## Title

Efficient Risk Calculation and Epistemic Uncertainty Propagation for Performance-Based Earthquake Engineering Distributed Systems



Authors
Lacour, Maxime
Abrahamson, norman

Publication Date
2025-12-05

DOI
10.55461/MFGK6090

PACIFIC EARTHQUAKE ENGINEERING RESEARCH CENTER

# Efficient Risk Calculation and Epistemic Uncertainty Propagation for Performance-Based Earthquake Engineering Distributed Systems

## A Report for the "Performance-based Earthquake Engineering Assessment Tool for Natural Gas Storage and Pipeline Systems" Project

**Maxime Lacour**
**Norman A. Abrahamson**

**Department of Civil and Environmental Engineering**
**University of California, Berkeley**

PEER Report 2025/08
Pacific Earthquake Engineering Research Center
Headquarters at the University of California, Berkeley
December 2025

**PEER 2025/08**
**December 2025**

# Efficient Risk Calculation and Epistemic Uncertainty Propagation for Performance-Based Earthquake Engineering Distributed Systems

## A Report for the "Performance-based Earthquake Engineering Assessment Tool for Natural Gas Storage and Pipeline Systems" Project


**Maxime Lacour**

**Norman A. Abrahamson**

Department of Civil and Environmental Engineering
University of California, Berkeley



PEER Report No. 2025/08
Pacific Earthquake Engineering Research Center
Headquarters, University of California, Berkeley
December 2025

# ABSTRACT

This report is one of a series of reports documenting the methods and findings of a multi-year, multi-disciplinary project conducted by the Pacific Earthquake Engineering Research Center (PEER) with the Lawrence Berkeley National Laboratory (LBNL) and funded by the California Energy Commission (CEC). The overall project is titled “Performance-based Earthquake Engineering As-sessment Tool for Natural Gas Storage and Pipeline Systems,” hereafter referred to as the ”Open-SRA Project.” The overall goal of the OpenSRA project is to create an open-source research-based seismic risk assessment tool for natural gas infrastructure that can be used by utility stakehold-ers to better understand state-wide risks, prioritize mitigation, plan new gas infrastructure, and help focus post-earthquake repair work. The project team includes researchers from LBNL, UC Berkeley, UC San Diego, University of Nevada Reno, the NHERI SimCenter at UC Berkeley, and Slate Geotechnical Consultants and its subcontractors Lettis Consultants International (LCI) and Thomas O’Rourke. Focused research to advance the seismic risk assessment tool was conducted by Task Groups, each addressing a particular area of study and expertise and collaborating with the other Task Groups. This report is the product of Task Group F: Synthesis of component fragilities into a system performance model.

**Keywords: risk, epistemic uncertainty, polynomial chaos**

# ACKNOWLEDGMENTS AND DISCLAIMER

This research study was funded by the California Energy Commission under Contract No. PIR 18-003. The opinions, findings, conclusions, and recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the California Energy Commission and its employees, the State of California, the Pacific Earthquake Engineering Research (PEER) Center, and the Regents of the University of California.

# CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# 1 Introduction

This report is one of a series of reports documenting the methods and findings of a multi-year, multi-disciplinary project conducted by the Pacific Earthquake Engineering Research Center (PEER) with the Lawrence Berkeley National Laboratory (LBNL) and funded by the California Energy Commission (CEC). The overall project is titled ”Performance-based Earthquake Engineering As-sessment Tool for Natural Gas Storage and Pipeline Systems,” hereafter referred to as the ”Open-SRA Project.”

The overall goal of the OpenSRA project is to create an open-source research-based seis-mic risk assessment tool for natural gas infrastructure that can be used by utility stakeholders to better understand state-wide risks, prioritize mitigation, plan new gas infrastructure, and help focus post-earthquake repair work.

The probabilistic seismic risk tool developed in this project follows the widely accepted risk methodology of Dr. A. Cornell (Cornell, 1968). A seismic source characterization is used to develop a suite of earthquake scenarios with associated rates of occurrence to represent the seismic hazard. Fault ruptures and the resulting ground deformation are generated for each earthquake scenario to represent the seismic loading, which includes a map of ground motion parameters. This scenario-based seismic parameter map is overlaid on the infrastructure system, and the seismic loading is combined with the capacities of the infrastructure to calculate the seismic performance of the natural gas system for the scenario. By repeating the process for all the scenarios in the suite, the tool can evaluate the seismic risk to the system.

A user-driven research approach was used to develop OpenSRA to be easily usable by regulators and utilities and to include updated models and methods for the seismic demands and capacities that control the seismic risk for natural gas systems. The project includes several innova-tive approaches that improve the basic methodology and distinguish this project’s approach from the standard approaches currently used. Current risk studies developed by the utilities use risk-scoring approaches that are highly subjective and qualitative. They do not properly incorporate the uncertainties in the seismic demand and in the fragility of the system and its components. Targeted research was conducted in this project to improve the characterization of uncertainty of key inputs to the seismic risk assessment tool. The seismic risk methodology employed in this project pro-vides quantitative estimates of the probabilistic seismic risk. For risk-informed decision-making processes, the reliability of the risk estimates needs to be considered because this can be significant, particularly for large, rare earthquakes.

The project team includes researchers from LBNL, UC Berkeley, UC San Diego, Univer-sity of Nevada Reno, the NHERI SimCenter at UC Berkeley, and Slate Geotechnical Consultants and its subcontractors Lettis Consultants International (LCI) and Thomas O’Rourke.

Focused research to advance the seismic risk assessment tool was conducted by Task Groups, each addressing a particular area of study and expertise and collaborating with the other Task Groups. The Task Groups are as follows:

- Task A: Fault Displacement
- Task B: Liquefaction-induced deformation and seismically induced slope displacement
- Task C: Performance of natural gas storage well casings and caprock
- Task D: Performance of gas storage and pipeline system surface infrastructure
- Task E: Smart gas infrastructure sensing of wells and pipeline connections performance
- **Task F: Synthesis of component fragilities into a system performance model**

This report is the product of the Task Group, as denoted in bolded text above. This report explains in detail the motivation, mathematical formulation, and numerical implementation of the development of a computationally very efficient tool for large-scale risk calculations of distributed systems, including the efficient propagation of the epistemic uncertainty in the fragility models.

# 2 Efficient Deterministic Performance-Based Earthquake Engineering Risk Calculation

## 2.1 GENERAL PBEE RISK FORMULATION

The methodology for performance-based earthquake engineering (PBEE) risk calculation was originally developed by the Pacific Earthquake Engineering Research (PEER) Center. This frame-work estimates risk as a function of ground shaking characterized by an intensity measure, IM (e.g., Peak Ground Acceleration, Peak Ground Velocity, etc.). The effect that the intensity measure has on a system is described by an engineering demand parameter, EDP (e.g., lateral spread displace-ment, liquefaction induced settlement, etc.), and the damage caused by the EDP is described by a damage measure, DM (e.g., pipe strain). Consequences of the damage are then assessed using a decision variable, DV (e.g., probability of pipeline rupture). The mathematical formulation of the risk following the PBEE methodology is given by:

$$\lambda_{dv}(DV > dv) = \iiint_{\mathcal{D}_{DM},\mathcal{D}_{EDP},\mathcal{D}_{IM}} P_{DV}(DV > dv|dm) f_{DM}(dm|edp) f_{EDP}(edp|im) f_{IM}(im) \, d(dm)\, d(edp)\, d(im) \tag{2.1}$$

The risk $\lambda_{dv}$ as given by Equation (2.1) represents, for one earthquake scenario, the probability of the decision variable DV to exceed a threshold value dv, conditioned on parameters IM, EDP, and DM. Because these parameters are random variables, the conditioning of DV by the parameters IM, EDP, and DM is mathematically expressed by probability density functions.

The function $f_{IM}$(im) represents the probability density function of the intensity measure IM at value IM = im. The intensity measure IM is typically a ground-motion parameter, such as spectral acceleration or spectral velocity, whose distribution is generally assumed to be lognormal and for which the median and standard deviation are estimated by ground-motion models (GMMs) for a given earthquake scenario.

The function $f_{EDP}$(edp|im) represents the conditional probability density function of the engineering demand parameter EDP at value EDP = edp, conditioned by the intensity measure IM at value IM = im. This conditioning captures the effect of the intensity measure IM on the probability of the engineering demand parameter EDP to occur.

Similarly, the function $f_{DM}(dm|edp)$ represents the conditional probability density function of the damage measure $DM$ at value $DM = dm$ conditioned by the engineering demand parameter value $EDP = edp$.

Finally, the function $P_{DV}(DV > dv|dm)$ represents the probability of the decision variable $DV$ to exceed the value $dv$ conditioned by the damage measure value $DM = dm$.

The risk $\lambda_{dv}(DV > dv)$ is further obtained by summing up the probabilities of occurrence of all the possible values $\{im, edp, dm\}$ of parameters $IM$, $EDP$ and $DM$ using their (conditional) probability density functions. This summation over all the values $\{im, edp, dm\}$ is mathematically represented by the triple integral in Equation (2.1).

An alternative of Equation (2.1) that is commonly used considers the DV to be a discrete variable, such as a damage state. In this case, the risk value is not given by a risk curve but by a single number, which is the probability that the decision variable $DV$ will occur. The mathematical formulation of the risk for this case is:

$$\lambda(DV) = \iiint_{\mathcal{D}_{DM},\mathcal{D}_{EDP},\mathcal{D}_{IM}} P_{DM}(DV|dm) f_{DM}(dm|edp) f_{EDP}(edp|im) f_{IM}(im) \, d(dm)\, d(edp)\, d(im) \tag{2.2}$$

In this report, we only consider the mathematical formulation of the risk curve following Equation (2.1); however, we show both formulations in the illustrative example section.

Computation using the above equations can be numerically very inefficient. To improve the numerical efficiency, we make some approximations of the risk formulation. We describe how to efficiently implement Equation (2.1). A Matlab code for the efficient risk calculation for (2.1) and (2.2) is available at `https://github.com/maxlacour/PC_Risk_Calculation`.

## 2.2 EFFICIENT REFORMULATION OF THE RISK EQUATION

Performance PBEE-based risk calculations presented in Equation (2.1) require the computation of a triple integral. This triple integral can be numerically approximated using traditional quadrature rules, such as the rectangle, trapezoidal, or Gauss quadrature rules. These numerical methods for integration can be computationally very efficient for evaluating one-dimensional integrals; however, they become exponentially more expensive in computation for evaluating multidimensional integrals. This increase in computation time with dimension is due to the exponentially increasing number of quadrature points required to discretize the multidimensional domain over which the integrals are evaluated.

One practical property of the PBEE-based risk calculations is that the conditional models of $IM, EDP, DM$, and $DV$ are typically one-dimensional functions of these parameters. For example, the intensity measure $IM$ is typically used for the conditioning of the median engineering demand parameter $EDP$ only and not for the conditioning of the other parameters ($DM$ and $DV$). Similarly, the engineering demand parameter $EDP$ is typically only used for the conditioning of the median damage measure $DM$, and the damage measure $DM$ is only used for the conditioning of the median decision variable $DV$.

We can take advantage of these one-dimensional relationships between parameters $IM$, $EDP$, $DM$, and $DV$ and rewrite the triple integral from Equation (2.1) with one-dimensional integrals only. Mathematically, the updated formulation of the risk becomes:

$$\begin{aligned}\lambda_{dv} = \int_{\mathcal{D}_{DM}} & P(DV > dv|dm)\ d(dm)\times \\ \int_{\mathcal{D}_{EDP}} & f(dm|edp)\ d(edp)\times \\ \int_{\mathcal{D}_{IM}} & f(edp|im) * f(im)d(im)\end{aligned} \tag{2.3}$$

in which a first integral is performed over the domain of $IM$ only, then a second integral is performed over the domain of $EDP$ only. Finally, a third integral is performed over the domain of $DM$ only.

Reformulating the risk equation as in Equation (2.3) allows us to avoid the expensive computation of a triple integral over the three-dimensional domain $\{IM, EDP, DM\}$ from Equation (2.1) using the quadrature rule. Instead, we now deal with three one-dimensional integrals over the domain of $\{IM\}$, $\{EDP\}$ and $\{DM\}$ separately. This reduces the computational complexity in the numerical integration from exponential to linear. Note that this reformulation of the risk from Equation (2.1) to Equation (2.3) is done without making any approximation on Equation (2.1) but by simply taking advantage of the mathematical properties of the PBEE-based risk analysis framework.

## 2.3 MOTIVATION FOR AN EFFICIENT APPROXIMATION OF THE PBEE RISK FOR LARGE-SCALE SYSTEMS

We reduced the computational complexity of the risk calculation from exponential to linear via the formulation of the main risk Equation (2.1) into Equation (2.3); however, we must still compute the three integrals from Equation (2.3). If they do not present analytical solutions, these integrals must still be approximated using the traditional quadrature rules described above. Evaluating integrals with quadrature rules is inefficient because it requires the repeated evaluation of the function inside the integral at the multiple quadrature points. When the functions inside the integrals involve computationally expensive terms such as exponentials or the Gaussian error function $\Phi$, as is the case in Equation (2.3), the approximation of the risk using quadrature rules becomes particularly inefficient.

Below, we illustrate the computational cost of evaluating Equation (2.1) for just the EDP hazard using the quadrature rule. That is, we only consider one integral in the main risk Equation (2.1), such as:

$$\lambda_{EDP}(EDP > edp) = \int_{\mathcal{D}_{IM}} P(EDP > edp|im)f(im)d(im) \tag{2.4}$$

For that example, we consider a single earthquake scenario of magnitude $M = 5.1$ and a single site at a distance $R$ = 20 km from the source. We use the quadratic rule to calculate the risk curve from Equation (2.4) for the EDP seismic slope displacement conditioned on the IM ground-motion spectral acceleration. For that earthquake scenario, the median and standard deviation of the ground motion are given by the NGA-West 2 GMMs and are equal to $\ln(SA)_{med} = -5.04$ and $\sigma(SA) = 0.77$.

The median of the seismic slope displacements is given by the model from Bray and Macedo (2019), which is shown in Figure 2.1.

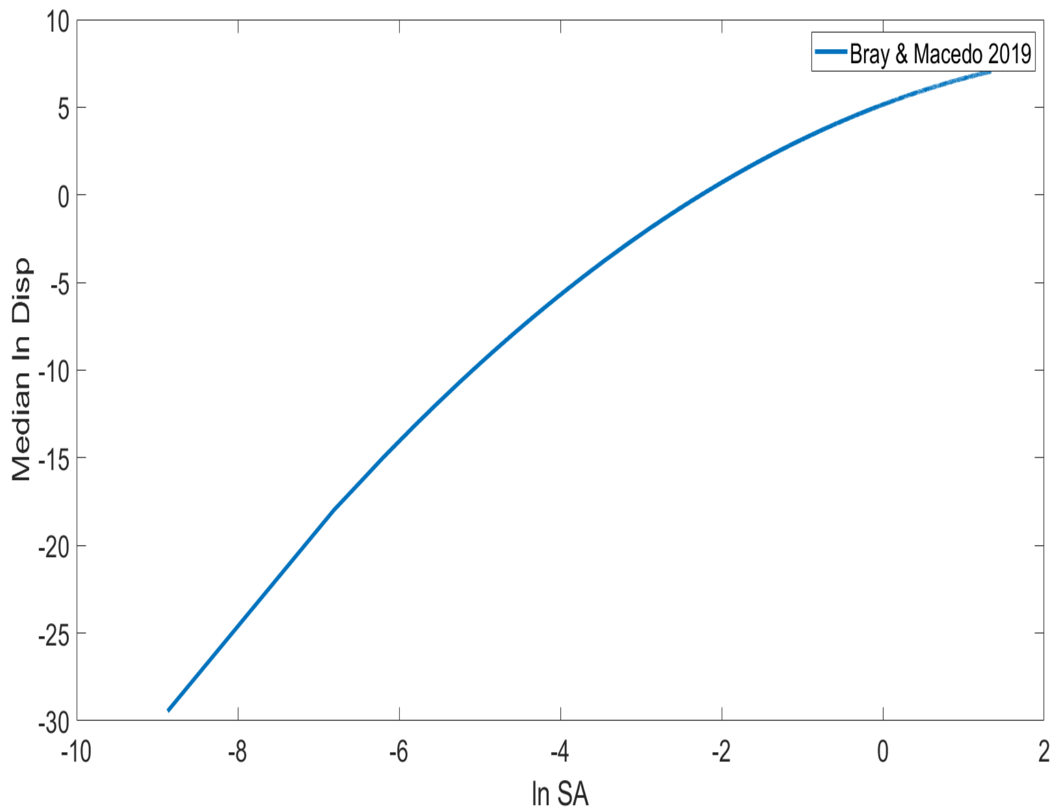


**Figure 2.1: Model From Bray and Macedo (2019) ($k_y = 0.1$, $T_s = 1s$)**

We calculate the conditional probability of exceedance (conditioned on the earthquake having occurred) using Equation (2.4) for the example earthquake scenario using the quadrature rule with 50, 100, and 200 quadrature points. The conditional probability of exceedance is complementary cumulative distribution function, CCDF, given that the scenario occurred. The CCDFs for the three quadratures are shown in Figure (2.2).

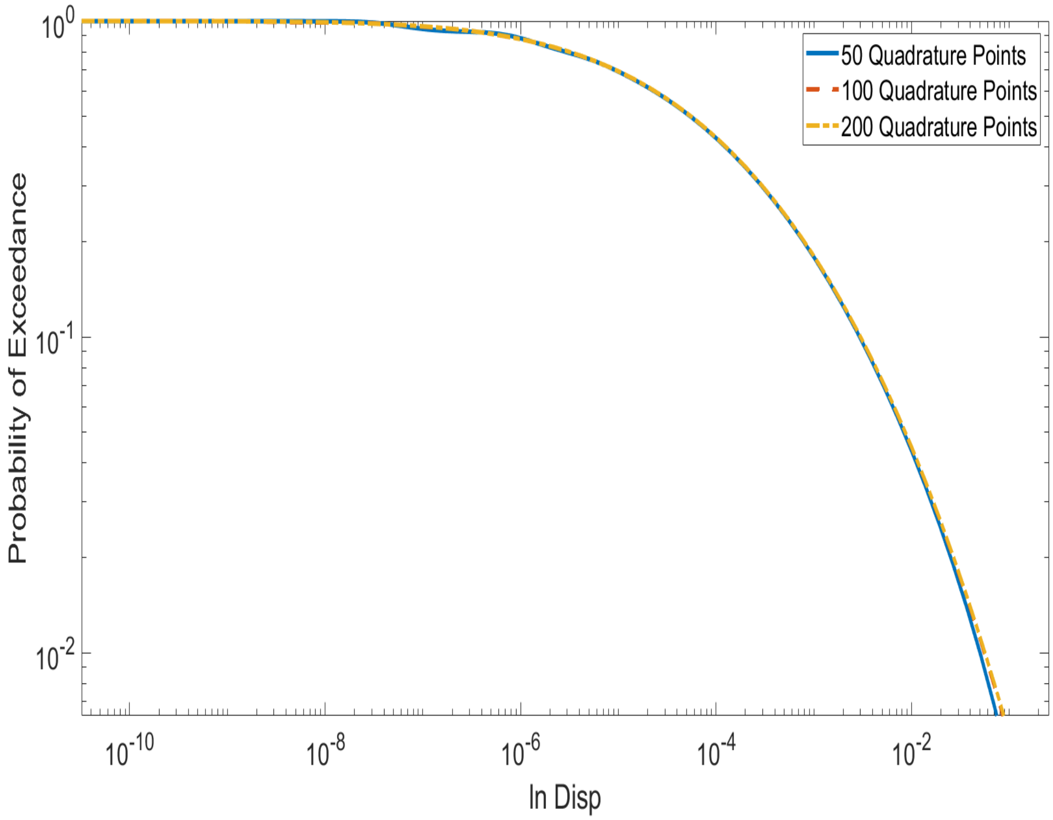


**Figure 2.2: Conditional probability of exceedance (CCDF) for one earthquake scenario using 50, 100, and 200 quadrature points)**

Figure 2.2 shows that with 50 quadrature points, we reach convergence of the CCDF with reasonable accuracy (less than 10% error) in the lower tail, (e.g., $10^{-2}$) when compared with the risk curve obtained with 100 and 200 quadrature points.

The computational times required to calculate the three CCDFs with 50, 100, and 200 quadrature points are approximately $t_{50} = 0.002s$, $t_{100} = 0.004s$, and $t_{200} = 0.008s$ in Matlab. This computational time represents the time required to calculate one CCDF at one site for one earthquake scenario and using only a single integral from the main risk Equation (2.1). However, for a regular hazard calculation, the number of earthquake scenarios easily reaches 1,000 for rupture scenarios along faults and 10,000 for background zones. Additionally, for large-scale systems, the number of locations at which we are interested in calculating the risk can also easily reach 10,000 site locations.

Scaling the computational time $t_{50}$ by the typical number of earthquake scenarios and number of sites and multiplying by 3 to account for the triple integral in Equation (2.1) leads to the following rough estimate of the total computational time of a risk calculation for large scale systems using the quadrature rule:

$$t_{quadrature50} = 0.002 \times 1,000 \times 10,000 \times 3 \simeq 16h \tag{2.5}$$

The rough estimate of a regular PBEE-based risk calculation for large-scale systems is on the order of a day. That computational time does not include the propagation of the epistemic uncertainty from the median models, which would increase the calculation time by another order of magnitude if the standard logic-tree approach is used. Therefore, more efficient numerical methods and/or approximations of the risk need to be employed to reduce that computational time to the order of minutes.

## 2.4 EFFICIENT PBEE RISK APPROXIMATION USING ANALYTICAL SOLUTIONS

### 2.4.1 Requirements for Analytical PBEE Risk Solutions

The high computational cost of calculating a CCDF using quadrature rules comes from the repeated evaluation of the function inside the integral at the multiple quadrature points. For the example of 50 points, that means that the function below the integral in Equation (2.4) needs to be evaluated 50 times to calculate the CCDF with some reasonable accuracy in the tail, which further explains the high computational cost of the total risk calculation.

The use of quadrature rules can be avoided if the integrals in Equation (2.3) could be solved analytically. In that case, the integrals in Equation (2.3) could be directly obtained by computing their analytical solution in just one numerical evaluation, further reducing the computational expenses by several orders of magnitude.

We present an approximation that allows us to have analytical solutions for the integrals in Equation (2.3) to gain computational efficiency. We then illustrate this approximation with the example of a single earthquake scenario and show the loss in accuracy and the reduction in computational time between the analytical and numerical evaluations of the integrals in the risk calculations.

To find analytical solutions for the PBEE-based risk calculation in Equation (2.3), we first replace the probability density functions by their mathematical expressions and then try to find under what conditions the integrals may have analytical solutions. The multiple random variables $IM, EDP, DM, DV$ involved in the risk are typically assumed to follow normal or lognormal distributions. For example, if they are assumed to be log-normally distributed, their probability density and distribution functions are given by:

$$f_X(x) = \frac{1}{\sqrt{2\,\pi}\sigma_X\, x}\, exp\left(-\frac{(\ln(x) - m_X)^2}{2\sigma_X^2}\right) \tag{2.6}$$

$$F_X(X < x) = \Phi\left(\frac{\ln(x) - m_X}{\sigma_X}\right) \tag{2.7}$$

in which $m_X$ and $\sigma_X$ represent the median and the standard deviation of the random variable $X$, and $\Phi$ is the cumulative distribution function of the standard normal variable given by:

$$\Phi(\epsilon) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\epsilon} e^{-0.5t^2} dt \tag{2.8}$$

For the conditional probability density functions, the conditioning of the random variable $X$ on the random variable $Y$ is typically only performed in the median $m_Y$ of the random variable $Y$. For the case in which $Y$ is log-normally distributed, the probability density function of Y conditioned by X takes the following form:

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\,\pi}\sigma_Y\; y}\; exp\left(-\frac{(\ln y - m_Y(x))^2}{2\sigma_Y^2}\right) \tag{2.9}$$

Substituting these probability density functions into the main risk equation 2.3, we obtain:

$$\begin{aligned}
\lambda_{dv}(DV > dv) = \int_{\mathcal{D}_{DM}} & \Phi\left(\frac{\ln{(dv)} - m_{DV}(dm)}{\sigma_{DV}}\right) d(dm) \\
& \int_{\mathcal{D}_{EDP}} \frac{1}{\sqrt{2\,\pi}\sigma_{DM}\; dm} exp\left(-\frac{(\ln{(dm)} - m_{DM}(edp))^2}{2\sigma_{DM}^2}\right)\; d(edp) \\
& \int_{\mathcal{D}_{IM}} \frac{1}{\sqrt{2\,\pi}\sigma_{EDP}}\; exp\left(-\frac{(\ln{(edp)} - m_{EDP}(im))^2}{2\sigma_{EDP}^2}\right) \\
& \frac{1}{\sqrt{2\,\pi}\sigma_{IM}\; im}\; exp\left(-\frac{(\ln{(im)} - m_{IM})^2}{2\sigma_{IM}^2}\right) d(im)
\end{aligned} \tag{2.10}$$

The integrals in Equation (2.10) have analytical solutions if and only if the terms in the exponent remain quadratic. That is because the integral of the function $x \mapsto e^{-x^2}$ has an analytical solution over the set of real numbers $\mathbb{R}$ (the Gaussian error function $\Phi$), but the integral of the function $x \mapsto e^{-x^3}$ does not. Consequently, the requirement for having analytical solutions for (2.10) is to keep the conditional median models $m_Y(x)$ to be polynomials of degree one (or linear functions) in the conditioning variable. In this way, the terms $(\ln(y) - m_Y(x)^2)$ in the exponents remain quadratic in the conditioning variable $X$, and analytical solutions for the integrals are available.

A linear conditioning of random variable $X$ on the median of random variable $Y$ by a first-order polynomial is given by:

$$m_Y(x) = a_Y * \ln{(x)} + b_Y \tag{2.11}$$

in which $a_Y$ and $b_Y$ are constant coefficients.

Replacing the linear conditioning above from Equation (2.11) for variables $EDP$, $DM$ and $DV$ into the risk equation (2.10) leads to:

$$
\begin{aligned}
\lambda_{dv}(DV > dv) = &\int_{\mathcal{D}_{DM}} \Phi\left(\frac{\ln(dv) - (a_{DV}\ln(dm) + b_{DV})}{\sigma_{DV}}\right) d(dm) \\
&\int_{\mathcal{D}_{EDP}} \frac{1}{\sqrt{2\,\pi}\sigma_{DM}\, dm} exp\left(-\frac{(\ln(dm)) - (a_{DM}\ln(edp) + b_{DM}))^2}{2\sigma_{DM}^2}\right)\, d(edp) \\
&\int_{\mathcal{D}_{IM}} \frac{1}{\sqrt{2\,\pi}\sigma_{EDP}}\, exp\left(-\frac{(\ln(edp)) - (a_{EDP}\ln(im) + b_{EDP}))^2}{2\sigma_{EDP}^2}\right) \\
&\frac{1}{\sqrt{2\,\pi}\sigma_{IM}\, im}\, exp\left(-\frac{(\ln(im) - m_{IM})^2}{2\sigma_{IM}^2}\right) d(im)
\end{aligned}
\tag{2.12}
$$

The integrals in the risk equation as in (2.12) all involve polynomials of degree two in the exponentials, and therefore analytical solutions are available for these integrals. Since all these integrals involve Gaussian density and distribution functions, the final analytical solution for the risk is itself a Gaussian distribution function. That is, the risk from Equation (2.12) is also equal to:

$$
\lambda_{dv}(DV > dv) = \Phi\left(\frac{\ln(dv) - m_{DV}^a}{\sigma_{DV}^a}\right) \tag{2.13}
$$

in which $m_{DV}^a$ and $\sigma_{DV}^a$ are the analytical mean and standard deviation of the random variable $DV$. The parameters $m_{DV}^a$ and $\sigma_{DV}^a$ are functions of all the input parameters in Equation (2.12), which are the mean, standard deviation, and conditioning parameters of the variables $IM$, $EDP$, $DM$, and $DV$.

The analytical solution for $\lambda_{dv}$ could, in theory, be obtained analytically by finding the direct mapping between $m_{DV}^a$ and $\sigma_{DV}^a$ and the parameters from Equation (2.12); however, this mapping will have a high computational cost because the three integrals from Equation (2.12) are nested, and their analytical solutions will involve many repeated terms from the solution of each integral individually. For that reason, it is much more efficient to numerically evaluate the analytical solution of each integral one at a time in Equation (2.12) in an iterative manner rather than trying to evaluate the analytical solution of the three nested integrals directly. In other words, the most efficient way to evaluate Equation (2.12) is to numerically evaluate the analytical solution of the first integral over $IM$, to use this numerical evaluation as an input to the analytical solution of the second integral over $EDP$, evaluate this second integral and use it as an input for the analytical solution of the third integral over $DM$ and finally evaluate the analytical solution of this third integral.

Therefore, we do not provide the direct analytical solution for the risk as in Equation (2.13) but instead provide the analytical solution of each integral separately. These analytical solutions are provided in Appendix (A).

### 2.4.2 Approximation Trade-off between Computational Efficiency and Accuracy

In practice, the conditional model from a random variable $X$ on a random variable $Y$ is not a linear function but is typically a polynomial function of higher degrees. For example, the model from Bray and Macedo (2019), as illustrated in Figure (2.1), gives the median seismic slope displacement $\hat{d}$ ($EDP$) as a polynomial function of degree 2 of the value of the ground-motion spectral acceleration $S_a$ ($IM$ ), as well as other geotechnical parameters. The model from Bray and Macedo (2019) is given by, for the case of an earthquake of magnitude $M > 7$:

$$\begin{aligned} \ln{(\hat{d})} =& a_1 - 2.83 \ln k_y - 0.333 (\ln k_y)^2 + 0.566 \ln k_y \ln(S_a) + \\ & 3.04 \ln S_a - 0.244 (\ln S_a)^2 + a_2 T_s + 0.278 (M - 7) \end{aligned} \tag{2.14}$$

When replacing the model from Bray and Macedo (2019) into the risk Equation (2.12), the quadratic term in the exponent leads to a polynomial of degree four in the variable $IM$, and therefore there is no analytical solution to the integrals from Equation (2.12).

Even if the model from Bray and Macedo (2019) is not a first-order polynomial in $S_a$, it can be well approximated locally by a linear function. The linear approximation of the median models is a reasonable approximation for two main reasons: 1) the median models vary quite smoothly with their input variables, and 2) we only use these models over a localized region of interest, not over their entire domain.

The region of interest of each parameter will be specific to each earthquake scenario. That means we do not replace the models by a linear function over the entire domain, but we do so only over a small region where the model matters for one particular earthquake scenario.

The region of interest can reasonably be defined to be two standard deviations around the median value of the conditioning parameter. The probability that the conditioning parameter falls inside this interval is about $95\%$. Outside the region of interest, the probability of the conditioning parameter to occur is close to zero, and the risk associated with the corresponding median values of the conditioned parameter will have a negligible contribution to the sum of the risk as given in Equation (2.10). Over each region of interest, the median models of $IM$, $EDP$, $DM$, and $DV$ can be reasonably well approximated by a linear function.

We choose to perform the linear approximation of the median models by taking their tangent at the center of the region of interest, which corresponds to the median value of the conditioning parameter. That corresponds to an exact representation of the model at the median of the conditioning parameter, which is the most likely value to occur. The error between the true model and the tangent approximation increases as we move further from the median. However, we could also focus on other regions of the model, for example, on the upper tail of the model, which would lead to a less accurate approximation by the tangent on the upper part of the risk curve but to a more accurate approximation on the lower tail of the risk curve.

We illustrate the linear approximation with the model from Bray and Macedo (2019) with a hypothetical earthquake scenario of magnitude $M = 5$ at a distance $R = 20km$ from a site. For

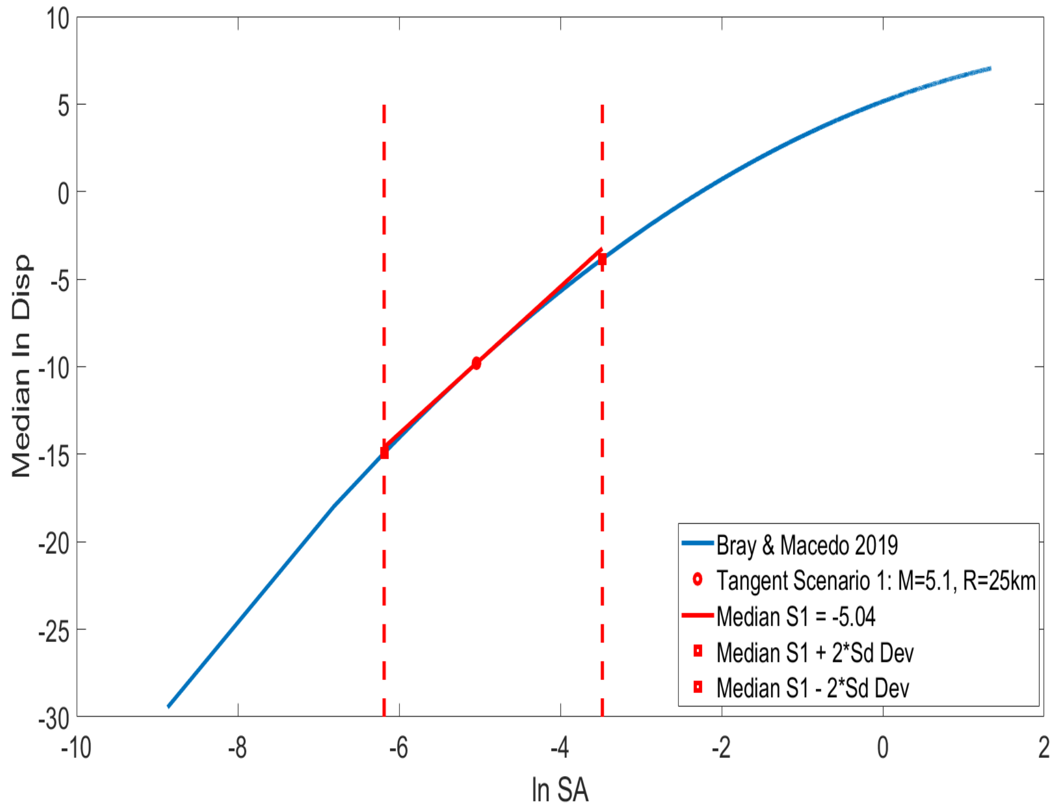


**Figure 2.3: Model from Bray and Macedo (2019) and its linear approximation at the median ground-motion for a given earthquake scenario** ($M = 5, R = 20km$)

that earthquake scenario, the median and standard deviation of the ground-motion are given by the NGA-West 2 GMMs (Gregor et al., 2014) and are equal to $\ln(SA)_{med} = -5.04$ and $\sigma(SA) = 0.77$. Figure (2.3) shows the model from Bray and Macedo (2019) and its tangent approximation at the median ground-motion value over the domain of interest, the $95\%$ confidence interval obtained by two standard deviations around the median value.

To check the accuracy of the linear approximation, we compute the CCDF from the model from Bray and Macedo (2019) with the quadrature rule using 100 points, and we compare it with the CCDF obtained from the linear approximation. The CCDF using the tangent approximation is obtained analytically from the equations derived in Appendix (A). These two curves are shown in Figure 2.4.

Figure 2.4 shows that the linear approximation looks reasonable over the domain of interest. The error between the risk curve from the true model from Bray and Macedo (2019) and its linear approximation starts growing as we reach the tail of the risk curve. This error comes from the difference between the true median model and its linear approximation at the right tail, as observed in Figure (2.3).

In terms of computational expenses, the calculation of the risk curve $t_{analytical}$ using the

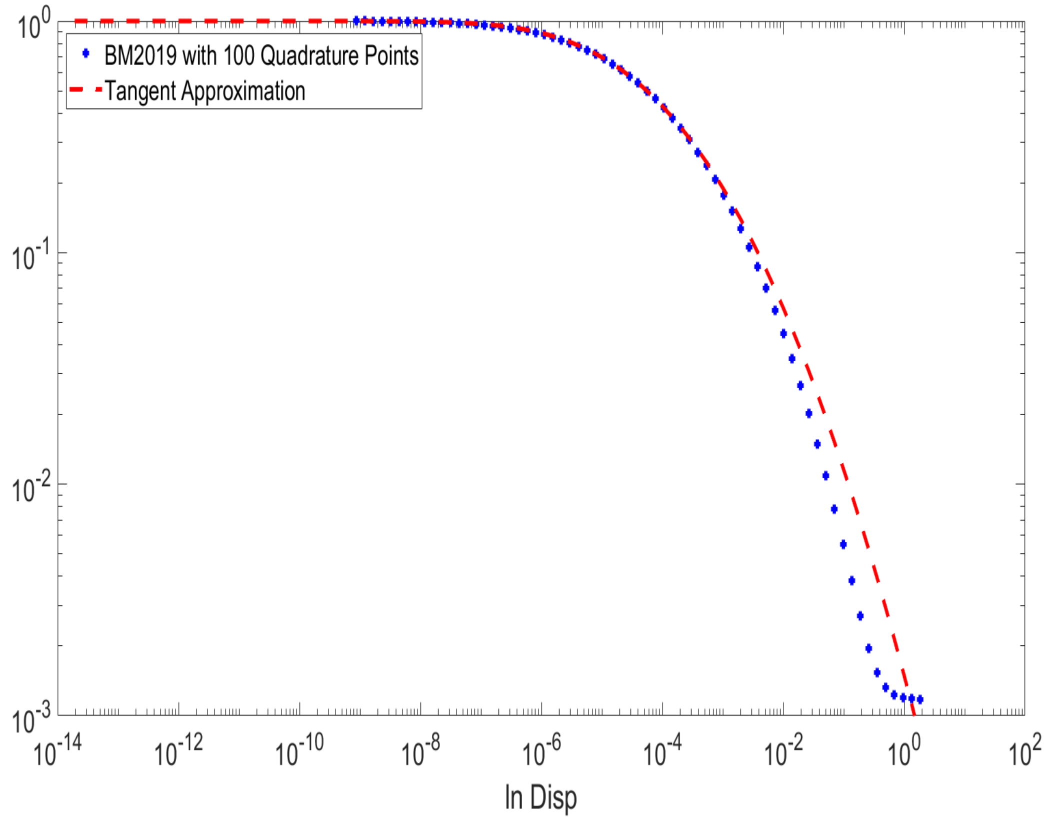


**Figure 2.4: Comparison of the CCDF using the Bray and Macedo (2019) model with 100 quadrature points and linear analytical approximation.**

analytical solution as given in Appendix (A) is approximately (also in Matlab):

$$t_{analytical} \simeq 0.0001s << t_{quadrature50} \simeq 0.008s$$

As a rough estimate, we can now predict that a risk calculation using analytical solutions will be faster by about two orders of magnitude than using the quadrature method, bringing the total computational time of a large-scale risk calculation to the order of minutes, which now becomes reasonable. The illustrative example section will give further details on the computational times of true risk calculations.

# 3 Efficient Propagation of Epistemic Uncertainty in Median Models into the PBEE Risk Equation

Epistemic uncertainty in the various models used in PBEE risk calculations arises from different sources, including alternative models of the behavior of the physical phenomena and a lack of data to constrain the inputs to the models.

The epistemic uncertainty in the models needs to be taken into account and propagated into risk calculation using numerical methods. The traditionally used approach to propagate epistemic uncertainties into civil engineering systems is the logic-tree approach, which consists of a weighted sampling of the alternative models and uncertain input parameters, including all combinations of the branches of the logic tree for the set of logic tree nodes. Considering all combinations of the logic-tree branches (called the full enumeration approach) can be computationally prohibitive for large logic trees. As an alternative, a Monte-Carlo approach that takes random paths through the full logic tree can be used. The Monte-Carlo approach still requires a large number of simulations before converging to stable results. The large number of simulations required does not make Monte-Carlo simulations practical for developing seismic designs for the pipeline, for which risk calculations should be obtained quickly to evaluate the performance of the alternative designs.

Other approaches were developed more recently to avoid the large computational requirements from Monte-Carlo simulations, particularly the spectral approach. The idea behind the spectral approach is to approximate the uncertain solution using a set of basis functions. When these basis functions are taken as polynomials, spectral approaches lead to the Polynomial Chaos (PC) approximation of the solution. Applications of the spectral approach for uncertainty quantification were first found in the field of numerical simulations, such as in the stochastic finite-element method, originally developed by Ghanem and Spanos (1991).

More recently, the spectral approach was also applied to seismic hazard calculations by Lacour and Abrahamson (2019) to efficiently propagate the epistemic uncertainty in median ground-motion models into hazard codes and avoid the heavy computational expenses of a large number of Monte-Carlo simulations. Similarly, Macedo et al. (2020) used PC to efficiently propagate the epistemic uncertainty in ground-motion models and seismic-slope displacement models into risk calculations.

We present here a new spectral approach based on the PC method from Lacour and Abrahamson (2019), which allows the efficient propagation of the epistemic uncertainty in median $IM$,

$EDP$, $DM$, and $DV$ models in the framework of PBEE risk calculations from Equation (2.1).

## 3.1 POLYNOMIAL CHAOS EXPANSION OF COMPLEMENTARY CUMULATIVE DISTRIBUTION FUNCTIONS AND PROBABILITY DENSITY FUNCTIONS

### 3.1.1 Polynomial Chaos Expansion of Complementary Cumulative Distribution Functions

Consider a random variable $X$, normally or log-normally distributed, with a median $\mu_X$ and a standard deviation $\sigma_X$. To model the epistemic uncertainty in the median model $\mu_X$, we make the assumption that the median is normally distributed, centered on the mean median, $m_{\mu X}$, with standard deviation $\sigma_{\mu_X}$, such that:

$$\mu_X = m_{\mu_X} + \sigma_{\mu_X}\, \xi \tag{3.1}$$

in which $\xi$ is a standard normal random variable.

The probability of exceedance of a log-normally distributed random variable $X$ with epistemic uncertainty in its median $\mu_X$ as defined above can be written as:

$$P(X > x) = 1 - \Phi\left(\frac{\ln(x) - (m_{\mu_X} + \sigma_{\mu_X}\xi)}{\sigma_X}\right) \tag{3.2}$$

The conditional probability of exceedance is the complementary cumulative distribution function (CCDF) of $X$. As an example, the CCDF for the spectral acceleration intensity measure, $SA$, is shown in Figure (3.1) for several values of $\xi$, which represent alternative samples of the epistemic uncertainty.

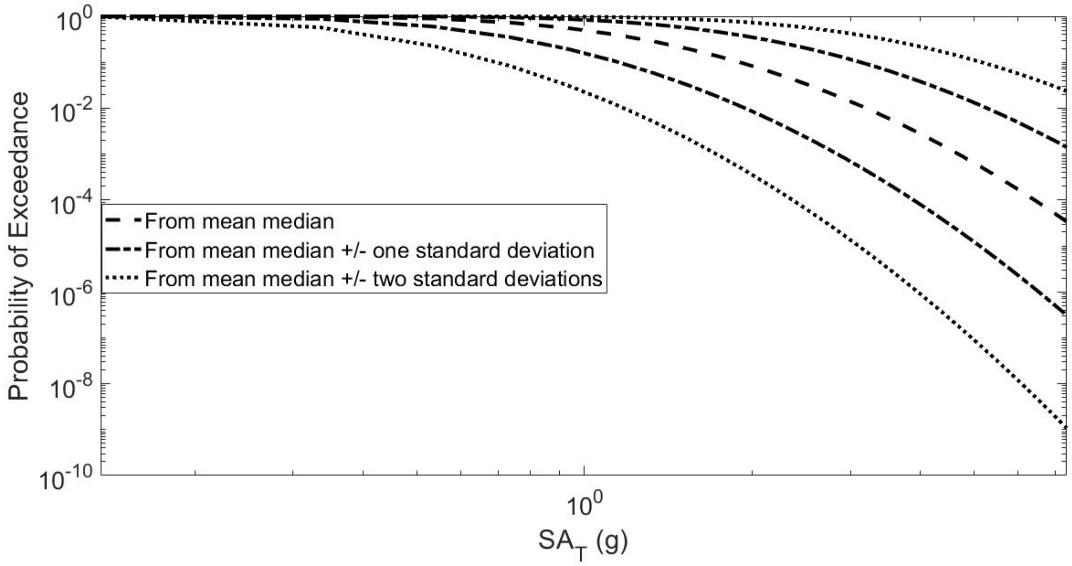


**Figure 3.1: Example of the conditional probability of exceedance ($P(SA > z|M, R)$ with epistemic uncertainty in the median with $m_\mu = 0$, and $\sigma = 0.5$ ln units**

The epistemic uncertainty in the median leads to fixed shifts in the CCDF along the X-axis. The y-axis values of the CCDF with epistemic uncertainty in the median can be approximated with the Hermite PC, as described in Lacour and Abrahamson (2019):

$$P(X > x) \approx \sum_{i=0}^{P} C_i(x)\Psi_i[\{\xi\}] \tag{3.3}$$

In Equation (3.3), the terms $\Psi_i[\xi]$ are the Hermite polynomials of the standard normal random variable $\xi$. The histograms of 10,000 realizations of the first four Hermite Polynomials are shown in Figure (3.2).

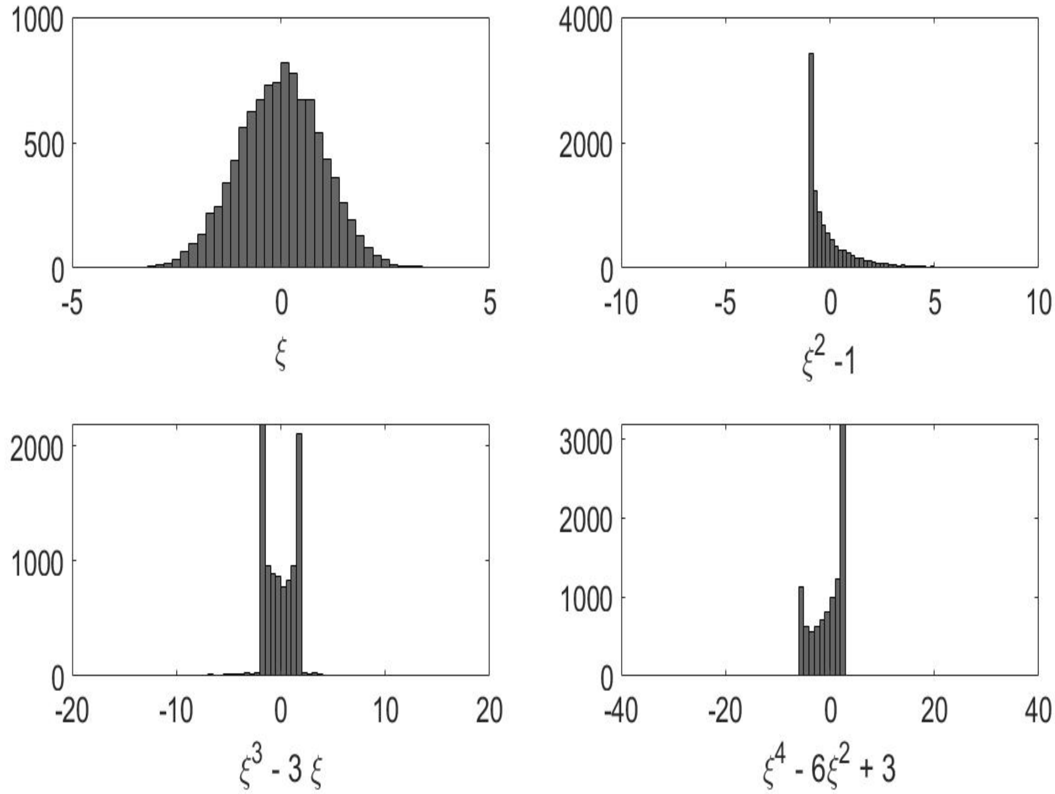


**Figure 3.2: Realizations of first, second, third, and fourth Hermite polynomials (N = 10 000)**

The CCDF of $X$ can be approximated by a linear combination of the Hermite polynomials following Equation (3.3), for which the coefficients $C_i(x)$ are given by the projection of the distribution on these polynomials:

$$\begin{aligned}\langle P(X > x) \cdot \Psi_i[\{\xi\}]\rangle = &\int_{\xi=-\infty}^{+\infty} \left[1 - \Phi\left(\frac{\ln(x) - (m_{\mu_X} + \sigma_{\mu_X}\xi)}{\sigma_X}\right)\right] \\ &\times \Psi_i[\{\xi\}] \times \frac{1}{\sqrt{(2\pi)}} e^{-\xi^2/2} d\xi\end{aligned} \tag{3.4}$$

Solving the integrals in Equation (3.4) using integration by parts, we obtain the coefficient of the Hermite PC expansion of a CCDF with epistemic uncertainty in the median as in Equation (3.2). The coefficients up to order 4 of the Hermite PC expansion of the CCDF of $X$ are given in Appendix A and are shown in Figure 3.3 for the spectral acceleration.

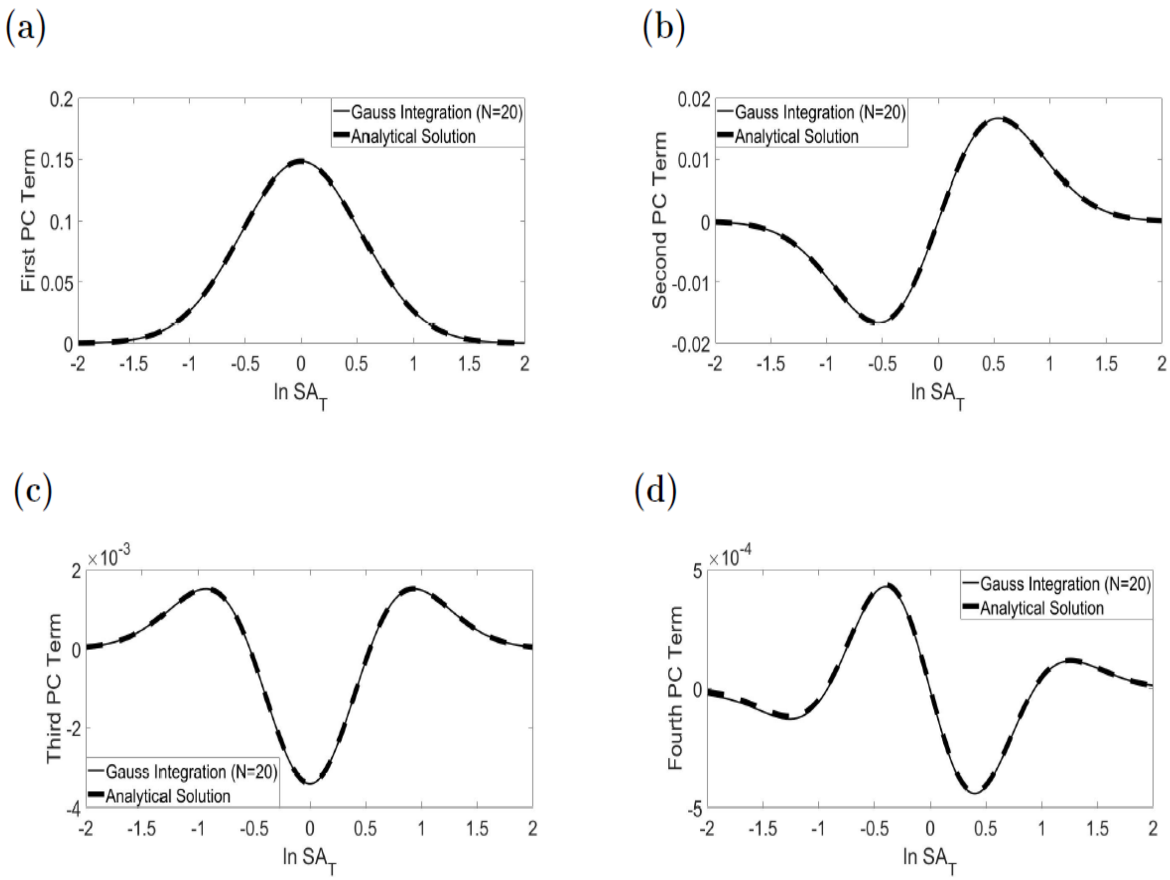


**Figure 3.3: Comparison of the first (a), second (b), third (c), and fourth (d) polynomial chaos terms of uncertain CCDF computed using the analytical solutions and by numerical integration. This example uses $m_\mu = 0$, $\sigma_\mu = 0.2$, and $\sigma = 0.5$, all in LN units.**

The fractiles of the PC approximation of a CCDF with epistemic uncertainty in the median are shown in Figure (3.4). As a reference, 10,000 realizations of the median were generated, and the corresponding hazard curves were computed using a Monte-Carlo approach, which we consider to be the true solution.

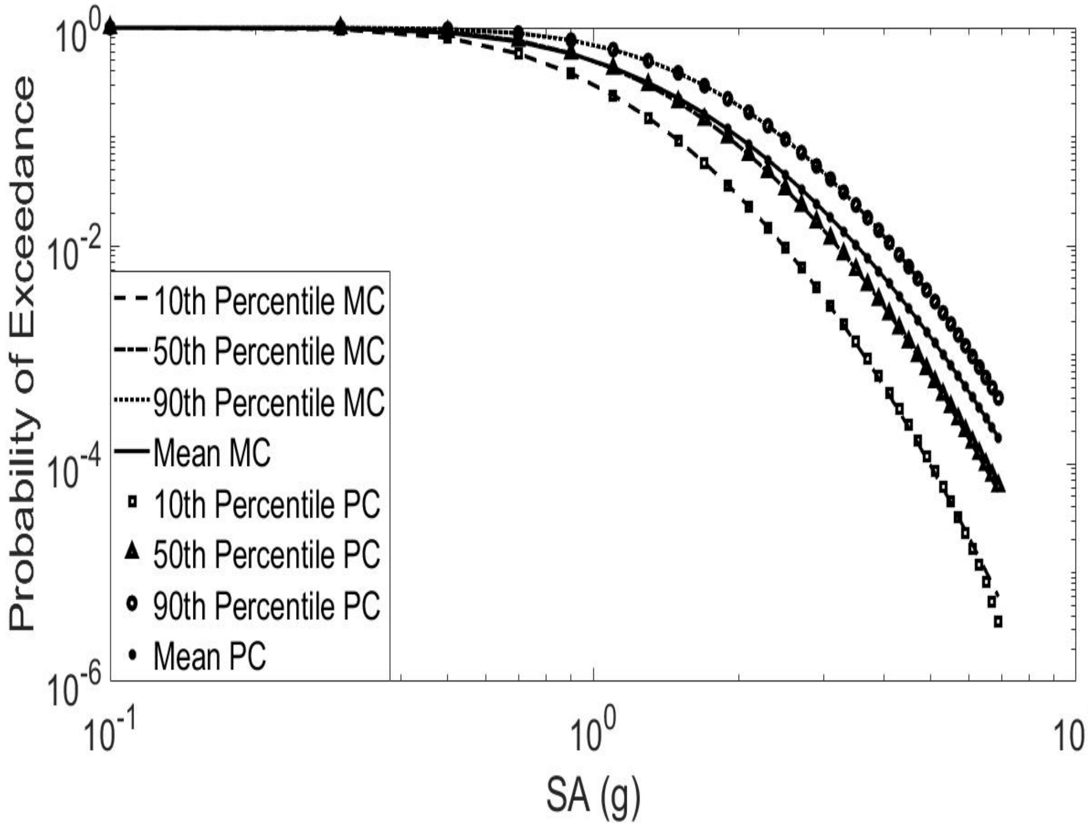


**Figure 3.4: Example of the CCDF for one scenario with uncertain median (**$m_\mu = 0.0, \sigma_\mu = 0.2, \sigma = 0.5$ **LN units. The epistemic fractiles are compared using Monte-Carlo (lines) and using the fourth-order polynomial chaos approximation (symbols)**

### 3.1.2 Polynomial Chaos Expansion of Probability Density Functions (PDF)

The formulation of the PBEE risk as in Equation (2.1) contains the CCDFs for the aleatory variability as well as the PDFs for the epistemic uncertainty of the $IM$, $EDP$, $DM$, and $DV$. To efficiently propagate the epistemic uncertainty of these median models into PBEE risk calculations, it is necessary to obtain a PC approximation for both the CCDF and the PDFs. The PC approximation of CCDF was described in the section 3.1.1. Here, we introduce the PC approximation of the PDFs for the epistemic uncertainty.

Similar to the case of the CCDF, we consider a log-normally distributed random variable X, with a median $\mu_X$ and a standard deviation $\sigma_X$ for the epistemic uncertainty in the median model $\mu_X$. We again make the assumption that this median is normally distributed, centered on the mean median $m_{\mu X}$ with standard deviation $\sigma_{\mu_X}$.

The probability density function of the log-normally distributed random variable $X$ with epistemic uncertainty in the median $\mu_X$ as defined above can be written as:

$$f(x) = \frac{1}{x\sigma_X\sqrt{2\pi}} exp\left(\frac{-(ln(x) - (m_{\mu_X} + \sigma_{\mu_X}\xi))^2}{2\sigma_X^2}\right) \tag{3.5}$$

The PC approximation of the PDF from Equation (3.5) is given by:

$$f(x) \approx \sum_{i=0}^{P} c_i(x)\Psi_i[\{\xi\}] \tag{3.6}$$

In Equation (3.6), the terms $\Psi_i[\xi]$ are the Hermite polynomials of the standard normal random variable $\xi$, and the PDF of $X$ can be approximated by a linear combination of these polynomials, for which the coefficients $c_i(x)$ are given by the following projection:

$$\begin{aligned}\langle f(x) \cdot \Psi_i[\{\xi\}]\rangle = &\int_{\xi=-\infty}^{+\infty} \frac{1}{x\sigma_X\sqrt{2\pi}} exp\left(-\frac{(\ln(x) - (m_{\mu_X} + \sigma_{\mu_X}\xi))^2}{2\sigma_X^2}\right) \\ &\times \Psi_i[\{\xi\}] \times \frac{1}{\sqrt{(2\pi)}} e^{-\xi^2/2} d\xi\end{aligned} \tag{3.7}$$

Solving the integrals in Equation (3.7), we obtain the coefficients of the Hermite PC approximation of a PDF with epistemic uncertainty in its median, as in Equation (3.5).

The coefficients up to order 4 of the Hermite PC expansion of the PDF of $X$ are given in Appendix (A) and are illustrated in Figure (3.5)

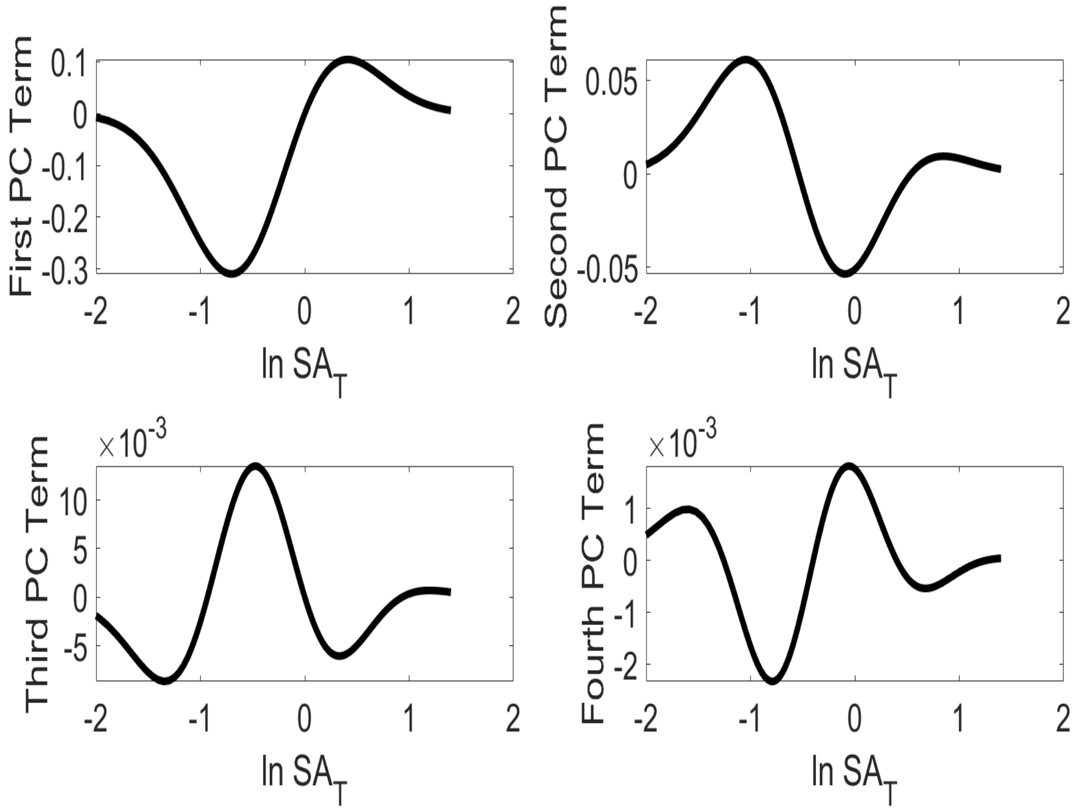


**Figure 3.5: First four polynomial chaos terms of a probability density function with epistemic uncertainty in the median. This example uses $m_\mu = 0, \sigma_\mu = 0.2, \sigma = 0.5$, all in LN units.**

The PC approximation of a PDF with epistemic uncertainty in the median is given in Figure (3.6).

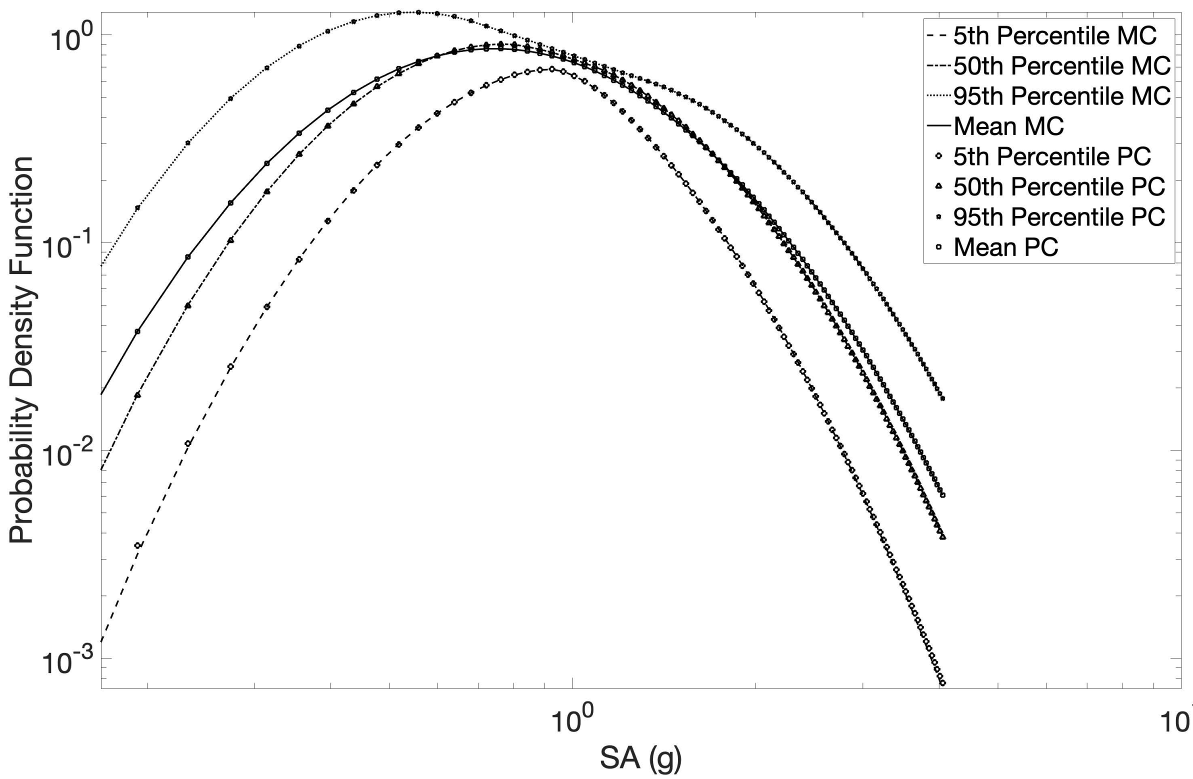


**Figure 3.6: Example of a probability density function with uncertain median ($m_\mu = 0.0, \sigma_\mu = 0.2, \sigma = 0.5$). The epistemic fractiles are compared using Monte-Carlo (MC) shown by the lines and using the fourth-order polynomial chaos (PC) approximation shown by the symbols)**

## 3.2 POLYNOMIAL CHAOS EXPANSION OF THE PBEE RISK EQUATION

The main Equation for PBEE risk calculation given in Equation (2.12) is for a single set of models and input parameters for $IM$, $EDP$. $DM$, and $DV$

The epistemic uncertainty in the median models for the $IM$, $EDP$, $DM$, and $DV$ can be easily modeled in a Monte-Carlo approach using a logic tree for each term. These logic trees can be assumed to be independent because the epistemic uncertainty in the median of one parameter should not depend on the epistemic uncertainty in the median of the other parameters. With the assumption of the four epistemic uncertainties being independent, one realization of the risk curve can be computed by picking one branch from each node of the logic tree. In the Monte-Carlo approach, the risk is computed for 1000s of samples of the branches of the logic tree.

Alternatively, these logic trees can be replaced by continuous distributions (Lacour and Abrahamson (2019)) with four independent random variables, $\xi_1$, $\xi_2$, $\xi_3$ and $\xi_4$, defined by:

$$
\begin{aligned}
\xi_1 &= \frac{\mu_{IM} - m_{\mu_{IM}}}{\sigma_{\mu_{IM}}} \\
\xi_2 &= \frac{\mu_{EDP} - m_{\mu_{EDP}}}{\sigma_{\mu_{EDP}}} \\
\xi_3 &= \frac{\mu_{DM} - m_{\mu_{DM}}}{\sigma_{\mu_{DM}}} \\
\xi_4 &= \frac{\mu_{DV} - m_{\mu_{DV}}}{\sigma_{\mu_{DV}}}
\end{aligned}
\tag{3.8}
$$

The epistemic uncertainty in the median $IM$ model is shown in Figure (3.7). This epistemic uncertainty comes from the alternative median ground motions from the NGA-West 2 ground-motion models (GMMs). Similarly, the epistemic uncertainty in the median $EDP$ with seismic-slope displacement from the model from Bray and Macedo (2019) is shown in Figure 3.8.

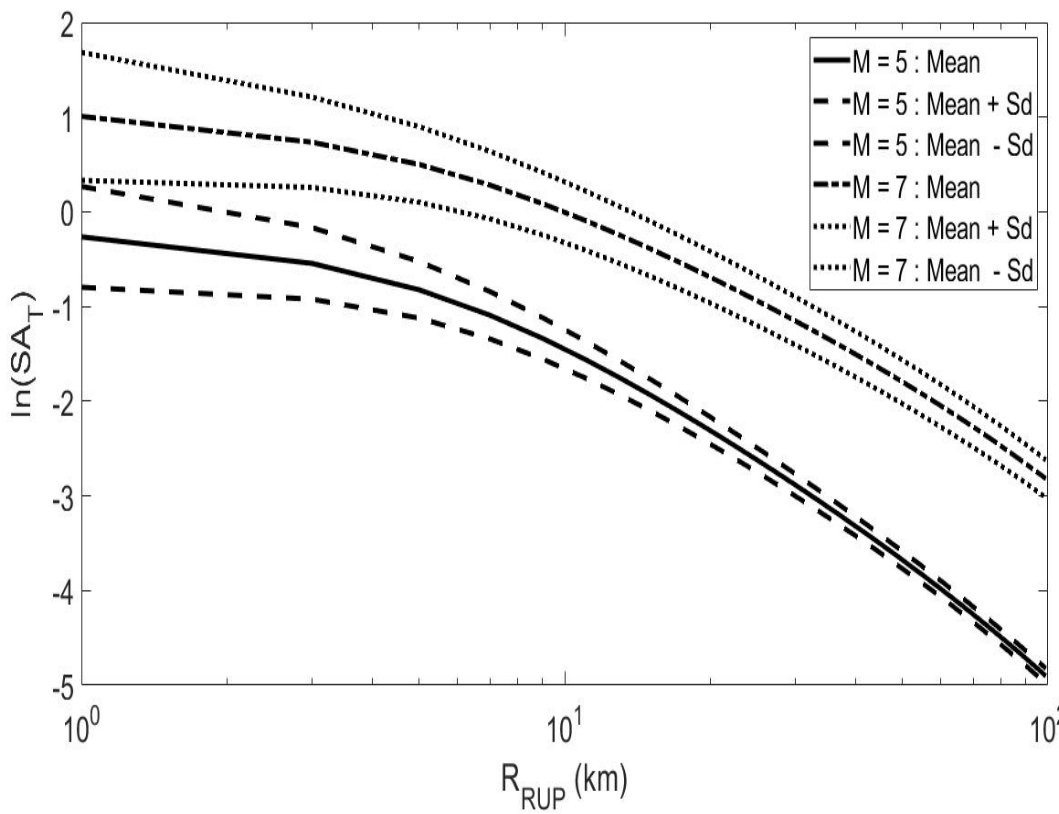


**Figure 3.7: Example of the epistemic uncertainty in the median ground motion for magnitude 5 and magnitude 7 earthquakes**

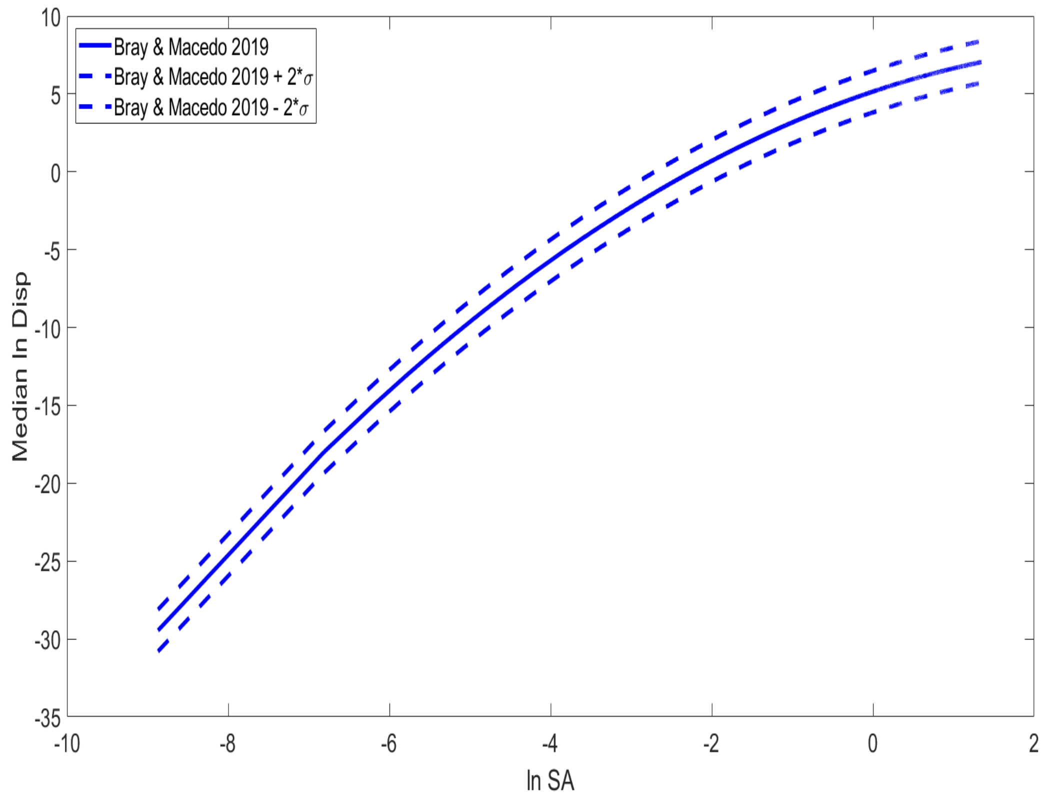


**Figure 3.8: Model from Bray and Macedo (2019) and the fractiles from its epistemic uncertainty ($\sigma_\mu = 0.7$)**

Adding the epistemic uncertainty in the median models of $IM$, $EDP$, $DM$ and $DV$ using the four independent standard normal variables $\{\xi_1, \xi_2, \xi_3, \xi_4\}$, the main risk Equation (2.12) becomes:

$$\lambda_{dv}(DV > dv) = \int_{\xi_4} \int_{\mathcal{D}_{DM}} \Phi \left( \frac{\ln(dv) - (m_{\mu_{DV}}(dm) + \sigma_{\mu_{DV}} \xi_4)}{\sigma_{DV}} \right) d(dm) d\xi_4 \tag{3.9}$$
$$\int_{\xi_3} \int_{\mathcal{D}_{EDP}} \frac{1}{\sqrt{2\,\pi} \sigma_{DM}\; dm} exp \left( - \frac{(\ln(dm) - (m_{\mu_{DM}}(edp) + \sigma_{\mu_{DM}} \xi_3))^2}{2\sigma_{DM}^2} \right) d(edp) d\xi_3$$
$$\int_{\xi_2} \int_{\mathcal{D}_{IM}} \frac{1}{\sqrt{2\,\pi} \sigma_{EDP}}\; exp \left( - \frac{(\ln(edp) - (m_{\mu_{EDP}}(im) + \sigma_{\mu_{EDP}} \xi_2))^2}{2\sigma_{EDP}^2} \right) d\xi_2$$
$$\int_{\xi_1} \frac{1}{\sqrt{2\,\pi} \sigma_{IM}\; im}\; exp \left( - \frac{(\ln(im) - (m_{\mu_{IM}} + \sigma_{\mu_{IM}} \xi_1))^2}{2\sigma_{IM}^2} \right) d(im) d\xi_1$$

Next, we can approximate each PDF and the CDF in the Equation (3.9) by their PC expansion from Equations (3.3) and (3.6). For one realization of $\xi_1$, $xi_2$, $\xi_3$, and $\xi_4$, the risk is given by:

$$\begin{aligned}\lambda_{dv}(DV > dv) = \int_{\mathcal{D}_{DM}} \sum_{i=0}^{P} C_i^{DV}(dm)\Psi_i[\xi_4]d(dm) \\ \int_{\mathcal{D}_{EDP}} \sum_{j=0}^{P} c_j^{DM}(edp)\Psi_j[\xi_3]\, d(edp) \\ \int_{\mathcal{D}_{IM}} \sum_{k=0}^{P} c_k^{EDP}(im) \\ \sum_{l=0}^{P} c_l^{IM}(im)\Psi_l[\xi_1]d(im)\end{aligned} \tag{3.10}$$

Reordering the sums and the integrals, we obtain:

$$\begin{aligned}\lambda_{dv}(DV > dv) = \sum_{i=0}^{P}\sum_{j=0}^{P}\sum_{k=0}^{P}\sum_{l=0}^{P} \int_{\mathcal{D}_{DM}} C_i^{DV}(dm)d(dm) \int_{\mathcal{D}_{EDP}} c_j^{DM}(edp)\, d(edp) \\ \int_{\mathcal{D}_{IM}} c_k^{EDP}(im)c_l^{IM}(im)d(im)\Psi_i[\xi_4]\Psi_j[\xi_3]\Psi_k[\xi_2]\Psi_l[\xi_1]\end{aligned} \tag{3.11}$$

The PC expansion of the main risk equation becomes:

$$\lambda_{dv}(DV > dv) = \sum_{i=0}^{P}\sum_{j=0}^{P}\sum_{k=0}^{P}\sum_{l=0}^{P} \lambda_{dv}^{i,j,k,l}\Psi_i[\xi_4]\Psi_j[\xi_3]\Psi_k[\xi_2]\Psi_l[\xi_1] \tag{3.12}$$

in which:

$$\begin{aligned}\lambda_{dv}^{i,j,k,l} = \int_{\mathcal{D}_{DM}} C_i^{DV}(dm)d(dm) \int_{\mathcal{D}_{EDP}} c_j^{DM}(edp)\, d(edp) \\ \int_{\mathcal{D}_{IM}} c_k^{EDP}(im)c_l^{IM}(im)d(im)\end{aligned} \tag{3.13}$$

The PC approximation of the main PBEE risk Equation (3.13) allows the propagation of the epistemic uncertainty in the median $IM$, $EDP$, $DM$, and $DV$ models in a much more efficient way than the Monte-Carlo sampling approach. The Monte-Carlo approach consists of sampling alternative branches of the logic trees $\xi_1, \xi_2, \xi_3, \xi_4$ from the median models and evaluating the main risk Equation (2.10) to obtain a statistical description of the epistemic uncertainty in the resulting risk. This approach requires hundreds or thousands of sample evaluations of the risk equation to provide a statistically robust estimation of the epistemic uncertainty in the risk. Alternatively, the PC approach only requires the calculation of the few PC terms in the approximation of the risk in Equation (3.13), and the sampling of the logic tree can be performed in an efficient post-processing phase where Hermite polynomials are evaluated outside of the risk calculation.

Similar to the deterministic case of Equation (2.12), the PC coefficients of the risk $\lambda_{dv}^{i,j,k,l}$ also involve three integrals and the PC coefficients $C_i(dm)$, $c_j(edp)$, $c_k(im)$, and $c_l(im)$ of the PC expansion of the CDF and the PDF of the random variables $IM$, $EDP$, $DM$, and $DV$. These integrals can be evaluated using the quadrature rule, but they can be much more efficiently evaluated if they have analytical solutions.

Because the PC coefficients inside the integrals include exponential terms for the different median models (the terms $\alpha$, as described in the Appendix (A)), the tangent approximation of the median conditioning models is also a necessary condition in order to have analytical solutions of the integrals in Equation (3.13), as it was the case for the deterministic PBEE risk calculation.

If the epistemic uncertainty in the conditioning models is assumed to be normally distributed, leading to constant shifts of the median model, the epistemic uncertainty of the tangent approximation remains a linear function of the conditioning parameters. Analytical solutions are available for the PC terms $\lambda_{dv}^{i,j,k,l}$. We provide the mathematical formulation of the analytical solutions of the PC terms of Equation (3.13) in the Appendix (A).

The PC approach combined with the tangent approximation for analytical solutions of the risk integrals provides an efficient tool for calculating risk and propagating the epistemic uncertainty in the median models into large-scale risk calculations.

## 3.3 ILLUSTRATIVE EXAMPLES

In this section, we illustrate the limited loss accuracy and the large gain of computational speed using the combined PC and tangent approximation approach compared to the direct Monte-Carlo sampling approach with numerical integration using the quadrature rule.

### 3.3.1 Single Integration of a Risk Curve with a Smooth Fragility Model

For the case of a single earthquake scenario and a single integral of the risk over $IM$, the risk as a function of $edp$ is given by:

$$
\begin{aligned}
\lambda_{EDP}(EDP > edp) &= \int_{\mathcal{D}_{IM}} P(EDP > edp|im) f(im) d(im) \\
&= \int_{\mathcal{D}_{IM}} \Phi\left(\frac{\ln(edp) - (m_{\mu_{EDP}}(im) + \sigma_{\mu_{EDP}}\xi_2)}{\sigma_{EDP}}\right) \times \\
&\frac{1}{\sqrt{2\,\pi}\sigma_{IM}\, im}\, exp\left(-\frac{(\ln(im) - (m_{\mu_{IM}} + \sigma_{\mu_{IM}}\xi_1))^2}{2\sigma_{IM}^2}\right) d(im)
\end{aligned} \tag{3.14}
$$

For the $IM$ parameter, we use the spectral acceleration and given by the five NGA-West 2 GMMs. The mean median is given by the average of the 5 different GMMs, and the corresponding epistemic uncertainty is given by the standard deviation of the median ground motions. The epistemic uncertainty is sampled using the standard normal random variable $\xi_1$.

For the $EDP$ parameter, we consider seismic slope displacement and use the model from Bray and Macedo (2019) as illustrated in Figure (3.9). The epistemic uncertainty of the Bray and Macedo (2019) model is assumed to be constant for all values of $\ln SA$ in the median model. The epistemic uncertainty in the median EDP model is sampled using the standard normal random variable $\xi_2$, which is assumed to be independent of $\xi_1$.

The risk curve and its tangent approximation are calculated for the case of a single earthquake scenario of magnitude $M = 5$ at a distance $R = 20km$ from a site. The mean median and standard deviation of the $ln(SA)$ for that scenario are $\ln \mu_{SA} = -5.04$ and $\sigma_{SA} = 0.78$ natural log units. The region of interest over which the tangent approximation is performed is taken to be +/- two standard deviations around the median value and is $\ln SA = [-6.60, -3.48]$. Over that range, the model from Bray and Macedo (2019) and its derivative are continuous, and it is reasonable to approximate it by a linear function of the $\ln SA$. The model from Bray and Macedo (2019) and its epistemic uncertainty along with its tangent approximation for that example earthquake scenario are shown in Figure (3.9).

The EDP hazard curve (assuming the rate of earthquakes = units) can be approximated by the tangent approximation:

$$
\begin{aligned}
\lambda_{EDP}(EDP > edp) &= \int_{\mathcal{D}_{IM}} P(EDP > edp|im) f(im) d(im) \\
&= \int_{\mathcal{D}_{IM}} \Phi(\frac{\ln(edp) - (a_{EDP}(im) + b_{EDP}(im) + \sigma_{\mu_{EDP}}\xi_2)}{\sigma_{EDP}}) \times \\
&\frac{1}{\sqrt{2\,\pi}\sigma_{IM}\, im}\, exp(-\frac{(\ln(im) - (m_{\mu_{IM}} + \sigma_{\mu_{IM}}\xi_1))^2}{2\sigma_{IM}^2}) d(im)
\end{aligned} \tag{3.15}
$$

in which $a_{EDP}(im)$ and $b_{EDP}(im)$ represent the slope and the intercept of the tangent approximation of the Bray and Macedo (2019) model at the median $IM$ value of $\ln \mu_{SA} = -5.04$.

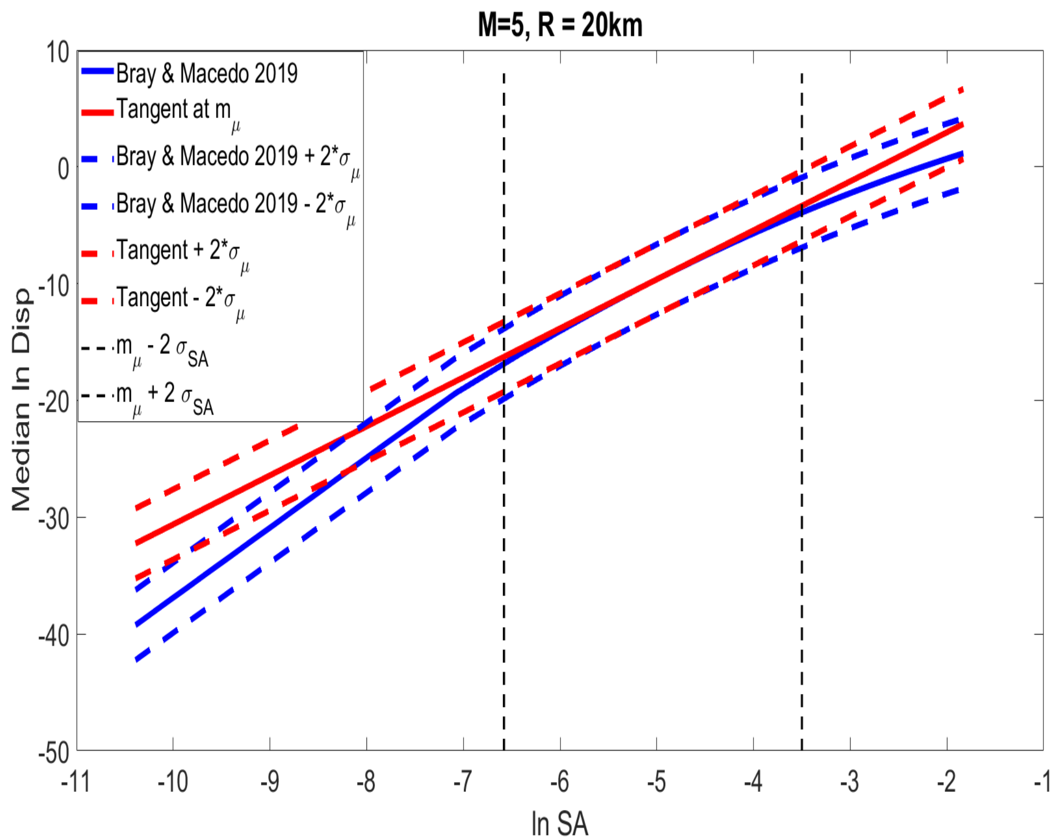


**Figure 3.9: Model from Bray and Macedo (2019) and the $+2\sigma$ fractiles from its epistemic uncertainty**

The tangent approximation of the Bray and Macedo (2019) model leads to analytical solutions of the PC terms $\lambda_{edp}^{i,j}$ in the main risk equation from Equation (3.13), which are given the Appendix A.

Next, we compute the risk curves and their epistemic uncertainty using both the traditional Monte-Carlo sampling approach and the PC method with the tangent approximation to propagate the epistemic uncertainty. The accuracy and computational time and the accuracy of the two methods are compared.

For the Monte Carlo method, 1000 samples of the median ground motion and the Bray and Macedo (2019) model are generated, and the corresponding sample risk curves are computed using 100 quadrature points.

For the PC-tangent method, the same 1000 median samples are propagated using PC order 4 for both the IM and EDP fragility models, leading to 15 PC terms, which are computed using their analytical solutions thanks to the tangent approximation. The sample risk curves are then generated during post-processing by multiplying the PC coefficients by the Hermite polynomials.

Figure (3.10) shows the fractiles of these risk curves. We observe that the two methods give similar results, with an increasing error in the fractiles as we approach the tails of the distribution. This error comes from both the PC approximation and the tangent approximation. The error coming from the PC expansion could be reduced by increasing the order of the PC expansion. The error from the tangent approximation could be reduced by performing the tangent approximation over multiple subdomains. In both cases, it is possible to increase the accuracy of the method but at the cost of more computational expenses.

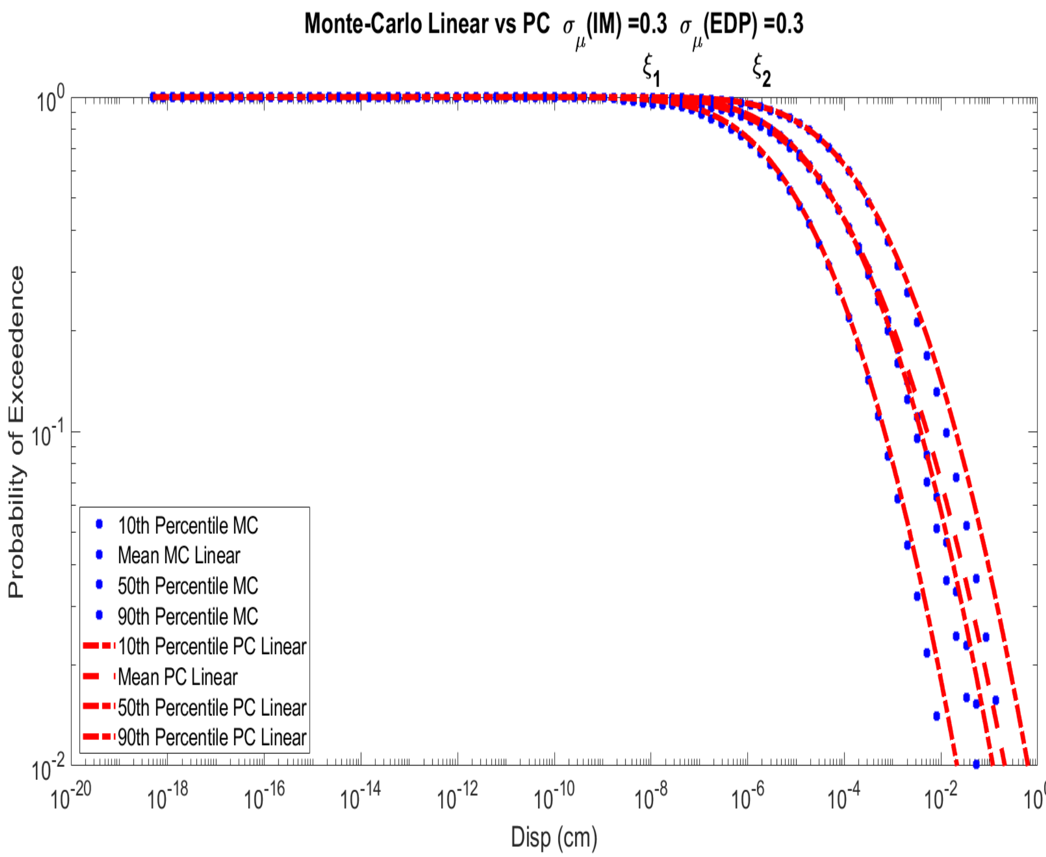


**Figure 3.10: Probability of exceeance curves and their epistemic fractiles from the Monte-Carlo and PC approaches**

The computational times of the quadrature-rule/Monte-Carlo approach and of the tangent-approximation/PC approach are shown in Figures (3.11) and (3.12). In Figure (3.11), the computational times of the two methods are plotted as a function of the number of medians sampled from $\xi_1$ and $\xi_2$. The computational time from the Monte-Carlo approach increases linearly with the number of branches; however, the computational time from the PC approach does not dependent on the number of branches in the logic tree because it uses analytical approximations for a continuous distribution. The sampling of the continuous distribution is conducted during post-processing, which is very fast.

In Figure (3.12), the computational times of the two methods are calculated for 1,000 hypothetical earthquake scenarios $M = 5, R = 20km$, to have an estimate of the total run time of a hypothetical hazard calculation. The computational times of the two methods are plotted as a function of the number of hypothetical sites at which we calculate the risk. The run times from the two methods scale linearly with the number of sites, and the PC method remains faster by about 4 orders of magnitude compared to the Monte-Carlo approach with 1000 sampled medians and 100 quadrature points.

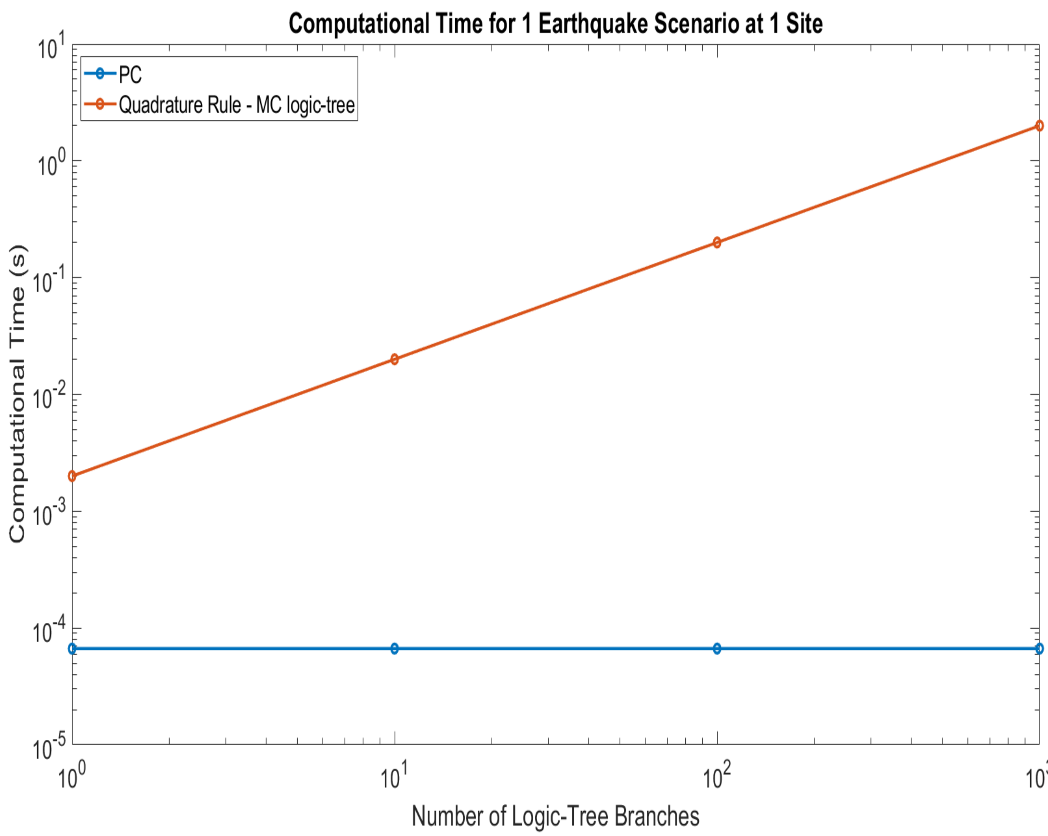


**Figure 3.11: Computational time of the risk curves from the Monte-Carlo and PC approaches for one earthquake scenario at one site as a function of the number of end branches from the logic-tree.**

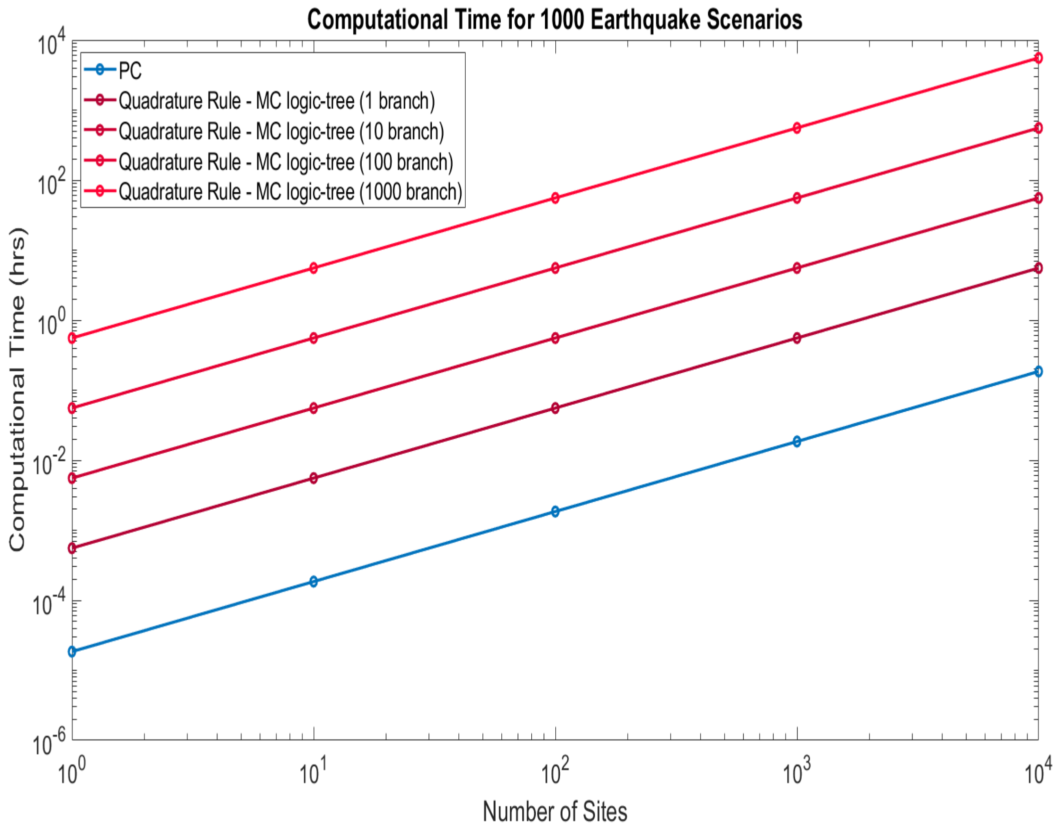


**Figure 3.12: Computational time of the risk curves from the Monte-Carlo and PC approaches for 1,000 earthquake scenarios as a function of the number of sites**

The PC-tangent method leads to a very efficient procedure to calculate risk and propagate the epistemic uncertainty in the different models for the $IM$, $EDP$, $DM$, and $DV$ parameters. For 1,000 sampled median models and 10,000 sites, the PC-tangent method takes about 10 minutes, while the direct approach would require months of calculations. This gain in computational efficiency justifies the reasonable error in the risk approximation coming from the tangent and the PC methods.

### 3.3.2 Triple Integration of a Risk Number with a Strongly Non-Linear Fragility Model

The previous example showed the risk calculation for smooth scaling in the models. For models with strong nonlinear behavior, the scaling may not be smooth. For example, three alternative models of the pipe strain from the Hutabarat et al. (2025) model are shown in Figure 3.13. There is a sharp change of slope in its median fragility model. This sharp change of slope makes the tangent approximation not accurate when the tangent is taken around that domain; however, when taking the average of multiple realizations of the model, the mean of the median DM does not have a sharp change of slope, as shown in Figure 3.14. This is because the multiple sharp changes of slopes average out over the domain of $EDP$. When averaging all these realizations, the mean median model becomes smooth, and it makes the tangent approximation reasonable. A key requirement for using the PC method with the tangent approximation is to account for the epistemic uncertainty in the input parameters of the fragility model and to take its average so that the mean fragility model is smooth and can be accurately approximated with the tangent approximation.

The result from the example risk calculation following Equation (3.9) is the probability of pipe rupture, which is a single number. Because there is epistemic uncertainty in the fragility models and in their input parameters, the resulting probability of pipe rupture also presents epistemic uncertainty. The probability of rupture and its epistemic uncertainty are displayed in the histogram from Figure (3.15). The results from the direct Monte-Carlo approach with numerical integration and the PC approach with analytical integration give similar results. The accuracy of the histogram could be improved by increasing the number of PC terms and the sub-domains over which the tangent approximation is performed, but in general, a PC order of 4 and a single tangent lead to acceptable accuracy. The computational time for the two approaches scales similarly to the previous case of a risk curve; however, because we now only calculate a number and not a curve, all the computational times obtained for Equation (3.9) are faster by one order of magnitude compared to the case of Equation (2.1).

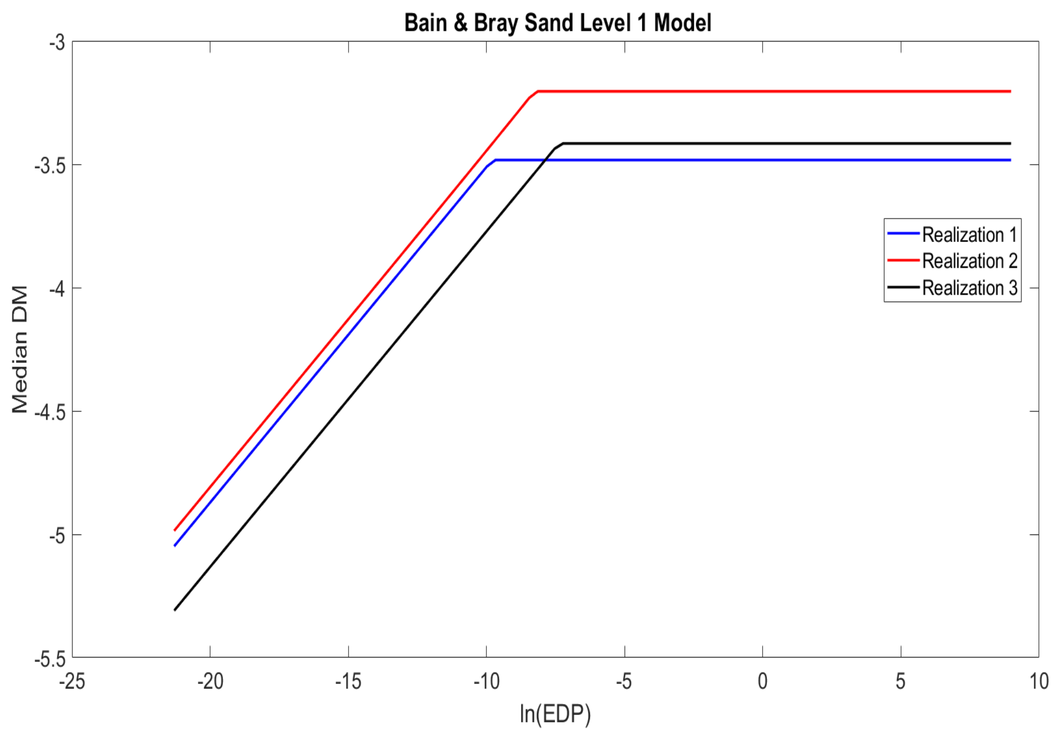


**Figure 3.13: Three Realizations of the Hutabarat et al. (2025) Model**

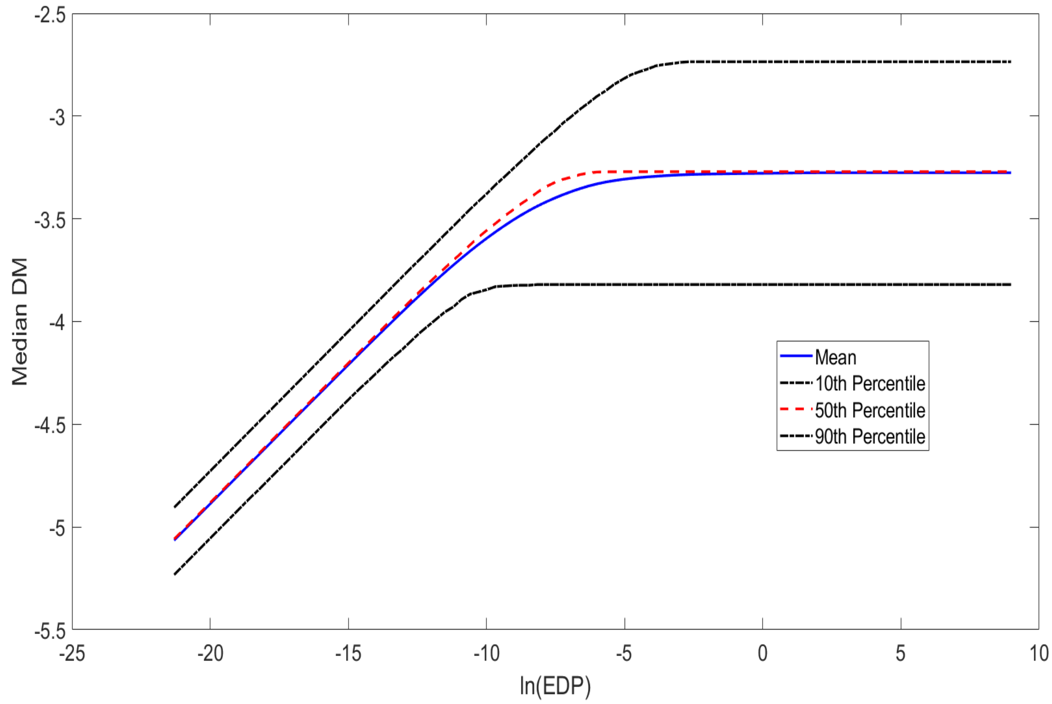


**Figure 3.14: Fractiles of the Median Hutabarat et al. (2025) Model ($m_\mu$, $\sigma_\mu$)**

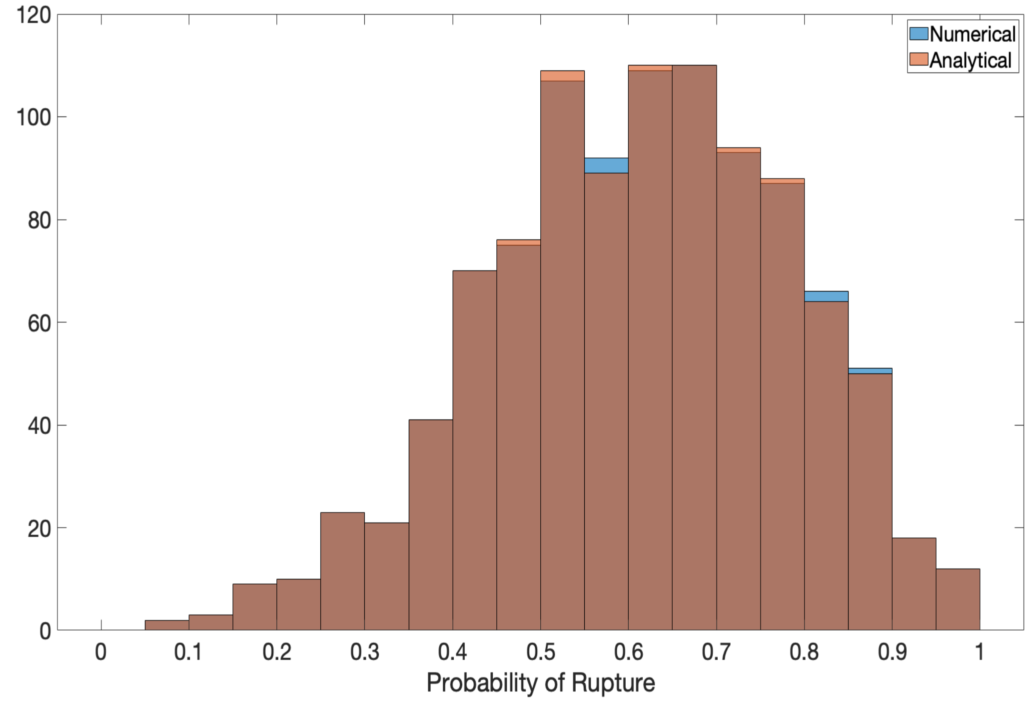


**Figure 3.15: Histogram of the probability of pipe rupture from the Monte-Carlo and PC approaches**

# 4 Conclusion

The new method based on Polynomial Choas with a tangent approximation for the scaling of the EDP, DM, and DV is a numerically efficient method for computing the risk based on the PEER risk framework. The PC method efficiently propagates the epistemic uncertainty for the EDP, DM, and DV models into the risk calculation by approximating the multiple risk curves from a logic tree with a set of functions given by the Hermite polynomials. The linear approximation of the EDP and models makes the risk calculation from complicated geotechnical models practical.

The method can be applied to large-scale, state-wide risk calculation. The increase in computational speed compared to the traditional approach (brute-force Monte-Carlo sampling with numerical integration) can be 4 orders of magnitude, which justifies the relatively small error from the approximations in the new approach. The very short calculation time makes the geotechnical engineering design and decision-making involved in the development of the gas pipeline network practical; however, the scope of applications of the developed method is not limited to geotechnical engineering and can be broadened to other fields, including other branches of risk calculation in civil engineering as well as in the insurance industry.

The Matlab code of the presented method for efficient PBEE risk calculation can be found at: `https://github.com/maxlacour/PC_Risk_Calculation`

# REFERENCES

Bray, J. D. and Macedo, J. (2019). "Procedure for estimating shear-induced seismic slope displacement for shallow crustal earthquakes." *Journal of Geotechnical and Geoenvironmental Engineering*, 145(12), 04019106.

Ghanem, R. and Spanos, P. (1991). *Stochastic Finite Elements: A Spectral Approach*. Springer-Verlag, <https://books.google.com/books?id=N3GwQgAACAAJ>.

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

Hutabarat, D., O'Rourke, T.D., Bray, J., Bain, C., and Linavall, S. (2025). *Fragility Analysis of Buried Pipelines Derived from Finite-Element Soil-Pipeline Interaction Analysis. (OpenSRA Project).* PEER Report No. 2025-11, Pacific Earthquake Engineering Center, University of California, Berkeley, CA. https://doi.org/10.55461/BDOO7217

Lacour, M. and Abrahamson, N. A. (2019). "Efficient Propagation of Epistemic Uncertainty in the Median Ground-Motion Model in Probabilistic Hazard Calculations." *Bulletin of the Seismological Society of America*, 109(5), 2063–2072.

Macedo, J., Candia, G., Lacour, M., and Liu, C. (2020). "New developments for the performance-based assessment of seismically-induced slope displacements." *Engineering Geology*, 277, 105786.

# Appendix A: Analytical Solution of Risk Integration of Polynomial Chaos Expansion of Probability Distributions and Density Functions

## A.1 POLYNOMIAL CHAOS COEFFICIENTS OF CDF

The complementary probability distribution function of a log-normally distributed random variable $X$ with mean median $m_X$, epistemic uncertainty $\sigma_{\mu_X}$ and standard deviation $\sigma_X$ is given:

$$P(X > x) = 1 - \Phi\left(\frac{\ln(x) - (m_{\mu_X} + \sigma_{\mu_X}\xi)}{\sigma_X}\right) \tag{A.1}$$

The Polynomial Chaos expansion of this probability distribution function is given by:

$$P(X > x) \approx \sum_{i=0}^{P} C_i(x)\Psi_i[\{\xi\}] \tag{A.2}$$

in which the coefficients $C_i$ are given by:

$$\begin{aligned}
C_0(x) =& \left[1 - \Phi\left(\frac{\ln(x) - m_{\mu_X}}{\sqrt{\sigma_X^2 + \sigma_{\mu_X}^2}}\right)\right] \\
C_1(x) =& \alpha_C \times \frac{1}{(-a(x))^{1/2}} \\
C_2(x) =& \frac{\alpha_C}{2} \times \frac{b(x)}{2(-a(x))^{3/2}} \\
C_3(x) =& \frac{\alpha_C}{6} \times \frac{-2a(x)(1 + 2a(x)) + b(x)^2}{4(-a(x))^{5/2}} \\
C_4(x) =& \frac{\alpha_C}{24} \times \frac{-b(x)(6a(x)(1 + 2a(x)) - b(x)^2)}{8(-a(x))^{7/2}}
\end{aligned} \tag{A.3}$$

and where:

$$\alpha_C = \frac{\sigma_{\mu_X}}{2\sigma_X\sqrt{\pi}} \times exp\left(c(x) - b(x)^2/(4a(x))\right) \tag{A.4}$$

and

$$\begin{aligned} a(x) &= -\frac{\sigma_{\mu_X}^2}{2\sigma_X^2} - \frac{1}{2} \\ b(x) &= (\ln(x) - m_{\mu_X})\frac{\sigma_{\mu_X}}{\sigma_X^2} \\ c(x) &= -(\ln(x) - m_{\mu_X})^2\frac{1}{2\sigma_X^2} \end{aligned} \tag{A.5}$$

## A.2 POLYNOMIAL CHAOS COEFFICIENTS OF PDF

The probability density function of a log-normally distributed random variable $X$ with mean median $m_X$, epistemic uncertainty $\sigma_{\mu_X}$ and standard deviation $\sigma_X$ is given:

$$f(x) = \frac{1}{x\sigma_X\sqrt{2\pi}} e^{-\dfrac{(\ln(x) - (m_{\mu_X} + \sigma_{\mu_X}\xi))^2}{2\sigma_X^2}} \tag{A.6}$$

The Polynomial Chaos expansion of this probability density function is given by:

$$f_X(x) \approx \sum_{i=0}^{P} c_i(x)\Psi_i[\{\xi\}] \tag{A.7}$$

in which the coefficients $c_i$ are given by:

$$\begin{aligned} c_0(x) =& \alpha_P \times \frac{1}{(-a(x))^{1/2}} \\ c_1(x) =& \frac{\alpha_P}{1} \times \frac{b(x)}{2(-a(x))^{3/2}} \\ c_2(x) =& \frac{\alpha_P}{2} \times \frac{-2a(x)(1+2a(x)) + b(x)^2}{4(-a(x))^{5/2}} \\ c_3(x) =& \frac{\alpha_P}{6} \times \frac{-b(x)(6a(x)(1+2a(x)) - b(x)^2)}{8(-a(x))^{7/2}} \\ c_4(x) =& \frac{\alpha_P}{24} \times \frac{12a^2(x)(1+2a(x))^2 - 12a(x)(1+2a(x))b(x)^2 + b(x)^4}{16(-a(x))^{9/2}} \end{aligned} \tag{A.8}$$

$$\alpha_P = \frac{1}{2x\,\sigma_X\sqrt{\pi}} \times e^{c(x)-b(x)^2/(4a(x))} \tag{A.9}$$

and

$$\begin{aligned} a(x) &= -\frac{\sigma_{\mu_X}^2}{2\sigma_X^2} - \frac{1}{2} \\ b(x) &= (\ln(x) - m_{\mu_X})\frac{\sigma_{\mu_X}}{\sigma_X^2} \\ c(x) &= -(\ln(x) - m_{\mu_X})^2\frac{1}{2\sigma_X^2} \end{aligned} \tag{A.10}$$

## A.3 ANALYTICAL INTEGRATION OF POLYNOMIAL CHAOS EXPANSIONS FOR RISK CALCULATIONS

Consider two random variables $X$ and $Y$, log-normally distributed, with mean median $m_X$, epistemic uncertainty $\sigma_{\mu_X}$, standard deviation $\sigma_X$ for variable $X$, and similarly, mean median $m_Y$, epistemic uncertainty $\sigma_{\mu_Y}$ and standard deviation $\sigma_Y$ for variable $Y$. The risk equation from the PEER framework becomes:

$$\begin{aligned} \lambda_X &= \int_{\mathcal{D}_Y} P(X > x|y) f(y) dy \\ &= \int_{\mathcal{D}_Y} \Phi(\frac{\ln(x) - (m_{\mu_X}(y) + \sigma_{\mu_X}\xi_2)}{\sigma_X}) \times \\ &\frac{1}{\sqrt{2\,\pi}\sigma_Y\; y}\; exp\left(-\frac{(\ln(y) - (m_{\mu_Y} + \sigma_{\mu_Y}\xi_1))^2}{2\sigma_Y^2}\right) dy \end{aligned} \tag{A.11}$$

The complementary cumulative distribution function of $X$ can be approximated by its Polynomial Expansion:

$$P(X > x) \approx \sum_{i=0}^{P} C_i(x)\Psi_i[\{\xi_1\}] \tag{A.12}$$

Similarly, the probability density function of $Y$ can be approximated by its PC expansion:

$$f_Y(y) \approx \sum_{i=0}^{P} c_i(y)\Psi_i[\{\xi_2\}] \tag{A.13}$$

Replacing these Polynomial Chaos expansions in Equation (A.11), we obtain:

$$\lambda_X = \int_{\mathcal{D}_Y} \sum_{i=0}^{P} \sum_{j=0}^{P} C_i(x) c_j(y) \Psi_i[\{\xi_1\}] \Psi_j[\{\xi_2\}] dy \tag{A.14}$$

Because the integration is performed over the domain $\mathcal{D}_Y$, the Hermite polynomials can be factorized outside of the integral, which simplifies into:

$$\lambda_X = \left( \int_{\mathcal{D}_Y} \sum_{i=0}^{P} \sum_{j=0}^{P} C_i(x) c_j(y) dy \right) \Psi_i[\{\xi_1\}] \Psi_j[\{\xi_2\}] \tag{A.15}$$

$$= \sum_{i=0}^{P} \sum_{j=0}^{P} I_{i,j} \Psi_i[\{\xi_1\}] \Psi_j[\{\xi_2\}] \tag{A.16}$$

We assume that the conditional median model $m_{\mu_X}(y)$ is a linear function of $\ln(y)$. Under that assumption and by replacing the coefficients $C_i$ and $c_j$ by their analytical expressions, the integral $I_{i,j}$ takes the form:

$$I_{i,j} = \int_{\mathcal{D}_Y} P_{i,j}(\ln(y)) \; e^{a_{i,j}(\ln(y))^2 + b_{i,j}\ln(y) + c_{i,j}} \; dy \tag{A.17}$$

in which $P_{i,j}(\ln(y))$ is a polynomial in $\ln(y)$ for which the degree will depend on the indices $i$ and $j$ in the double summation, and where $a_{i,j}, b_{i,j}, c_{i,j}$ are constants. The integral $I_{i,j}$ in the previous Equation presents an analytical solution because it is composed of the product of a polynomial in $\ln(y)$ times the exponential of a degree two polynomial in $\ln(y)$.

The analytical solution of the integral $I_{i,j}$ for polynomials $P_{i,j}$ of degrees up to 4 are given by:

$$\int_{\mathcal{D}_Y} (p_0) \; e^{a_{i,j}(\ln(y))^2 + b_{i,j}\ln(y) + c_{i,j}} \; dy = \alpha_I \; \times \frac{1}{(-a_{i,j})^{1/2}} \, p_0 \tag{A.18}$$

$$\int_{\mathcal{D}_Y} (p_0 + p_1 \ln(y)) \; e^{a_{i,j}(\ln(y))^2 + b_{i,j}\ln(y) + c_{i,j}} \; dy = \alpha_I \; \times \frac{-2a_{i,j}p_0 + b_{i,j}p_1}{(-2a_{i,j})^{3/2}} \tag{A.19}$$

$$\begin{aligned} &\int_{\mathcal{D}_Y} (p_0 + p_1 \ln(y) + p_2(\ln(y))^2) \; e^{a_{i,j}(\ln(y))^2 + b_{i,j}\ln(y) + c_{i,j}} \; dy \\ &\qquad = \alpha_I \; \times \frac{4a_{i,j}^2 p_0 + b_{i,j}^2 p_2 - 2a_{i,j}(b_{i,j}p_1 + p_2)}{(-4a_{i,j})^{5/2}} \end{aligned} \tag{A.20}$$

$$\int_{\mathcal{D}_Y} (p_0 + p_1 \ln(y) + p_2(\ln(y))^2 + p_3(\ln(y))^3) \; e^{a_{i,j}(\ln(y))^2 + b_{i,j}\ln(y) + c_{i,j}} \; dy = \tag{A.21}$$
$$\alpha_I \; \times \frac{-8a_{i,j}^3 p_0 + 4a_{i,j}^2(b_{i,j}p_1 + p_2) + b_{i,j}^3 p_3 - 2a_{i,j}b_{i,j}(b_{i,j}p_2 + 3p_3)}{(-8a_{i,j})^{7/2}}$$

$$\int_{\mathcal{D}_Y} (p_0 + p_1 \ln(y) + p_2(\ln(y))^2 + p_3(\ln(y))^3 + p_4(\ln(y))^4) \; e^{a_{i,j}(\ln(y))^2 + b_{i,j}\ln(y) + c_{i,j}} \; dy = \tag{A.22}$$
$$\alpha_I \; \times \frac{16a_{i,j}^4 p_0 - 8a_{i,j}^3(b_{i,j}p_1 + p_2) + b_{i,j}^4 p_4}{(-16a_{i,j})^{9/2}} \tag{A.23}$$
$$+ \, \alpha_I \; \times \frac{4a_{i,j}^2(b_{i,j}^2 p_2 + 3b_{i,j}p_3 + 3p_4) - 2a_{i,j}b_{i,j}^2(b_{i,j}p_3 + 6p_4))}{(-16a_{i,j})^{9/2}}$$

in which:

$$\alpha_I = \sqrt{\pi} e^{-c_{i,j} - b_{i,j}^2/4a_{i,j}} \tag{A.24}$$

The same procedure applies in an iterative manner to solve for the double and triple integrals in the main risk equation (2.12).

The Pacific Earthquake Engineering Research Center (PEER) is a multi-institutional research and education center with headquarters at the University of California, Berkeley. Investigators from over 20 universities, several consulting companies, and researchers at various state and federal government agencies contribute to research programs focused on performance-based earthquake engineering.

These research programs aim to identify and reduce the risks from major earthquakes to life safety and to the economy by including research in a wide variety of disciplines including structural and geotechnical engineering, geology/seismology, lifelines, transportation, architecture, economics, risk management, and public policy.

PEER is supported by federal, state, local, and regional agencies, together with industry partners.

**PEER Core Institutions**

University of California, Berkeley (Lead Institution)
California Institute of Technology
Oregon State University
Stanford University
University of California, Davis
University of California, Irvine
University of California, Los Angeles
University of California, San Diego
University of Nevada, Reno
University of Southern California
University of Washington

Pacific Earthquake Engineering Research Center
University of California, Berkeley
325 Davis Hall, Mail Code 1792
Berkeley, CA 94720-1792
Tel: 510-642-3437
Email: peer_center@berkeley.edu

ISSN 2770-8314
https://doi.org/10.55461/MFGK6090