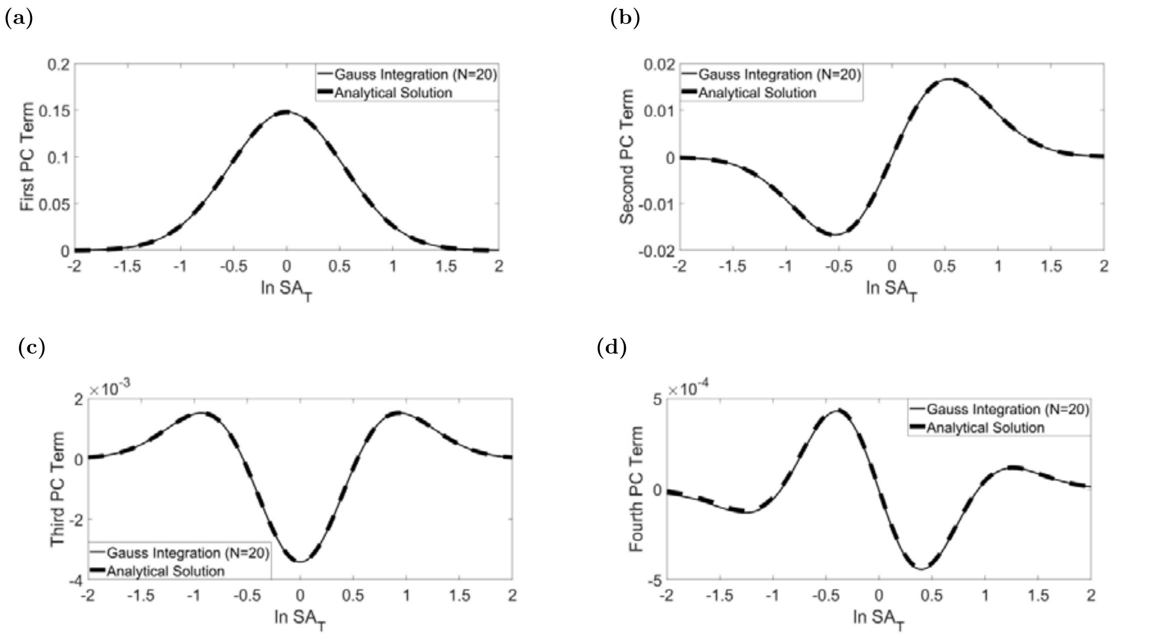

Figure 3: Comparison of the first (a), second (b), third (c), and fourth (d) polynomial chaos terms of uncertain hazard curve computed using the analytical solutions and by numerical integration. The scenario in this example uses $rate = 1, m_\mu = 0, \sigma_\mu = 0.2, \sigma = 0.5$.

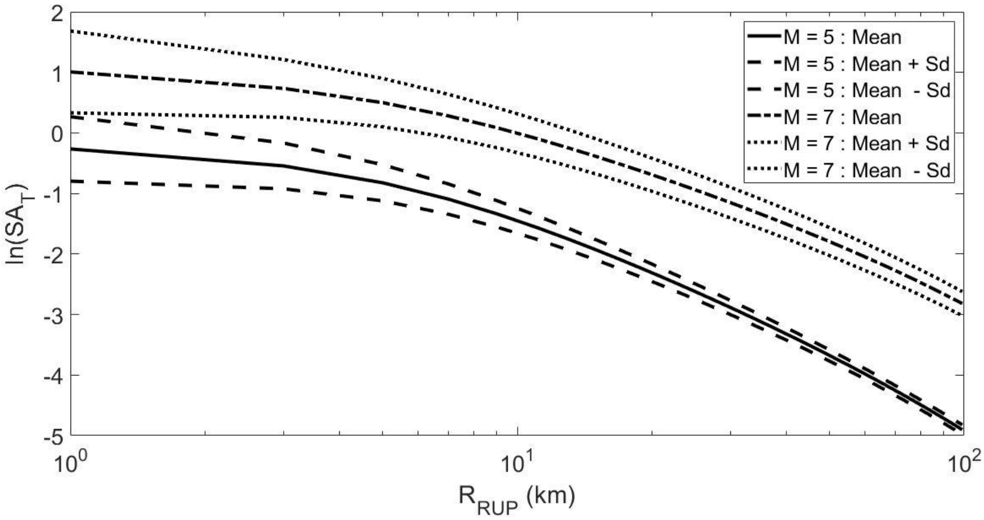

Figure 4: Example of the median GMM and the epistemic uncertainty in the median for magnitude 5 and magnitude 7 scenarios.

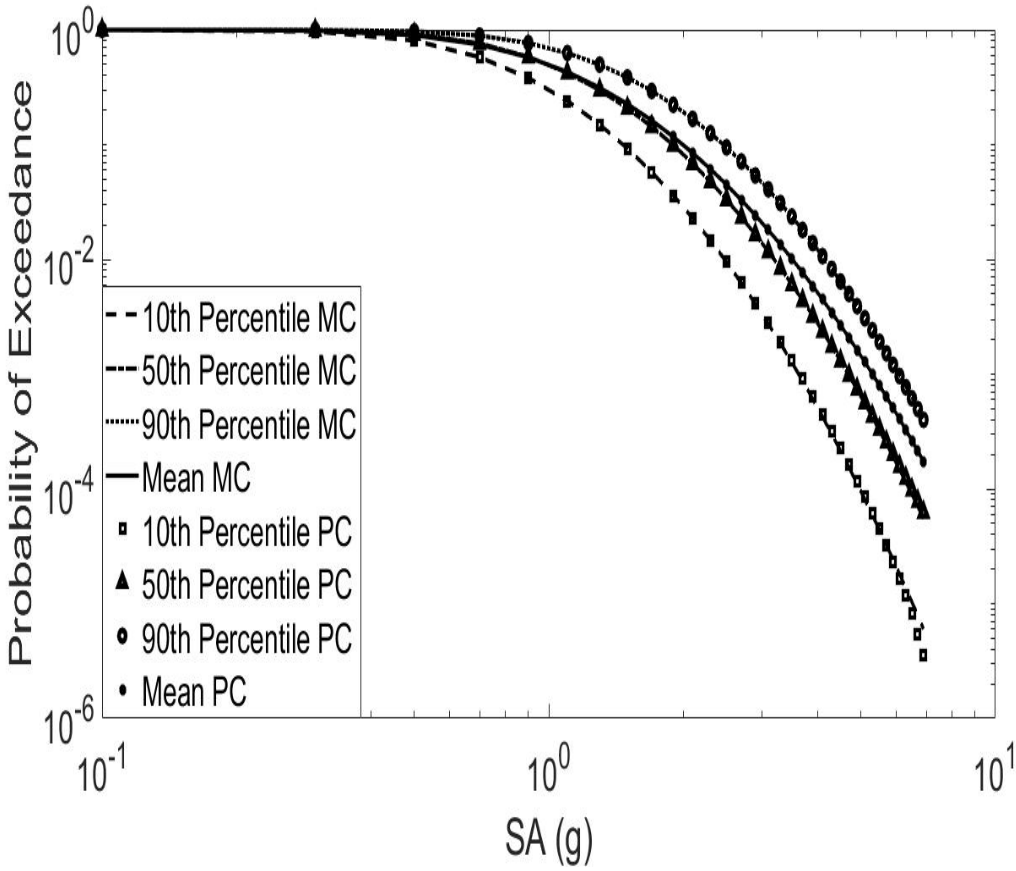

Figure 5: Example of the hazard curve from one scenario with uncertain median ($m_\mu = 0.0, \sigma_\mu = 0.2, \sigma = 0.5$. The epistemic fractiles are compared using Monte-Carlo (lines) and using the fourth-order PC approximation (symbols)

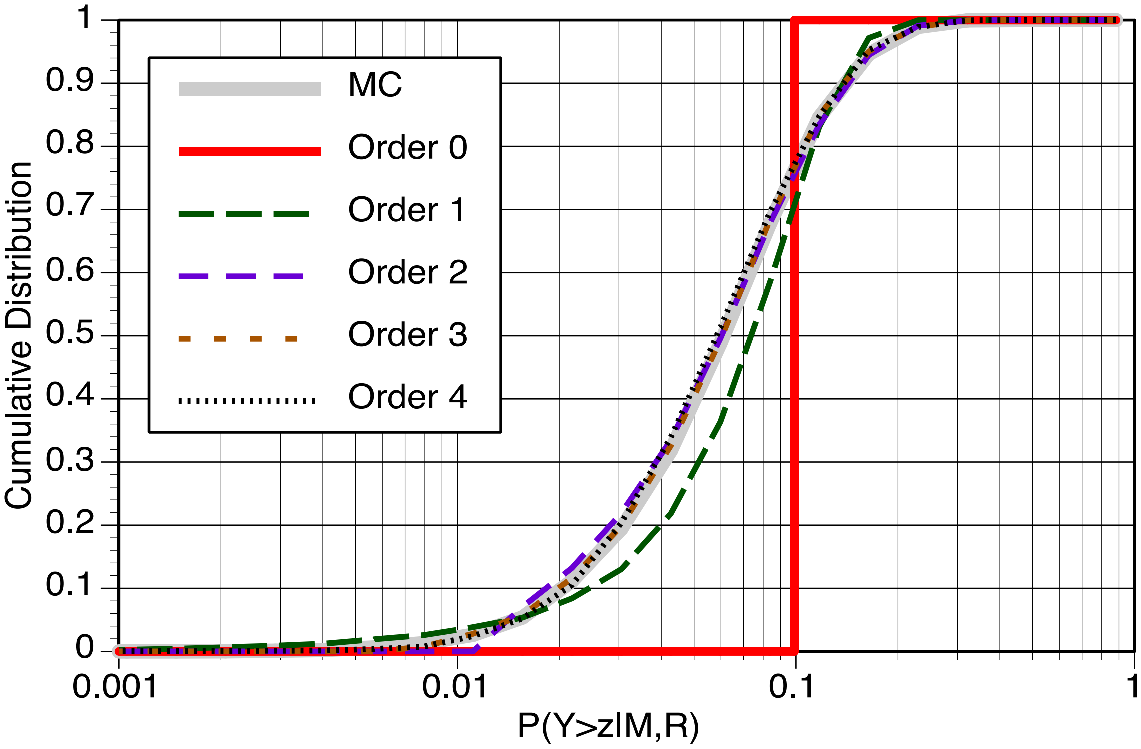

Figure 6: Comparison of the cumulative distribution of the epistemic fractiles of $P(Y > z|M, R)$ using PC of orders 0, 1, 2, 3, and 4. The correct distribution based on Monte Carlo (MC) for $m_\mu = 0$, $\sigma_\mu = 0.2$ and $\sigma = 0.5$ is shown by the thick line.

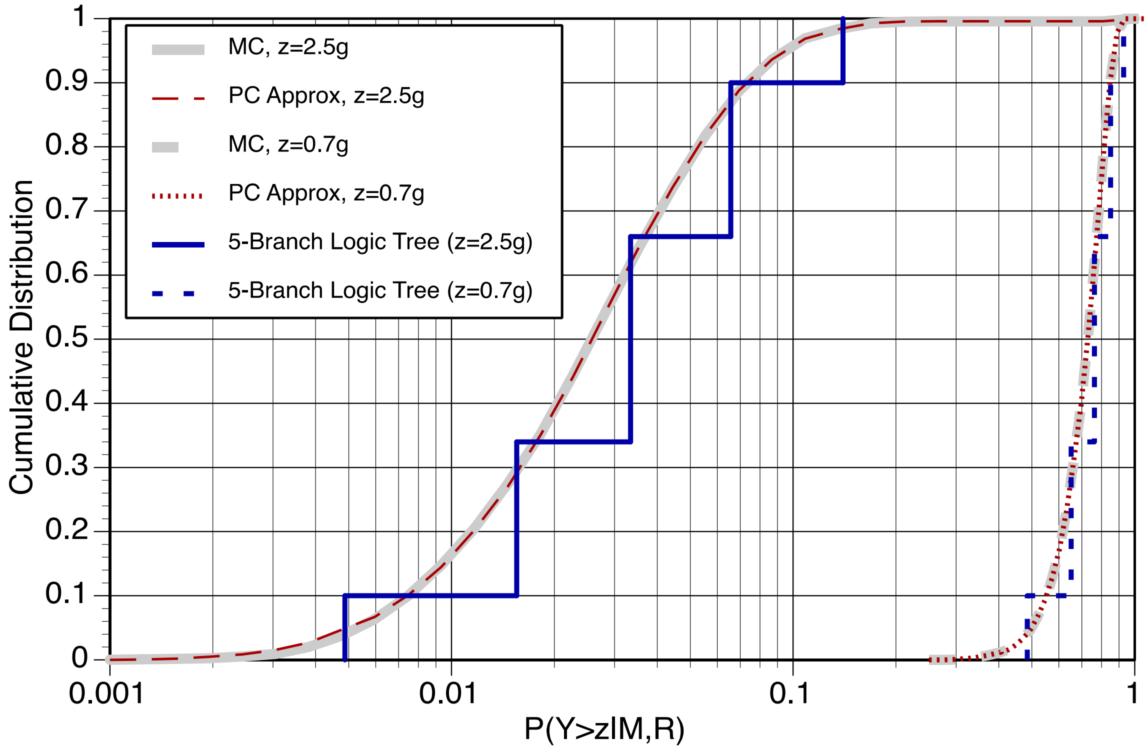

Figure 7: Comparison of the cumulative distribution of the epistemic fractiles of $P(Y > z|M, R)$ using fourth-order PC and using Monte Carlo (MC) for a single scenario for $m_\mu = 0$, $\sigma_\mu = 0.2$ and $\sigma = 0.5$.

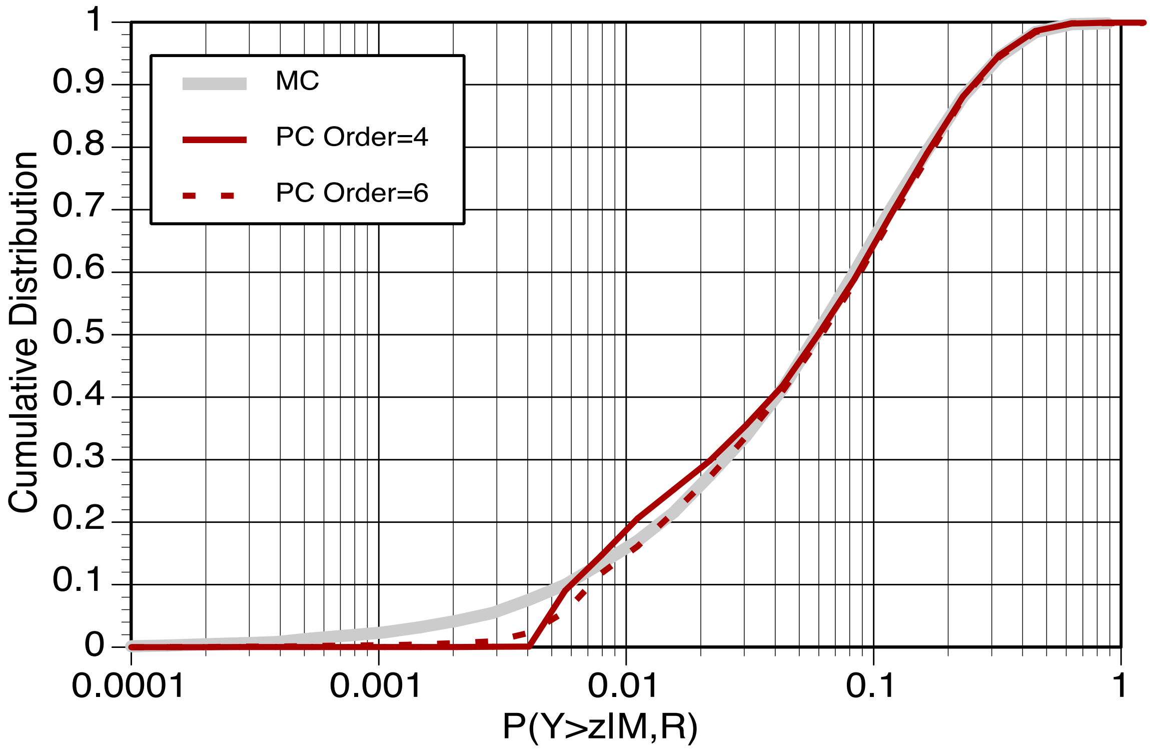

Figure 8: Comparison of the cumulative distribution of the epistemic fractiles of $P(Y > z|M, R)$ with larger epistemic uncertainty using the PC approximation with orders 4 and 6 for a single scenario for $z = 2g$, $m_\mu = 0$, $\sigma_\mu = 0.4$ and $\sigma = 0.5$.

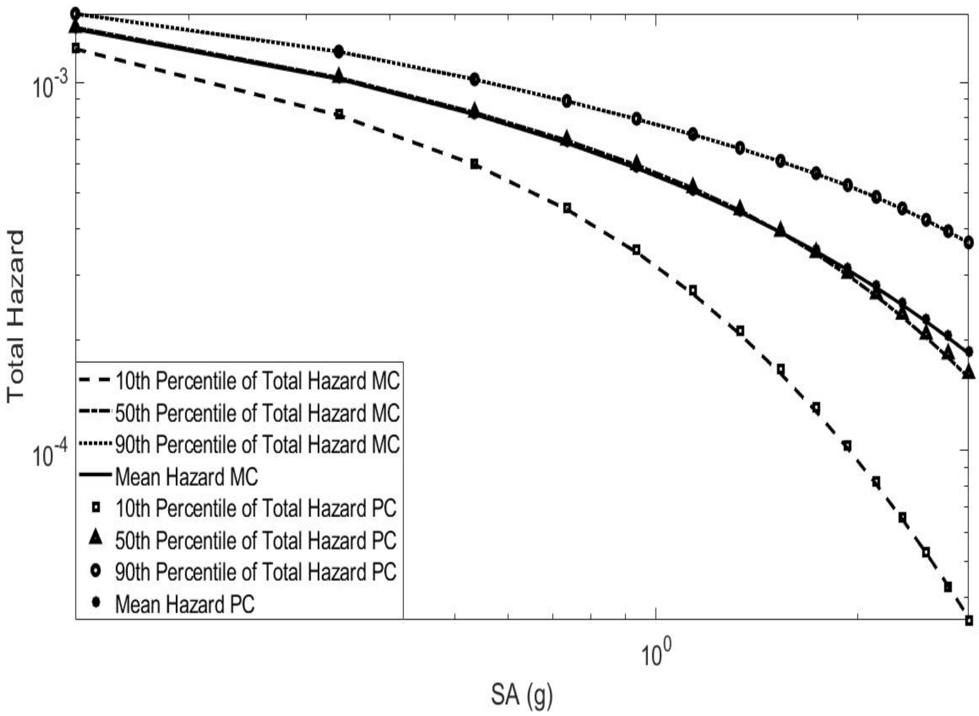

Figure 9: Percentiles of total hazard: Monte-Carlo vs polynomial chaos expansion

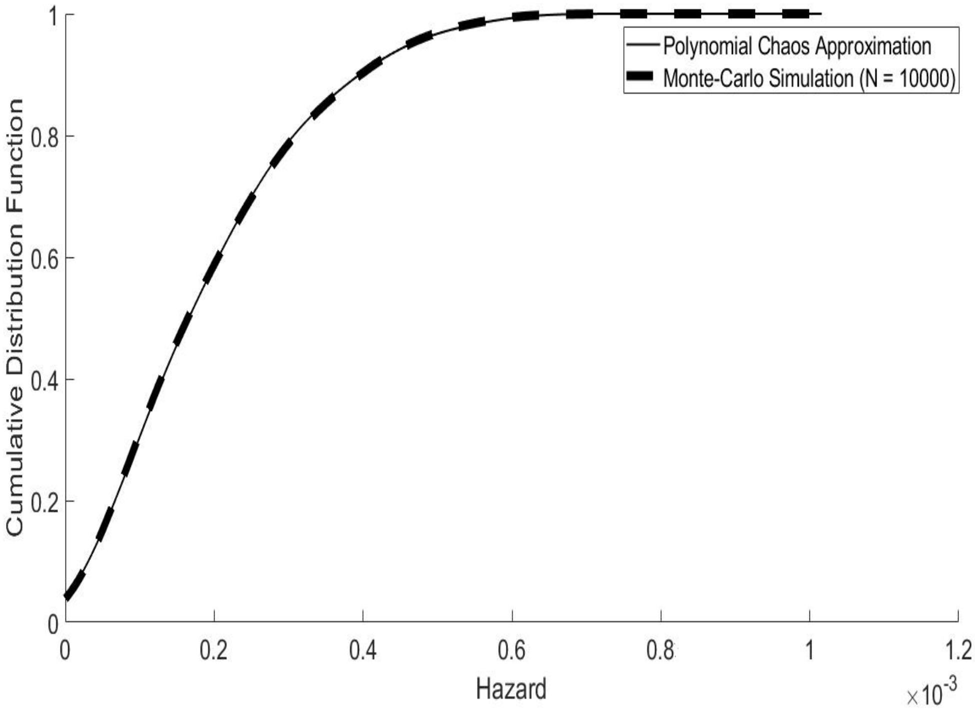

Figure 10: Comparison of the CDF of Total Hazard at z = 3g from the fourth-order PC expansion, and Monte-Carlo simulation

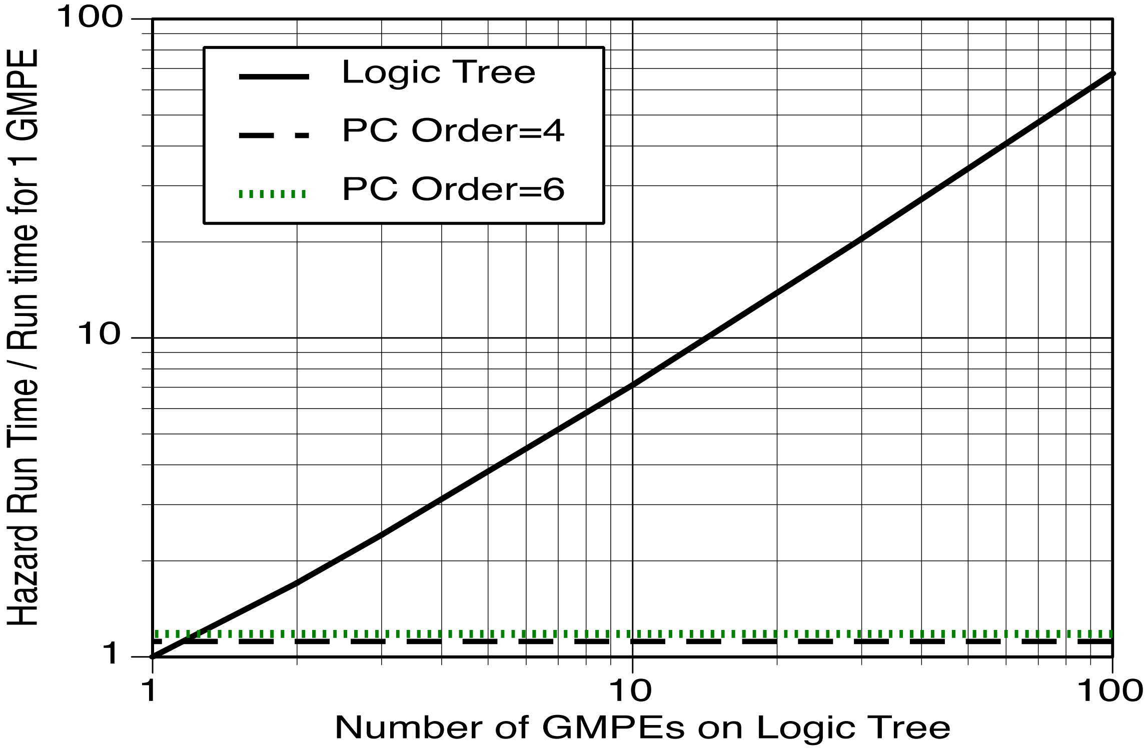

Figure 11: Comparison of hazard calculation time using the fourth-order PC expansion and logic trees as a function of the number of GMPE branches on the logic tree.

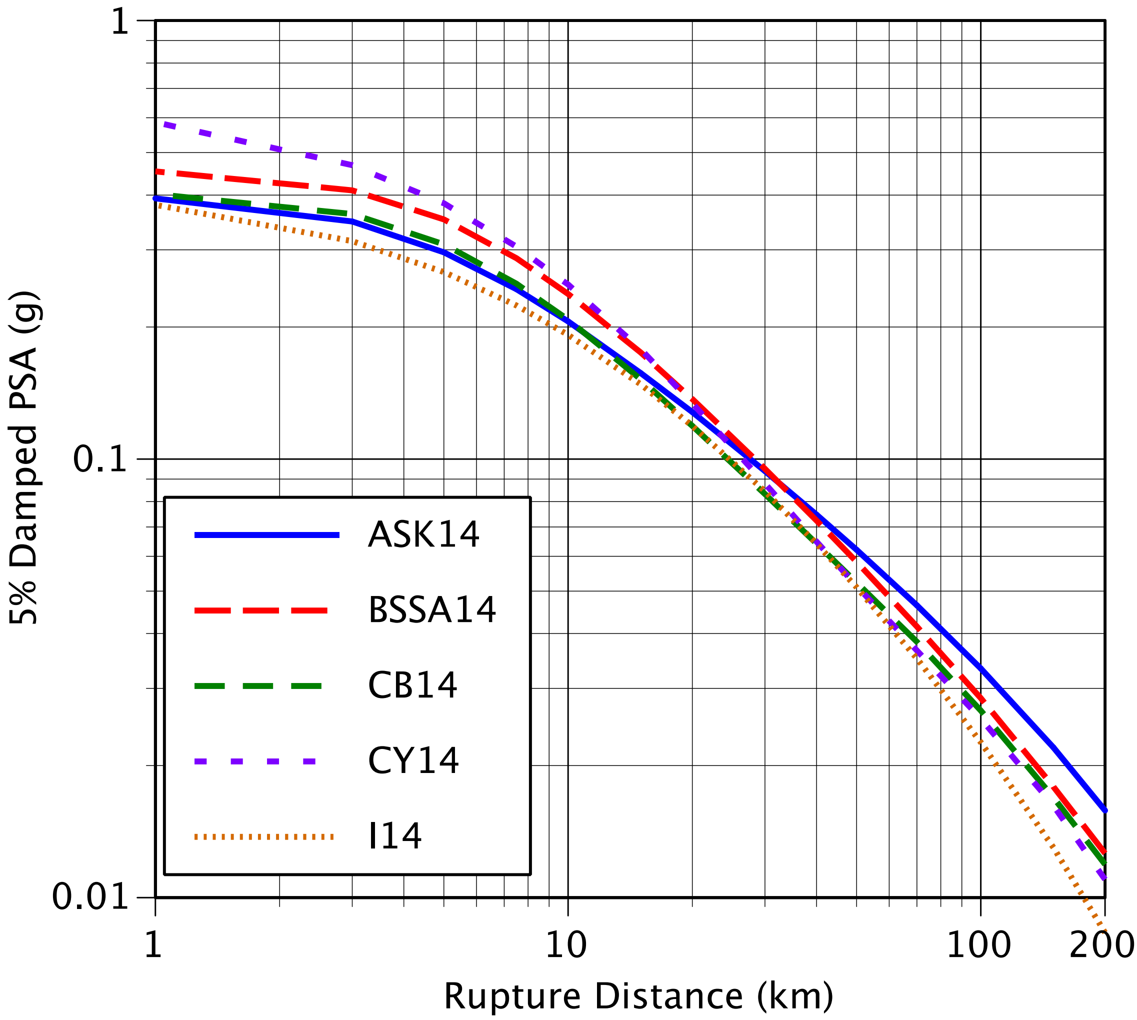

Figure 12: Example of the distance scaling for the five NGA-West2 ground-motion models for Magnitude 7 at a spectral period of 0.75 sec. The crossing of the curves shows the deviation from full correlation.

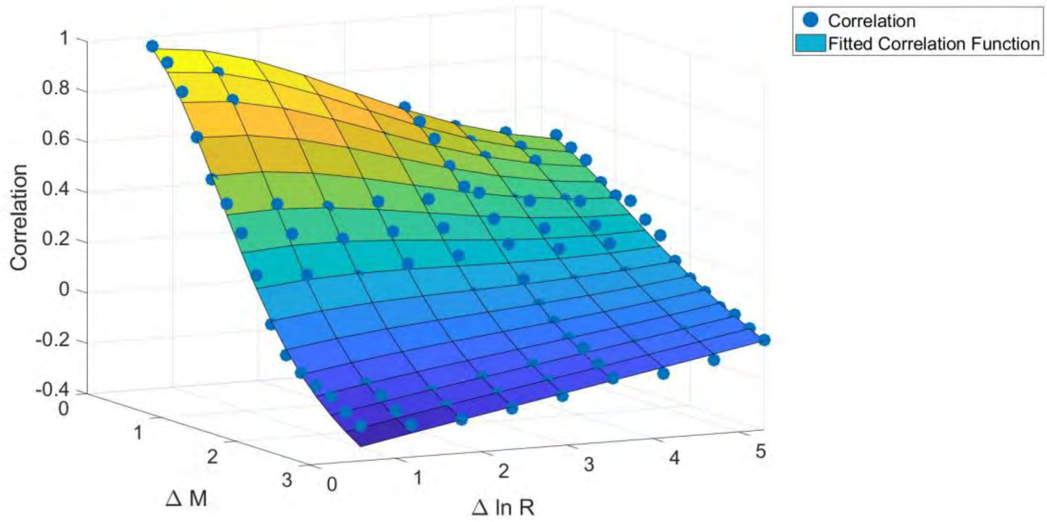

Figure 13: Point-estimates of the correlation and the fitted 2-D correlation function.

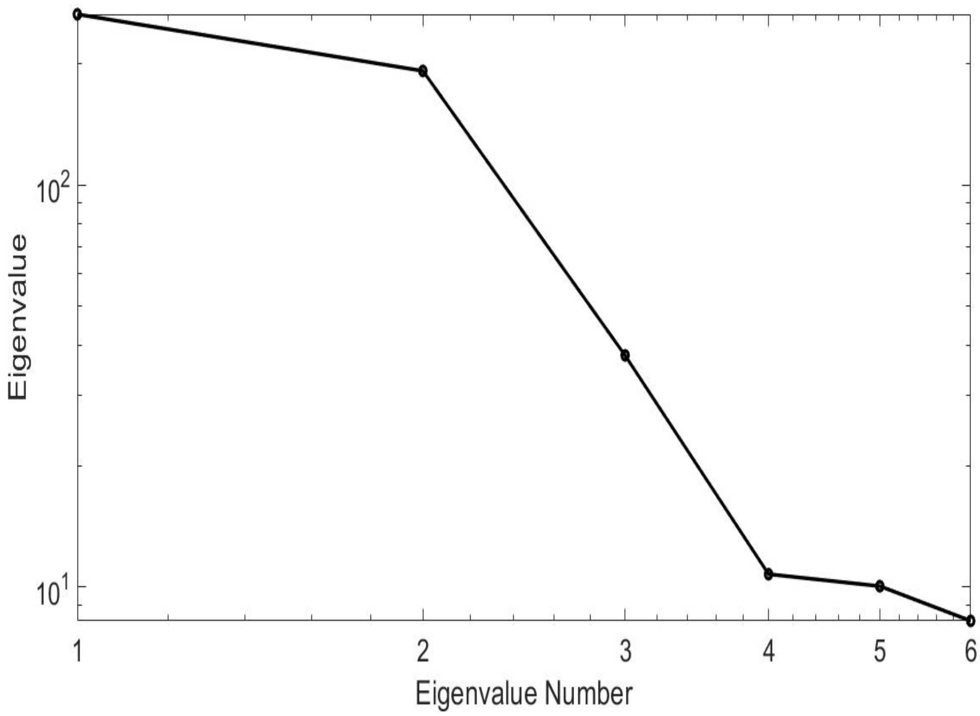

Figure 14: Largest six eigenvalues from the solution of finite-element analysis of equation (33)

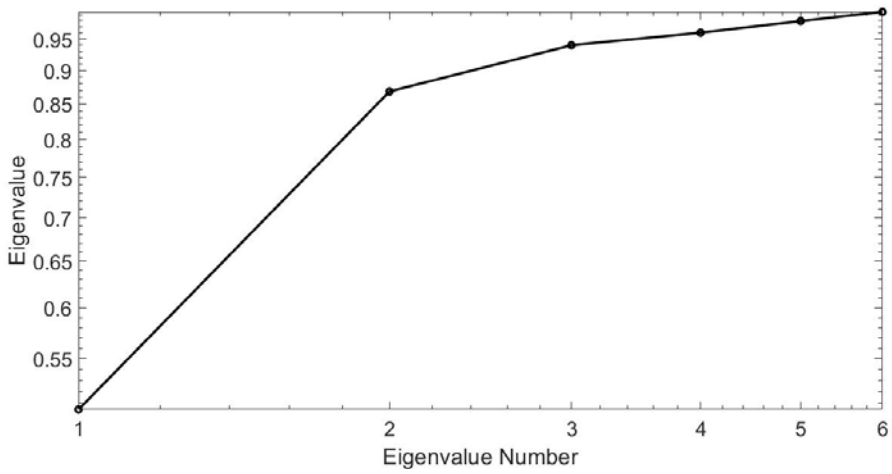

Figure 15: Cumulative Sum of First Six Eigenvalues

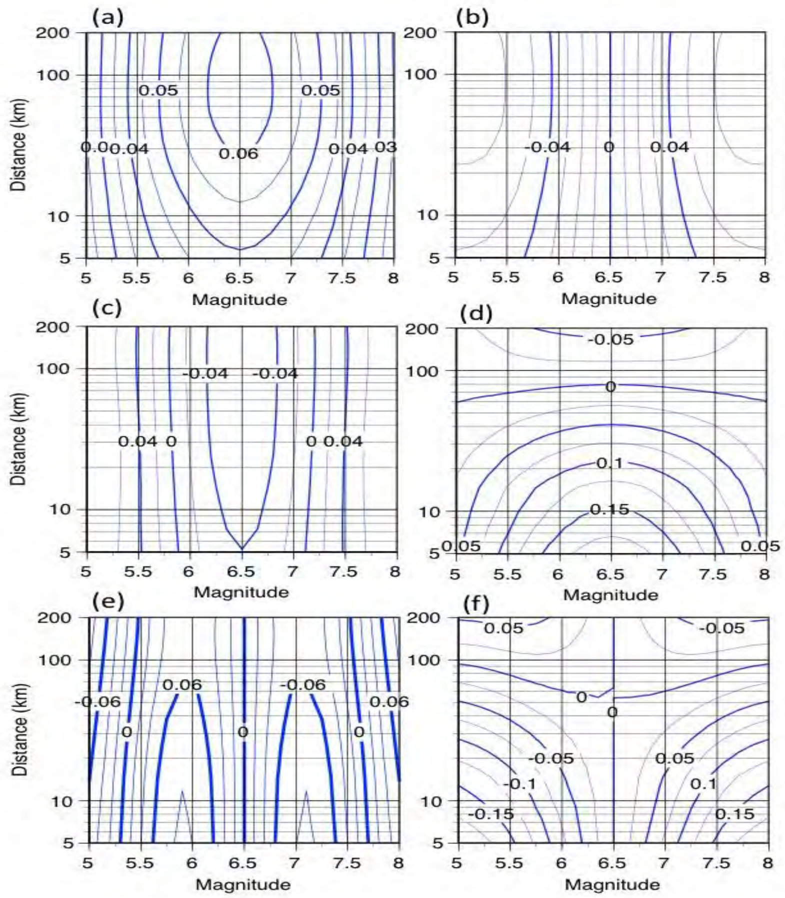

Figure 16: First Six Eigenfunctions

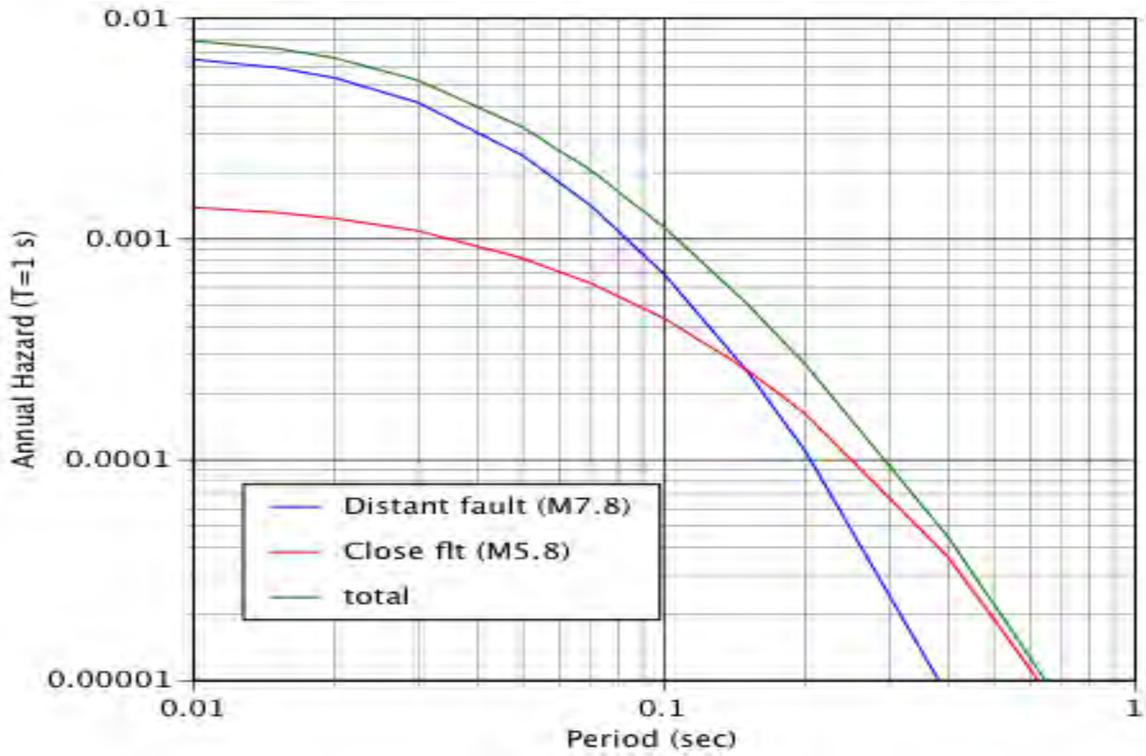

Figure 17: Hazard from the two sources for the example.

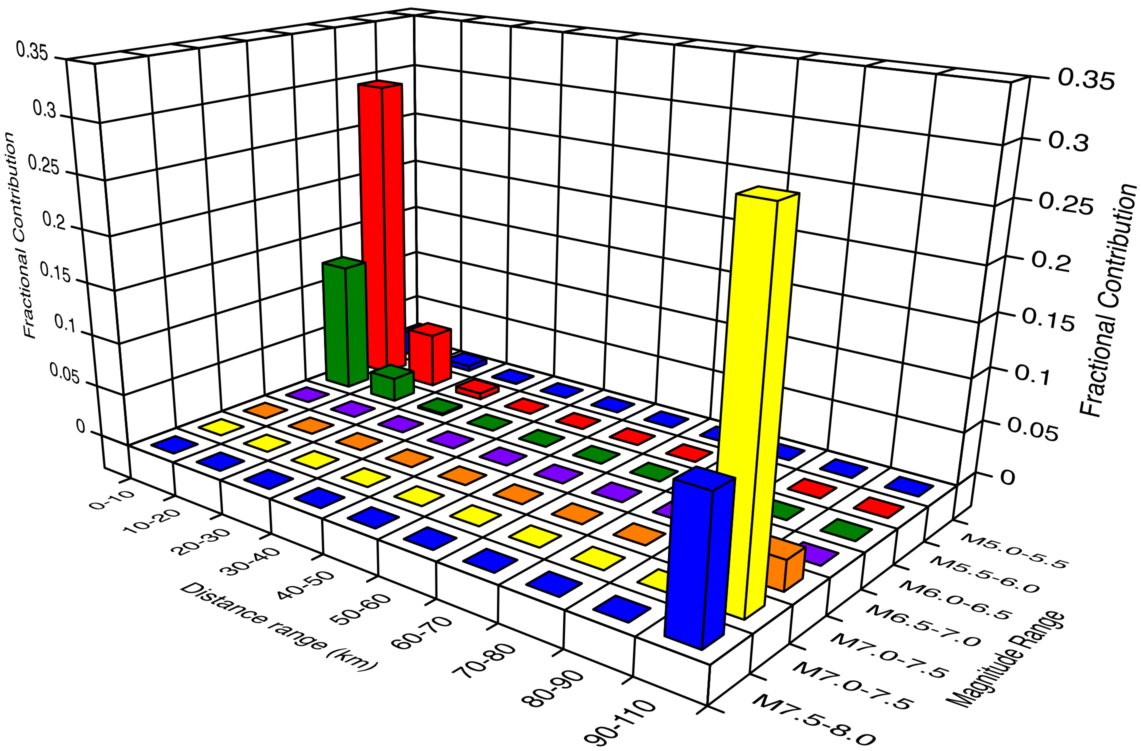

Figure 18: M-R disaggregation of the hazard at the 4E-4 hazard level.

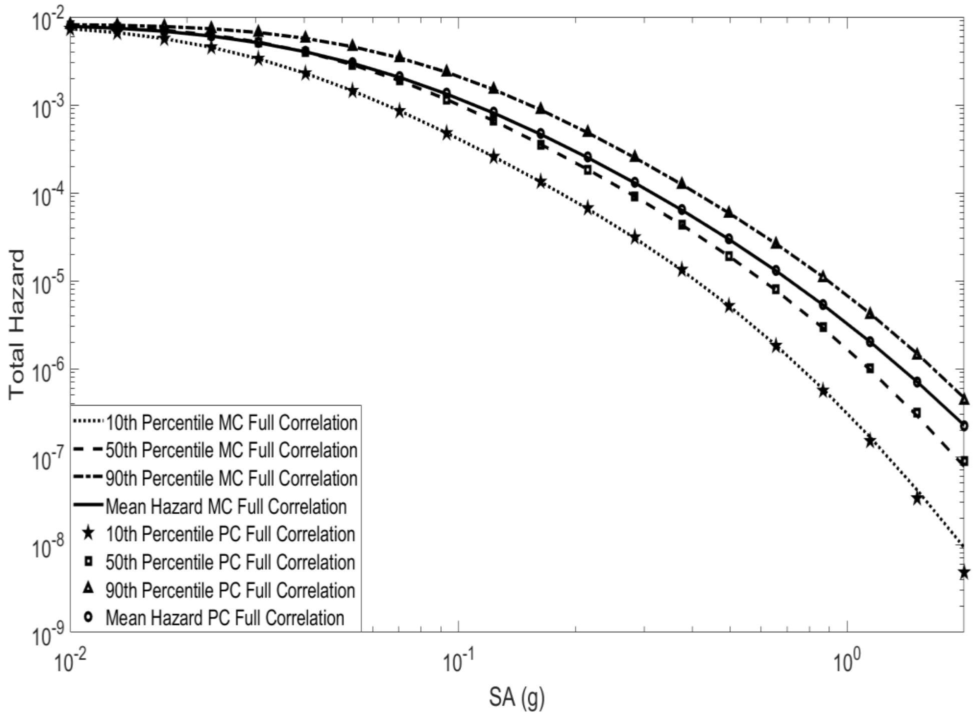

Figure 19: Percentiles of Total Hazard: Monte-Carlo with full correlation and with target correlation in GMMs across scenarios

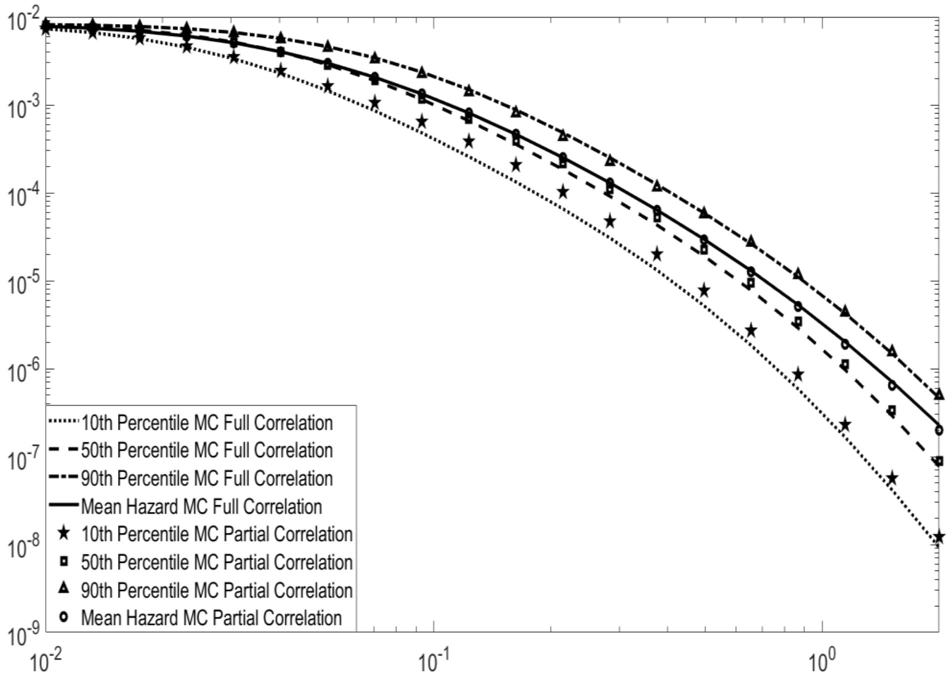

Figure 20: Percentiles of Total Hazard: Monte-Carlo with full and target correlation in GMMs across scenarios

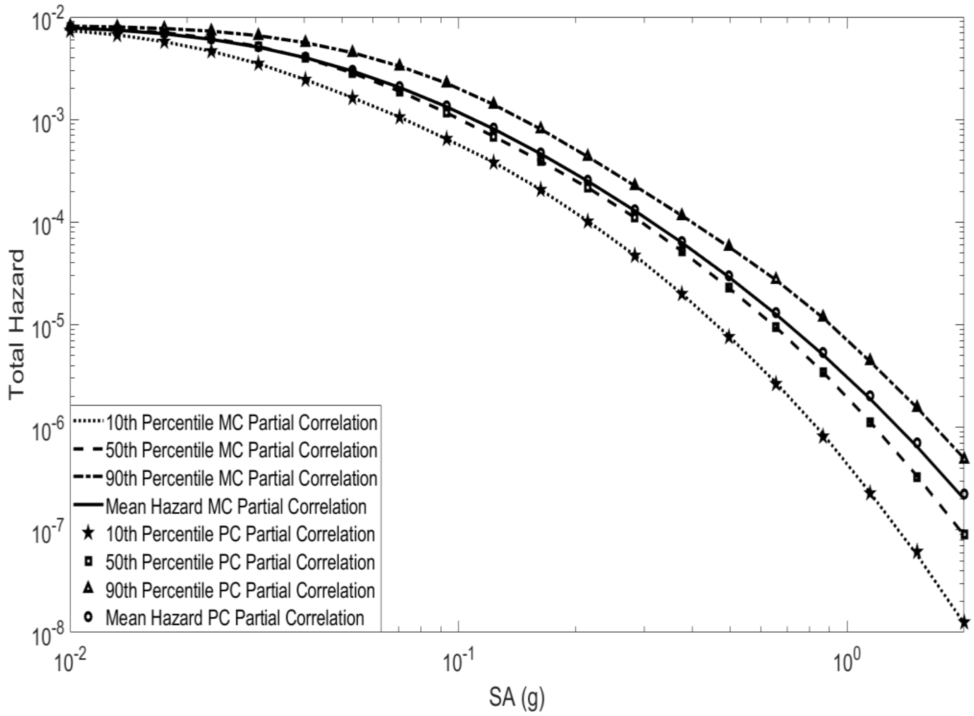

Figure 21: Percentiles of Total Hazard: Monte-Carlo vs Polynomial Chaos expansion with target correlation in GMMs across scenarios