



# Abstract

Geothermal energy is a renewable resource that currently draws significant attention in Germany. However, its side effect, geothermally induced earthquakes, must be considered when planning a power plant near residential areas. In the thesis, two questions are answered. First, will those earthquakes damage the building or make residents uncomfortable? Second, what is the predominant seismic parameter causing the negative effect on the structures and residents?

The following approaches are proposed to answer these two questions. First, a scenario-based ground motion generator is introduced that models the seismic event in the Insheim area, Germany. This generator can generate representative ground motion with stochasticity using seismic parameters such as moment magnitude and rupture distance. Once the generator is built, the responses of the structure excited by the ground motion are simulated and evaluated. Second, when performing serviceability and comfortability analysis, a large number of simulations are usually required, which will be highly time-consuming. To accelerate the assessment, a surrogate model, Stochastic Polynomial Chaos Expansion, is utilized. This surrogate model can efficiently re-interpret the relationship between inputted seismic parameters and outputted response, including the system's stochasticity.

Finally, the significance of each input seismic parameter will be analyzed using global sensitivity analysis, and the potential damage and adverse effects on residents will be evaluated using serviceability and comfortability analysis.

Keywords:

- Seismic risk assessment
- Geothermal-induced ground motion
- Scenario-based ground motion
- Probability-based surrogate model

- Stochastic Polynomial Chaos Expansion
- Global sensitivity analysis
- Serviceability and comfortability analysis
- Fragility curve

# Erklärungen

## Danksagung

An erster Stelle möchte ich Dr. Aditi Kumawat und M.Sc. Felix Schneider für die Betreuung meiner Arbeit danken. Ihre Tür stand bei Fragen oder Problemen immer offen. Wenn ich nicht weiterkam, habe sie sich die Zeit genommen, mit mir über eine Lösung zu diskutieren. Ebenfalls danke ich meiner Familie, meinen Freunden, Eva Huang und Ya-Wei Xiao für die moralische Unterstützung. Dabei gilt mein Dank insbesondere Dr. Aditi Kumawat und Chan-Hua Yeh für die kritische Durchsicht des Manuskripts.

## Erklärung

Hiermit versichere ich, die vorliegende Arbeit selbstständig und ohne fremde Hilfe angefertigt zu haben. Die verwendete Literatur und sonstige Hilfsmittel sind vollständig angegeben.

München, August 13, 2024

Wei-Teng Kao



# Contents

<b>Abstract</b>	<b>III</b>
<b>Erklärungen</b>	<b>V</b>
<b>List of Figures</b>	<b>X</b>
<b>Nomenclature</b>	<b>XV</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 State of research . . . . .	2
1.3 Outline . . . . .	4
<b>2 Modeling of induced ground motion</b>	<b>5</b>
2.1 General Procedure . . . . .	5
2.2 Assumptions for geothermally induced ground motion modeling . . . . .	7
2.3 Generation of artificial ground motion . . . . .	8
2.3.1 Kanai-Tajimi model . . . . .	8
2.3.2 Hu-Zhou model . . . . .	11
2.3.3 Time modulating function . . . . .	13
2.3.4 Artificial ground motion generator . . . . .	15
2.3.5 Fitting the model parameter from the database . . . . .	16
2.4 Relationship of seismic parameters and model parameters . . . . .	17
2.4.1 Linear mixed-effects model . . . . .	17
2.4.2 Fitting LME model and estimating the variance of error . . . . .	18
2.5 Baseline correction . . . . .	19
2.6 Model validation through response spectral analysis . . . . .	20
<b>3 Surrogate model</b>	<b>22</b>
3.1 Introduction of possible surrogate model . . . . .	22
3.2 Stochastic polynomial chaos expansion . . . . .	24
3.2.1 Polynomial chaos expansions . . . . .	24
3.2.2 Extension of PCE for stochastic simulators . . . . .	32
3.3 Validation and error measurement . . . . .	37
3.4 Global sensitivity analysis . . . . .	38
<b>4 Serviceability and comfortability analysis</b>	<b>39</b>
4.1 DIN standards 4150-2 . . . . .	39

4.2	DIN standards 4150-3 . . . . .	41
4.3	Fragility curve . . . . .	42
<b>5</b>	<b>Overall framework of seismic assessment</b>	<b>43</b>
5.1	Extension from existing framework . . . . .	43
5.2	Flowchart of the proposed framework . . . . .	44
<b>6</b>	<b>Numerical example</b>	<b>47</b>
6.1	Scenario-based ground motion generator . . . . .	47
6.1.1	Analysis of recorded ground motion . . . . .	47
6.1.2	Fitted LME results of seismic parameters . . . . .	53
6.1.3	Response spectral analysis - SBGMG Validation . . . . .	59
6.2	Numerical example: SDOF system . . . . .	60
6.2.1	SBGMG input parameters . . . . .	60
6.2.2	Training result and verification . . . . .	64
6.2.3	Convergence study of error . . . . .	70
6.2.4	Runtime comparison of SPCE . . . . .	71
<b>7</b>	<b>Case study: Geothermal energy site in Taufkirchen</b>	<b>76</b>
7.1	Building description . . . . .	76
7.1.1	Building parameters . . . . .	76
7.1.2	Soil-structure interaction . . . . .	78
7.2	Simulation and training result . . . . .	79
7.3	Sensitivity analysis . . . . .	85
7.4	Serviceability and comfortability assessment . . . . .	87
<b>8</b>	<b>Conclusion and Outlook</b>	<b>92</b>
8.1	Conclusion . . . . .	92
8.2	Outlook . . . . .	93
<b>A</b>	<b>Appendix</b>	<b>95</b>
A.1	Software . . . . .	95
A.1.1	MATLAB . . . . .	95
A.1.2	PYTHON . . . . .	95
A.1.3	ANSYS® . . . . .	95
A.2	Example flowchart of artificial ground motion generator with fitting procedure	96
A.3	Stochastic polynomial chaos expansion framework . . . . .	100
A.4	Overall flowchart of simulation and training . . . . .	101
A.5	Additional plots of numerical example . . . . .	102
A.5.1	Histogram of model parameters . . . . .	102
A.5.2	Quantile plots of model parameters . . . . .	103
A.5.3	Fitting result of Linear mixed effect model . . . . .	103
A.5.4	Scatter plots of residuals of fitting linear mixed effect model . . . . .	108
A.5.5	Two further examples of response spectral analysis . . . . .	111
A.6	Additional plots of case study . . . . .	113
A.6.1	Different perspectives of Building model . . . . .	113

---

A.6.2	Response surfaces of the trained SPCE model . . . . .	115
A.7	Database of recorded ground motion . . . . .	117
<b>Bibliography</b>		<b>124</b>



## List of Figures

2.1	General procedure of SBGMG, including the LME, AGMG, and Baseline correction packages. $X_{sb}$ are the seismic parameters and $X_a$ are the modal parameters. AGMG is formed by the Hu model and time-modulating function. After generation, ground motion's acceleration, velocity, and displacement will be exported simultaneously. . . . .	6
2.2	Schematic diagram of K-T model, according to [Chen et al 2022]. . . . .	10
2.3	The FRF of the Hu model, formulated using equation (2.16). . . . .	13
2.4	The time-modulating function, formulated using equation (2.19). . . . .	15
3.1	Sketch of the different truncation sets, modified from [Zhu and Sudret 2023].	28
5.1	Overall framework used for this thesis, extend and modify from the existing framework in [Zhu et al 2023]. . . . .	46
6.1	Collection of 240 recorded ground motions in the $x$ -direction from the database in the frequency domain. One example of ground motion occurred at 21:47:07 on 9 <sup>th</sup> November 2020, recorded by INS3 station in $x$ -direction, and is presented as a blue line. . . . .	48
6.2	Collection of recorded ground motions from the database in the frequency domain. The yellow lines represent 240 ground motions used for identifying the model parameters. The orange line shows the average of ground motions, and the blue line represents the smoothing version of the average one, which is calculated by the moving average function with a window size of 250. . . .	49
6.3	Histogram of $M_L$ , $D$ and $R$ . The frequency in the histogram is normalized by the relative probability and the height of bin value $v_i$ is calculated as $v_i = \frac{c_i}{N}$ , where $c_i$ is the number of elements in the bin, and $N$ is the total number of input data, $N = 240$ . . . . .	50
6.4	Histograms of model parameters $\omega_g/2\pi$ , $\zeta_g$ , $r_h$ , $PGA$ in $z$ -direction. The frequency in the histogram is normalized by the relative probability, where the $PGA$ is presented on the logarithm scale. . . . .	52
6.5	Quantile plots of the model parameter in $z$ -direction. Hollow circles indicate the first and the third quartiles. . . . .	53
6.6	Scatter plots of residuals of $\ln(PGA)$ against $M_L$ and $R$ in $z$ -direction . . . .	55
6.7	Scatter plots of residuals of $\omega_f$ against $\omega_g$ and $R$ in $z$ -direction . . . . .	56
6.8	Scatter plots of residuals of $\zeta_g$ against $\omega_g$ and $R$ in $z$ -direction . . . . .	56

6.9	Response spectral analysis. The top figure shows the velocity response, and the bottom shows the acceleration response. The blue line is the spectral response of the ground motion, recorded at 20:40:15 on November 3 <sup>rd</sup> , 2021, from the seismic station INS8, Insheim, Germany. The orange lines show the 50 realizations of spectral response from SBGMG, where the seismic parameters $[M_L, R, \omega_g, \zeta_g] = [0.9, 12.64, 59.12, 0.32]$ , where $M_L, R$ are provided by the station and $\omega_g, \zeta_g$ are obtained by the fitting procedure. . . . .	60
6.10	Histogram of rupture distance $R$ , where the statistical distribution of $R_x$ and $R_y$ assumed are $\sim \mathcal{U}(8, 7.5)$ . The bounds of distribution is refer to figure 6.3. . . . .	62
6.11	Histogram of moment magnitude $M_L$ . The log-normal distribution is assumed, with a mean value of -0.39 in a 95 percent confidence interval of $[-0.46, -0.33]$ and a standard deviation of 0.56 in a 95 percent confidence interval of $[0.51, 0.61]$ . The p-value of a null hypothesis test is 5.58e-12, testing by Chi-square goodness-of-fit test . . . . .	62
6.12	Histogram of predominant frequency of GM in three directions $\omega_{g,x}$ , $\omega_{g,y}$ and $\omega_{g,z}$ . The log-normal distributions are assumed for $\omega_g$ in all directions, where the mean values, standard deviations with their 95 percent confidence interval, and p-value of a null hypothesis test, testing by Chi-square goodness-of-fit test will be presented in table 6.3 . . . . .	63
6.13	Baseline correction. The realization of ground motion is generated by SBGMG, where the input parameters $[M_L, R_x, R_y, \omega_g/2\pi, \omega_{sdo}/2\pi, \zeta_{sdo}] = [0.74, 0.75, 14.89, 17.78, 13.47, 0.04]$ . The blue line is the ground motion before baseline correction, and the orange line is after baseline correction. The black dashed line is the zero baseline. . . . .	65
6.14	Response of the SDOF system excited by ground motion generated by the same input parameters as in figure 6.13. The Orange line is the ground motion, and the blue is the corresponding response. . . . .	66
6.15	Numerical example. Histogram of maximum velocity $v_{max}$ (mm/s) in $z$ -direction in logarithm scale. The total number of realizations is 500. The input parameters for SBGMG are shown in table 6.4 and 6.5. . . . .	66
6.16	Comparison of validation data and SPCE prediction of maximum velocities, $\mathbf{X} = X_{sb}$ . . . . .	68
6.17	Comparison of validation data and SPCE prediction of maximum velocities at sample points $X_{sb,i}, i = 1 \cdots 4$ . . . . .	69
6.18	Comparison of the convergence of the surrogate model with different training size $N$ . . . . .	71
6.19	Time comparison of different sizes of training data $N$ , $p = 4$ , $q_{norm} = 1$ . . . .	74
6.20	Time comparison of different combinations of $p$ and $q_{norm}$ . . . . .	74
6.21	Time comparison of running in single/multiple cores. . . . .	75
7.1	Building Geometry at geothermal energy site in Taufkirchen. 1 <sup>st</sup> Perspective. . . . .	77
7.2	Top view of the building, the black lines inside the red area are the placement of walls. The black dot is the target point recording the FRFs, whose coordinate is $[8.75, 4.8]$ (m) for $x$ and $y$ directions; the $z$ coordinates correspond to the height of each floor. . . . .	80

7.3	500 frequency response function of example buildings in the vertical direction at each floor, where Young's modulus and damping ratio are uncertain, following the probability distribution shown in table 7.1. The parameter for SSI is included in table 7.2. . . . .	81
7.4	500 frequency response function of example buildings in both horizontal directions at the topmost floor, where Young's modulus and damping ratio are uncertain, following the probability distribution shown in table 7.1. The parameter for SSI is included in table 7.2. . . . .	82
7.5	3D Response function built by SPCE against $M_L$ and $R$ in z-direction. The black-mesh surfaces are the standard deviation computed from 100 replications . . . . .	83
7.6	2D Response function built by SPCE against $M_L$ and $R$ in z-direction . . . .	83
7.7	Comparison of validation data and SPCE prediction of maximum velocities in 3 directions . . . . .	84
7.8	QoI-based Sobol's indices in three directions. The orange bars indicate the top two significant parameters where the large SPCE coefficients are set. The blue bars are the comparably less significant parameters affecting the system. Here, the label " $(R_x, R_y)$ " represents the higher-order index, and the label "Others" is the remaining indices that are insignificant. . . . .	87
7.9	3D Fragility curve for Comfortability assessment, following the standard of DIN-4150-2 for daytime, where $A_0 = 3$ and $c_F = 0.8$ . The z-axis describes the probability $p_f$ of whether the response will exceed the limit state or not. . . . .	89
7.10	2D Fragility curve for Comfortability assessment, following the standard of DIN-4150-2 for daytime, where $A_0 = 3$ and $c_F = 0.8$ . The z-axis describes the probability $p_f$ of whether the response will exceed the limit state or not. . . . .	89
7.11	3D Fragility curve for Comfortability assessment, following the standard of DIN-4150-2 for night, where $A_0 = 0.2$ and $c_F = 0.8$ . The z-axis describes the probability $p_f$ of whether the response will exceed the limit state or not. . . . .	90
7.12	2D Fragility curve for Comfortability assessment, following the standard of DIN-4150-2 for night, where $A_0 = 3$ and $c_F = 0.8$ . The z-axis describes the probability $p_f$ of whether the response will exceed the limit state or not. . . . .	90
7.13	3D Fragility curve for Serviceability assessment, following the standard of DIN-4150-3, where $v_{max} = 15$ (mm/s). The z-axis describes the probability $p_f$ of whether the response will exceed the limit state or not. . . . .	91
7.14	2D Fragility curve for Serviceability assessment, following the standard of DIN-4150-3, where $v_{max} = 15$ (mm/s). The z-axis describes the probability $p_f$ of whether the response will exceed the limit state or not. . . . .	91
A.1	Step 1: Fitting the FRF of Hu model with recorded ground motion . . . . .	96
A.2	Step 2: Transfer the GWN from the time domain to the frequency domain. . . . .	96
A.3	Step 3: Generate the ground motion by the Hu model. . . . .	97
A.4	Step 4: Generate the artificial ground motion based on Hu model and time modulating function . . . . .	98
A.5	Step 5: The comparison of artificial ground motion and recorded one. . . . .	99

A.6	Framework of SPCE model, including the use of nested optimization procedure of Gaussian process and L-BFGS-B, modified from [Zhu and Sudret 2023]. . . . .	100
A.7	Overall flowchart of the simulation and training. The simulation has three parts: the pre-processor, solver, and post-processor. The modelings and simulations for building are run in ANSYS ®. . . . .	101
A.8	Histograms of model parameters $\omega_g$ , $\zeta_g$ , $r_h$ , $PGA$ in x-direction. The frequency in the histogram is normalized by the relative probability, where the $PGA$ is presented on the logarithm scale. . . . .	102
A.9	Histograms of model parameters $\omega_g$ , $\zeta_g$ , $r_h$ , $PGA$ in y-direction. The frequency in the histogram is normalized by the relative probability, where the $PGA$ is presented on the logarithm scale. . . . .	102
A.10	Quantile plots of the model parameter in x-direction. Hollow circles indicate the first and the third quartiles. . . . .	103
A.11	Quantile plots of the model parameter in y-direction. Hollow circles indicate the first and the third quartiles. . . . .	103
A.12	Scatter plots of residuals of $\ln(PGA)$ against $M_L$ and $R$ in x-direction . . . .	108
A.13	Scatter plots of residuals of $\omega_f$ against $\omega_g$ and $R$ in x-direction . . . . .	108
A.14	Scatter plots of residuals of $\zeta_g$ against $\omega_g$ and $R$ in x-direction . . . . .	109
A.15	Scatter plots of residuals of $\ln(PGA)$ against $M_L$ and $R$ in y-direction . . . .	109
A.16	Scatter plots of residuals of $\omega_f$ against $\omega_g$ and $R$ in y-direction . . . . .	110
A.17	Scatter plots of residuals of $\zeta_g$ against $\omega_g$ and $R$ in y-direction . . . . .	110
A.18	Response spectral analysis. The top figure shows the velocity response, and the bottom shows the acceleration response. The blue line is the spectral response of the ground motion, recorded at 14:18:43 on March 27 <sup>th</sup> , 2021, from the seismic station INS3, Insheim, Germany. The orange lines show the 50 realizations of spectral response from SBGMG, where the seismic parameters $[M_L, R, \omega_g, \zeta_g] = [1.3, 8.07, 62.38, 0.39]$ , where $M_L, R$ are provided by the station and $\omega_g, \zeta_g$ are obtained by the fitting procedure. . . . .	111
A.19	Response spectral analysis. The top figure shows the velocity response, and the bottom shows the acceleration response. The blue line is the spectral response of the ground motion, recorded at 16:28:45 on February 26 <sup>th</sup> , 2021, from the seismic station TMO20, Insheim, Germany. The orange lines show the 50 realizations of spectral response from SBGMG, where the seismic parameters $[M_L, R, \omega_g, \zeta_g] = [1.2, 8.06, 56.80, 0.22]$ , where $M_L, R$ are provided by the station and $\omega_g, \zeta_g$ are obtained by the fitting procedure. . . . .	112
A.20	Building Geometry at geothermal energy site in Taufkirchen. 2 <sup>nd</sup> Perspective. . . . .	113
A.21	Building Geometry at geothermal energy site in Taufkirchen. 3 <sup>rd</sup> Perspective. . . . .	114
A.22	Building Geometry at geothermal energy site in Taufkirchen. 4 <sup>th</sup> Perspective. . . . .	114
A.23	3D Response function built by SPCE against $M_L$ and $R$ in x-direction. The black-mesh surfaces are the standard deviation computed from 100 replications . . . . .	115
A.24	2D Response function built by SPCE against $M_L$ and $R$ in x-direction . . . .	115

A.25 3D Response function built by SPCE against $M_L$ and $R$ in y-direction. The black-mesh surfaces are the standard deviation computed from 100 replications . . . . .	116
A.26 2D Response function built by SPCE against $M_L$ and $R$ in y-direction . . . .	116

# Nomenclature

## General symbols

$\mathcal{N}$		Normal distribution
$\mathcal{LN}$		Log normal distribution
$\mathcal{U}$		Uniform distribution
$\omega_{s dof}/2\pi$	Hz	Natural frequency of single degree of freedom system
$\zeta_{s dof}$		Damping ratio of single degree of freedom system
$l_f$	m	Length of the foundation
$b_f$	m	Width of the foundation
$\rho_b$	$\frac{\text{kg}}{\text{m}^3}$	Density of building material
$\rho_s$	$\frac{\text{kg}}{\text{m}^3}$	Density of soil
$\nu_b$		Poisson's ratio of building
$\nu_s$		Poisson's ratio of soil
$E_b$	$\frac{\text{N}}{\text{m}^2}$	Young's modulus
$G$	$\frac{\text{N}}{\text{m}^2}$	shear modulus
$C$		Dashpot coefficient
$\gamma$		Dimensionless dashpot coefficient
$M$		Mass
$\mu$		Dimensionless mass
$K$		Static stiffness
$\zeta_b$		Modal damping ratio of building
$N$		Number of input datasets
$\mathcal{X}$		Input datasets
$\omega$	$\frac{\text{rad}}{\text{s}}$	Angular frequency
$\mathbb{E}[\cdot]$		Mean value
$\mathbf{E}(\cdot)$		Square sum of difference

## Symbols for ground motion generation

$\mathbf{X}_{hu}$		Hu-Zhou model variables vector
$\mathbf{X}_{gmf}$		Parameters of time modulating function
$\mathbf{X}_a$		Model parameters vector of artificial ground motion generator
$\mathbf{X}_{a,opt}$		Optimized model parameters vector of artificial ground motion generator
$\mathbf{X}_{sb}$		Seismic parameters vector of Scenario-based ground motion generator
$\mathbf{M}_{K-T}$		K-T model
$\mathbf{M}_{hu}$		Hu-Zhou model
$\mathbf{S}_{K-T}$		Power spectral density function K-T model
$\mathbf{S}_{hu}$		Power spectral density function of Hu-Zhou model
$\mathbf{\Lambda}$		Vector of potential seismic parameters
$\phi$		Coefficients of linear mixed effect model
$\hat{\eta}$		Inter-event residuals
$\hat{\epsilon}$		Intra-event residuals
$\tau_{lme}$		Standard deviation of Inter-event residuals
$\sigma_{lme}$		Standard deviation of Intra-event residuals
$\mathbf{L}$		Lower triangular matrix of covariance matrix
$\mathbf{\Sigma}$		Covariance matrix
$\mathbf{I}$		Identical matrix
$\mathbf{1}$		Matrix of all ones
$\mathbf{0}$		Matrix of all zeros
$\omega_g/2\pi$	Hz	Predominate frequency of ground motion, introduced by Hu-Zhou model
$\omega_h/2\pi$	Hz	Cut off frequency of the high-pass filter, introduced by Hu-Zhou model
$\zeta_g$		Damping ratio of site, introduced by Hu-Zhou model
$M_L$		Moment magnitude
$R$	km	Rupture distance
$R_x$	km	Rupture distance in X direction
$R_y$	km	Rupture distance in Y direction
$v_s$	$\frac{m}{s}$	Shear wave velocity

$GM_{acc}$	$\frac{\text{mm}}{\text{s}^2}$	Ground motion in acceleration	
$GM_{vel}$	$\frac{\text{mm}}{\text{s}}$	Ground motion in velocity	
$GM_{disp}$	mm	Ground motion in displacement	
$PGA$	$\frac{\text{mm}}{\text{s}^2}$	Peak ground acceleration	
$PGV$	$\frac{\text{mm}}{\text{s}}$	Peak ground velocity	
$PGD$	mm	Peak ground displacement	
$T_0$	s	Start time of ground motion	
$t_5$	s	Time at 5 percentile of the percentile of the accumulated value of ground motion	
$t_{95}$	s	Time at 95 percentile of the percentile of the accumulated value of ground motion	
$t_{mid}$	s	Time at 50 percentile of the percentile of the accumulated value of ground motion	
$t_d$	s	Duration of ground motion	
$r_{mid}$		Ratio of $t_{mid}$ compared to $t_d$	$r_{mid} = \frac{t_{mid}}{t_d}$
$r_{freq}$		Ratio of $\omega_h$ compared to $\omega_g$	$r_{freq} = \frac{\omega_h}{\omega_g}$
$\varepsilon_P$		Error between two nonlinear square least fitting results	
$\epsilon_{res}$		Residuals of linear mixed effect model	
$\Omega$	Hz	Circular frequency	
$\ddot{U}$	$\frac{\text{mm}}{\text{s}^2}$	Acceleration of the rock cite	
$\Gamma$		Gamma function	
$\{u, d\}$	mm	Displacement	
$\{\dot{u}, v\}$	$\frac{\text{mm}}{\text{s}}$	Velocity	
$\{\ddot{u}, a\}$	$\frac{\text{mm}}{\text{s}^2}$	Acceleration	
$\{\tilde{a}, \tilde{v}, \tilde{d}\}$		Corrected by baseline correction	
$\{\hat{a}, \hat{v}, \hat{d}\}$		Corrective terms for baseline correction	
$\{p_a, p_v, p_d\}$		Polynomial formed for baseline correction	
$\{c_a, c_v, c_d\}$		Coefficients of $\{p_a, p_v, p_d\}$	

## Symbols for Stochastic Polynomial Chaos Expansion

$\mathbf{X}$	Random vector
$Y$	Quantity of interest (QoI)
$Y_X$	Response of input parameter set $x$
$\hat{Y}$	Prediction response



$\hat{Y}_X$	Prediction response of input parameter set $x$
$\mathcal{Y}$	Realization vector
$\mathcal{M}$	Computational model
$\mathcal{M}^{PC}$	PCE model
$\mathcal{M}_s$	Stochastic simulator
$\delta$	Kronecker symbol
$c_j$	Coefficients of $j$ -th polynomials
$\hat{c}_j$	Optimized coefficients of $j$ -th polynomials
$Z_j$	Sets of random variables, forming the $j$ -th bases of the Hilbert space
$f_{X_i}$	Marginal distributions of $i$ -th parameters
$f_{\mathbf{X}}$	Joint distributions of all parameters
$F$	Cumulative distribution function
$Z$	Latent variable
$\phi$	Arbitrary function
$\pi$	Orthogonal polynomials
$\psi$	Univariate orthonormal polynomials
$\alpha$	Polynomial term of multivariate polynomial
$\Psi_\alpha$	multivariate polynomial
$\mathcal{A}$	Polynomial basis
$\mathcal{A}^*$	Optimized polynomial basis
$\text{card}\mathcal{A}$	Number of polynomial terms in truncated series
$M$	Number of input parameters
$p$	Order of polynomial
$q$	Q-norm of hyperbolic truncation scheme
$\varepsilon$	Residual due to truncation
$\epsilon$	Noise of stochastic PCE
$\sigma$	Standard deviation of $\epsilon$
$\varphi(\cdot)$	Probability density function
$l$	Likelihood
$N_Q$	Number of integration points
$z_j$	$j$ -th integration point
$w_j$	weight of numerical integration
$\mathbf{A}$	Information matrix
$\Delta$	Difference of PCE model and realization

$\widehat{Err}_{LOO}$	Leave-one-out error
$\hat{\epsilon}_{LOO}$	Normalized leave-one-out error
$\epsilon_{val}$	error between simulation results and predictions for validation
$d_{ws}^2$	Wasserstein distance of order two
$S$	Sobol's indices

## Symbols for Serviceability and Comfortability analysis

$KB$	Weighted velocity
$KB_{Fmax}^*$	Root mean square weighted velocity
$c_F$	Empirical values for constant for different types of vibration
$LS$	Limit-state or damage state
$IM$	Intensity measurement

## Acronym

AGMG	Artificial ground motion generator
APDL	Ansys Parametric Design Language
BFGS	Broyden–Fletcher–Goldfarb–Shanno-algorithm
CDF	Cumulative distribution function
CP model	Clough-Penzien model
CV	Cross-validation
DIN	German national standard (German)
DOF	Degrees of freedom
EDP	Engineering demand parameter
FEA	Finite element analysis
FFT	Fast Fourier transform
FRF	Frequency response functions
GSA	Global sensitivity analysis
GWN	Gaussian white noise
Hu model	Hu-Zhou model

IFFT	Inverse fast Fourier transform
KDE	Kernel density estimator
K-T model	Kaiani-Tajimi model
LAR	Least-angle regression
L-BFGS-B	Limited-memory- Broyden–Fletcher–Goldfarb–Shanno- algorithm
LHS	Latin hypercube sampling
LME	Linear mixed-effect model
LOOCV	Leave-one-out cross-validation
LPM	Lumped-parameter model
MLE	Maximum likelihood estimation
OLS	Ordinary least square
PCA	Principal components analysis
PCE	Polynomial Chaos Expansion
PDF	Probability density function
QoI	Quantities of interest
SBGMG	Scenario-based ground motion generator
SFEM	Stochastic finite element method
SPCE	Stochastic Polynomial Chaos Expansion

# 1 Introduction

## 1.1 Motivation

Due to the lack of energy resources, the use of renewable energy increased globally in recent decades. In Germany, more and more attention is focused on geothermal energy, which is critical in transitioning to sustainable energy sources. However, every technology has its pros and cons. With the development of this new technology, its side effects on the environment also need to be considered. According to the research, it is shown that weak to moderate-scaled earthquakes may be induced during the operation of geothermal power plants [Khansefid et al 2022; Küperkoch et al 2018].

Geothermal power plants utilize steam from high-temperature water extracted from the Earth's crust to rotate the turbines and convert the kinetic energy into electricity. In this process, underground hot water is extracted continuously, and underground water resources might be destroyed without refill, leading to further problems. To keep the underground water resource balanced, cool water will be injected at high pressure back periodically after high-temperature water is extracted. This cyclic operation might affect the stress balance between the underlying Earth's layers, creating or propagating cracks in the crustal rocks [Khansefid et al 2022]. The fractured rocks may eventually induce small to moderate earthquakes [Kraft et al 2009]. Although currently targeted induced seismic events in Insheim area, Germany, have relatively small moment magnitudes  $M_L$ , i.e.,  $M_L \leq 2.4$  [Küperkoch et al 2018], many geothermal power plants are located near populated areas, and their induced earthquakes might still be frequently perceptible, which those long-term, highly recurrent induced earthquakes could disrupt the residents' daily lives or damage the building's structure.

To evaluate the above potential issue, it is crucial to analyze the induced earthquakes and understand the impact of those seismic events on the residents' comfort and property. One conventional way to fulfill this is to perform a time history seismic analysis to predict the dynamic response of the building's structure excited by the artificial ground motion [Lagaros

et al 2013] or by the recorded ground motions from the online database, e.g., Pacific Earthquake Engineering Research Center (PEER) ground motion database, database formed by Geophysical Observatory Fürstfeldbruck (GOF) in LMU or Observatories & Research Facilities for European Seismology (ORFEUS) then perform seismic risk assessment afterward. These methods are straightforward; however, there are some limitations and drawbacks. One of the most significant issues is that performing the time history seismic analysis for the building response is time-consuming. Additionally, the analysis requires a large number of responses to obtain valuable results. On the aspect of the usage of ground motions, it is also questionable to represent the geothermal-induced ground motions by the earthquakes from the database due to the difference of the magnitude and site characteristics; for instance, most of the geothermal-induced ground motions have a low-moderate magnitude i.e.  $M_L \leq 2.4$ , while the magnitude range of PEER database is  $3 \leq M_L \leq 7.9$  [Ancheta et al 2014]. Additionally, it will be a challenge if one would like to analyze the significance of each seismic parameter following this framework. Therefore, an improved framework and more advanced methods are introduced to avoid the mentioned problems.

This thesis introduces the Probability-based surrogate model, Stochastic polynomial chaos expansion (SPCE), to efficiently represent the relationship between the parameterized geothermally induced ground motion and the maximum response of the building. Furthermore, ground motions recorded in the Insheim geothermal site with their characteristic parameters will be analyzed and utilized to train the scenario-based ground motion generator (SBGMG) to generate artificial ground motion as the excitation of the system. The seismic parameters reported from the seismic stations, such as magnitude ( $M_L$ ), rupture distance ( $R$ ), and site characteristics, will be treated as the inputs for SBGMG. Finally, sensitivity analysis, comfortability, and serviceability assessment will be performed within the scope of this thesis.

## 1.2 State of research

Although the use of artificial ground motion as excitation and surrogate models to reduce computation time is well-established in seismic risk assessment, researchers are focusing on developing more advanced and efficient methods. These improvements aim to enhance performance and provide more comprehensive insights into the investigation. This section will state the current method of generating artificial ground motion and probability-based surrogate models.

Firstly, Housner proposed to describe the ground motion by a spectral-stationary stochastic

process model and modeled the seismic acceleration by stationary white noise [Housner et al 1953]. Following this concept, Kaiani and Tajimi proposed the well-known Kaiani-Tajimi model (K-T model), which applied the Gaussian filter to the stationary white noise. [Kanai 1957; Tajimi 1960]. In this model, the ground is modeled as the SDOF system, and the excitation underground is assumed to be Gaussian-filtered white noise. However, there are some limitations, such as it cannot present the ground motion well at a low-frequency range, and the singular point appears at zero frequency, violating the continuous quadratic integrability condition. To overcome the mentioned issue, the Clough-Penzien model (CP model) [Clough 1975] and the Hu-Zhou model (Hu model) [Hu and Zhou 1962] are introduced. The concepts of these two methods are similar in that both apply an extensive high-pass filter to weaken the signal in a low-frequency range and remove the singular points at zero frequency.

On the other hand, some studies have tried to develop non-stationary ground motion. Priestley has introduced the evolutionary PSD function [Priestley 1967], a widely-used method to model the non-stationary stochastic processes. Afterward, [Rezaeian and Der Kiureghian 2010] developed a new model to introduce temporal and spectral non-stationarity. In the study, temporal non-stationarity is introduced by a gamma-distribution-shape-like time-modulating function. Moreover, the relationship between model parameters and seismic parameters from the seismic stations, e.g., rupture distance  $R$ , moment magnitude  $M_L$ , shear wave velocity  $v_s$ , etc., are analyzed. [Broccardo and Dabaghi 2017] introduced a non-parametric method to enhance the flexibility of the time-modulating function. However, this approach significantly increases the number of variables, which can complicate the training of the surrogate model.

Furthermore, a probability-based surrogate model is considered to replace the time-consuming simulation. In this thesis, the selected approach is a polynomial chaos expansion-based model (PCE). PCE was first introduced in [Wiener 1938] in 1938, using Hermite polynomials to model stochastic processes. Thereafter, several relative methods are developed. In [Xiu and Karniadakis 2002], PCE is extended by introducing the Askey scheme, making PCE work with Gaussian random variables and other random variable types. To make the algorithm more efficient, [Blatman and Sudret 2010] developed an adaptive algorithm to build up the sparse PCE, which reduces the number of polynomials by neglecting the terms with less significance. [Blatman and Sudret 2011] improved this algorithm by utilizing the least angle regression method instead of the ordinary least square method for training. Lastly, the whole framework was summarized in [Sudret 2014]. Even though the mentioned PCE surrogate model is accurate and efficient enough, there is a limitation in that it only supports the uncertainty in input variables, not in a stochastic simulator. Hence, the stochastic polynomial

chaos expansion (SPCE) model is proposed [Zhu and Sudret 2023], which can model the stochastic system efficiently.

## 1.3 Outline

The thesis aims to assess the seismic risk due to the geothermal-induced earthquake using a ground motion database recorded in the Insheim area, Germany. The introduction of SBGMG is first done in Chapter 2, which includes its assumption, overall framework, and mathematical expression. Next, in Chapter 3, the theoretical concept and the adaption of the polynomial chaos expansion (PCE) method are introduced, followed by the algorithm of stochastic polynomial chaos expansion (SPCE), which captures the stochasticity from the stochastic simulator. In Chapter 4, DIN standards on serviceability and comfortability assessment are introduced, and how to represent them by the fragility curve is also described. Then, the overall framework of seismic assessment for this thesis will be explained in Chapter 5. Following this, a simple numerical example is implemented in Chapter 6, where the performance of SPCE will also be evaluated, and the case study on building structure in Taufkirchen, Germany, is conducted in Chapter 7. The building's model and harmonic analysis are implemented in Mechanical APDL by ANSYS®. Sensitivity analysis and further comfortability and serviceability assessment will be performed based on the built surrogate model. Finally, the conclusion and outlook are written in Chapter 8.

## 2 Modeling of induced ground motion

In this chapter, the proposed scenario-based ground motion generator is described. The general procedure, including the flowchart and its assumption, is introduced, followed by its mathematical expression. Furthermore, the validation method - response spectral analysis is briefly depicted.

### 2.1 General Procedure

This section explains the process for forming the scenario-based artificial ground motion (SBGMG). The method is based on the ground motion model proposed by [Hu and Zhou 1962] and the framework introduced by [Rezaeian and Der Kiureghian 2010]. The procedure can be separated into the following steps:

1. In the research, artificial ground motion is assumed spectral stationarity. All the signals in the database are first transferred to the frequency domain by Fast Fourier Transform (FFT) for further tasks.
2. Fit the artificial ground motion generator (AGMG) with recorded ground motions from the database by the nonlinear least square method. Once the fitting procedure is finished, the statistical distribution of the model parameters of the ground motions  $\mathbf{X}_a$  will be evaluated and prepared for the further fitting procedure.
3. Build the Hu model. The model is the function of the variables,  $\mathbf{X}_{hu} = \{\omega_g, \zeta_g, r_h\}$ . Furthermore, the time-modulating function is considered for temporal non-stationarity, and three additional variables,  $\mathbf{X}_{gmf} = \{PGV, t_d, r_{mid}\}$ , are introduced. Thus, the AGMG is formed as follows:

$$GM_{vel} = \mathbf{AGMG}(\mathbf{X}_a)$$

$$\mathbf{X}_a = [\mathbf{X}_{hu}, \mathbf{X}_{gmf}]^T \quad (2.1)$$



research collects over 200 ground motions recorded at the Insheim Geothermal Site, Germany where its induced seismicity has been continuously monitored since 2012 over a time span of more than four years [Küperkoch et al 2018]

## 2.2 Assumptions for geothermally induced ground motion modeling

It is crucial to define the stationarity of the ground motion when designing the model. In this thesis, temporal non-stationarity describes the variation of the ground motion in time, while spectral non-stationarity represents how the intensity in the frequency domain varies with time.

Several state-of-the-art methods assumed the non-stationarity of earthquake ground motion in both time and frequency domains and state that the spectral non-stationarity of the ground motion can depict the seismic waves recorded at a site better than a stationary one; however, it might not be necessary to utilize such complex approach if the goal is to evaluate the structure response excited by the geothermal-induced earthquake. The reason is the following.

In [Rezaeian and Der Kiureghian 2010], it is mentioned that typically, earthquakes could be separated into three parts, wherein the initial few seconds of motion, high-frequency P waves tend to dominate the ground motion, and the moderate-frequency S waves will appear in the following strong-motion phase; finally, the low-frequency, long-wavelength surface waves will replace it, and Liu shows that the frequency content may be different in the initial and intermediate stages, but the frequency content could be treated as constant during the strong-motion stage [Liu 1970], which implies that if the analysis focuses on the response of the structure under the strong-motion stage, the spectral stationarity assumption might be enough.

Plus, it has been mentioned that satisfactory accuracy could be achieved when analyzing the seismic response of linear structures using the stationary stochastic model [Hu and Zhou 1962]. Although the result is approximate, it is practical and efficient in solving the seismic assessment problem and designing the structure. Therefore, this thesis will assume spectral stationarity to simplify the system.

Additionally, in the thesis, it is assumed that:

- The artificial ground motion generated is spectral stationarity.
- The artificial ground motion generated is temporal non-stationarity, controlled by a time-modulating function.
- The artificial ground motion is based on Gaussian white noise (GWN), and its repeatability is controlled by the random seed.
- To make the analysis more efficient and consistent, the durations of the artificial ground motion are fixed at 15 seconds, and the parameters controlling the shape of the time-modulating function are the same; only the amplitude will be modified based on input *PGV*. The shape of the time-modulating function has no apparent impact on the results due to the spectral stationary assumption.
- The structure's response is analyzed by harmonic analysis instead of nonlinear time history analysis, which implies that the system has to satisfy all the assumptions of Harmonic analysis. The system can be treated as a linear structure fulfilling the mode superposition assumption.

## 2.3 Generation of artificial ground motion

Most of the state-of-the-art stochastic stationary ground motion simulators are based on the concept of the K-T model. Therefore, it is essential to first introduce the K-T model concept. This section will present an overview of the K-T model and the proposed Hu model, which is an adaptation of the K-T model.

### 2.3.1 Kanai-Tajimi model

The assumptions of the K-T model are:

1. The movement of the rock excited the system is GWN.
2. The mean value of GWN is zero.
3. The soil layers above the rock site are treated as a linear single DOF system; the schematic diagram is shown as figure 2.2.

The mathematical expression is deduced from the governing equation of the vibration excited by the acceleration of the base (equation (2.3)).

$$\ddot{u}(t) + 2\zeta_g\omega_g\dot{u}(t) + \omega_g^2u(t) = -\ddot{U}(t) \quad (2.3)$$

where  $u$  is the displacement of the system,  $\zeta_g$ ,  $\omega_g$  are the damping ratio and natural frequency of the ground motion, and  $\ddot{U}$  denotes the rock acceleration, which is GWN. Let's convert equation (2.3) to the frequency domain by Fourier transformation, equation (2.3) will be formed as equation (2.4).

$$-\Omega^2u(\Omega) + i2\Omega\zeta_g\omega_gu(\Omega) + \omega_g^2u(\Omega) = -\ddot{U}(\Omega) \quad (2.4)$$

After sorting the equation (2.4),  $u(\Omega)$  can be written as equation (2.5)

$$u(\Omega) = \frac{\ddot{U}(\Omega)}{(\Omega^2 - \omega_g^2) - i2\Omega\zeta_g\omega_g} \quad (2.5)$$

The ground acceleration  $\ddot{u}_g(t)$  can be written as equation (2.6);

$$\ddot{u}_g(t) = \ddot{u}(t) + \ddot{U}(t) = -2\zeta_g\omega_g\dot{u}(t) - \omega_g^2u(t) \quad (2.6)$$

Its formulation in the frequency domain is also stated.

$$\ddot{u}_g(\Omega) = -i2\Omega\zeta_g\omega_gu(\Omega) - \omega_g^2u(\Omega) \quad (2.7)$$

By inserting equation (2.5) to equation (2.7), the ground acceleration will be made as the function of rock acceleration, shown in equation (2.8). It is also the definition of the K-T model  $\mathbf{M}_{K-T}$ .

$$\begin{aligned} \ddot{u}_g(\Omega) &= \frac{-i2\Omega\zeta_g\omega_g - \omega_g^2}{(\Omega^2 - \omega_g^2) - i2\Omega\zeta_g\omega_g} \ddot{U}(\Omega) \\ &= \mathbf{M}_{K-T}(\Omega, \omega_g, \zeta_g) \end{aligned} \quad (2.8)$$

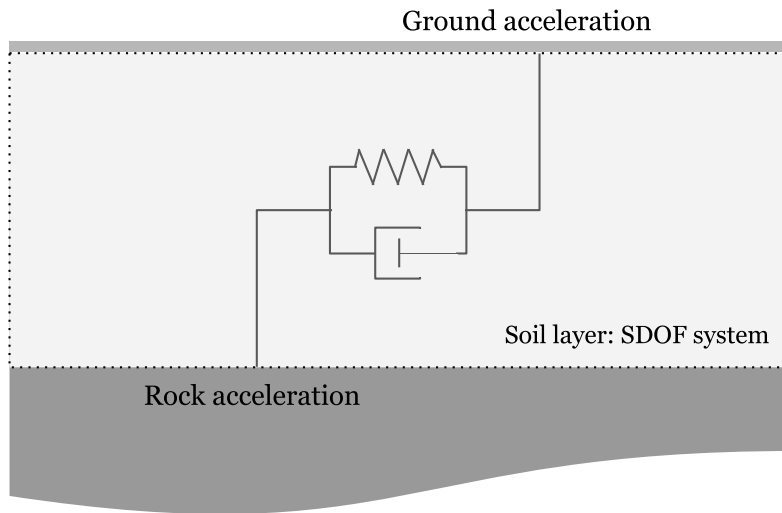
The power spectral density function of the K-T model  $S_{K-T}$  is then expressed as the equation

(2.9)

$$\mathbf{S}_{\text{K-T}}(\Omega) = \frac{(2\Omega\zeta_g\omega_g)^2 + \omega_g^4}{(\Omega^2 - \omega_g^2)^2 + 2\Omega\zeta_g\omega_g)^2} S_0 \quad (2.9)$$

where  $\mathbf{S}_{\text{K-T}}(\Omega)$  is the power spectral intensity of the ground and  $S_0$  is the constant spectral intensity of the rock motions. In the generating process, the parameters,  $\Omega, \omega_g, \zeta_g$  are obtained by fitting equation (2.9) to recorded ground motion. Those parameters will then be used in equation (2.8) to generate the artificial ground motion.

As shown earlier, The K-T model has a clear physical significance and is easy to understand. However, two shortcomings of the model in the model must be improved. First, it might overestimate the amplitude in the low-frequency range. This might lead to an unreasonable response if the system is identified as a low-frequency structure. Second, the model has a singular point at zero frequency, which causes the acceleration to not be continuous and leads to infinite amplitude in velocity or displacement term if integrated [Chen et al 2022]. To address these two issues, a high-pass filter is proposed to adjust the amplitude in the low-frequency range of the K-T spectral model. Additionally, the Hu model, a modified version of the K-T model, is introduced to further refine the approach.



**Figure 2.2:** Schematic diagram of K-T model, according to [Chen et al 2022].

### 2.3.2 Hu-Zhou model

Compared to the K-T model, using GWN as excitation from the rock site, the Hu model replaces it with color noise by applying a high pass filter on GWN. The frequency range of the color noise is controlled by spectral parameters  $\omega_h$  or the ratio  $r_h = \frac{\omega_h}{\omega_g}$ . Here, the spectral parameter  $\omega_h$  has been used for convenience.

The filter equation of the color noise is as follows:

$$\begin{aligned} \ddot{y}(t) + \omega_h^3 y(t) &= u_g(t) \\ \ddot{u}_{hu}(t) &= \ddot{y}(t) \end{aligned} \quad (2.10)$$

where  $\ddot{u}_{hu}(t)$  is the response of the Hu model,  $y(t)$  is the response of the filter,  $\omega_h$  is the cutoff frequency for the Hu model and  $u_g(t)$  is the input, obtained from equation (2.10).

Let transform equation (2.10) to the frequency domain

$$-i\Omega^3 y(\Omega) + \omega_h^3 y(\Omega) = u_g(\Omega) \quad (2.11)$$

and

$$\ddot{u}_{hu}(\Omega) = -i\Omega^3 y(\Omega) \quad (2.12)$$

By sorting equation (2.11) and substituting to equation (2.12)

$$\ddot{u}_{hu}(\Omega) = \frac{-i\Omega^3}{-i\Omega^3 + \omega_h^3} u_g(\Omega) \quad (2.13)$$

Following this, recalling the K-T model (equation (2.8)) and multiplication it with equation (2.13), the Hu model  $\mathbf{M}_{hu}$  is formed.

$$\begin{aligned} \ddot{u}_{hu}(\Omega) &= \frac{-i\Omega^3}{-i\Omega^3 + \omega_h^3} \frac{-i2\Omega\zeta_g\omega_g - \omega_g^2}{(\Omega^2 - \omega_g^2) - i2\Omega\zeta_g\omega_g} \ddot{U}(\Omega) \\ &= \mathbf{M}_{hu}(\Omega, \omega_g, \zeta_g, \omega_h) \end{aligned} \quad (2.14)$$

The power spectral density function of the Hu model  $\mathbf{S}_{hu}$  is then expressed as the equation (2.15)

$$\mathbf{S}_{hu}(\Omega) = \frac{\Omega^6}{\Omega^6 + \omega_h^6} \frac{(2\Omega\zeta_g\omega_g)^2 + \omega_g^4}{(\Omega^2 - \omega_g^2)^2 + (2\Omega\zeta_g\omega_g)^2} S_0 \quad (2.15)$$

To better describe the relationship between  $\omega_g$  and  $\omega_h$ , their ratio  $r_h = \frac{\omega_g}{\omega_f}$  is introduced and used to substitute the  $\omega_h$  in the equation (2.14) and (2.15). The FRF of the Hu model is presented in figure 2.3.

$$\begin{aligned} \ddot{u}_{hu}(\Omega) &= \frac{-i\Omega^3}{-i\Omega^3 + (r_h\omega_g)^3} \frac{-i2\Omega\zeta_g\omega_g - \omega_g^2}{(\Omega^2 - \omega_g^2) - i2\Omega\zeta_g\omega_g} \ddot{U}(\Omega) \\ &= \mathbf{M}_{hu}(\Omega, \omega_g, \zeta_g, r_h) \end{aligned} \quad (2.16)$$

and power spectral density function reads:

$$\mathbf{S}_{hu}(\Omega) = \frac{\Omega^6}{\Omega^6 + (r_h\omega_g)^6} \frac{(2\Omega\zeta_g\omega_g)^2 + \omega_g^4}{(\Omega^2 - \omega_g^2)^2 + (2\Omega\zeta_g\omega_g)^2} S_0 \quad (2.17)$$

The Hu model is now in the frequency domain. Finally, the signal needs to be transformed back to the time domain. This process can be done quickly by performing the IFFT technique in the in-built MATLAB toolbox.

$$\mathbf{M}_{hu}(t, \omega_g, \zeta_g, r_h) = \text{IFFT}(\mathbf{M}_{hu}(\Omega, \omega_g, \zeta_g, r_h)) \quad (2.18)$$

The Hu model, however, only controls the behavior of ground motion in the spectral domain. A time-modulating function is necessary to model the non-stationarities in the temporal domain.

In this thesis, the time at a certain percentile of the accumulated value of ground motion  $t_p$  and maximum velocity value design the shape of the ground motion. The default design points are:

1. the maximum velocity (PGV) occurs at  $t_{mid} = 5$  seconds, where  $t_{mid}$  is the 45 percentile of accumulated value.
2. there is no amplitude from 0 sec to start time ( $T_0$ ).
3. the start time ( $T_0$ ) occurs at  $t_5 = 2.5$  seconds
4. the end time occurs at  $t_{95} = 10$  seconds
5. the default duration is 20 seconds.

By fitting the gamma PDF and matching the percentile of the accumulated value of ground motion to the time information, i.e., start and end time, the parameters  $a$  and  $b$  are obtained inversely. Additionally, a property of gamma PDF can also be utilized to adjust the time where PGV occurred in the time-modulating function: The mean of gamma PDF or the highest amplitude of the gamma-modulating function is located at the multiplication of parameter  $a$  and  $b$  (equation (2.20)).

$$\arg \max_{\tau} q(\tau, a, b) = \mathbb{E}[q(\tau, a, b)] = ab = t_{mid} \quad (2.20)$$

By specifying the  $t_{mid}$  at a certain time and updating the fitted parameter  $b = \frac{t_{mid}}{a}$ , the modulating function can shift the maximum velocity to the desired  $t_{mid}$ , thus meeting the default design points mentioned above. The time-modulating function is illustrated in figure 2.4

where,

$$\begin{aligned}\mathbf{X}_{hu} &= [\omega_g, \zeta_g, r_{freq}]^T \\ \mathbf{X}_{gmf} &= [PGV, t_d, r_{mid}]^T\end{aligned}\tag{2.23}$$

### 2.3.5 Fitting the model parameter from the database

Since the model is a nonlinear function of its model parameters,  $\mathbf{X}_a = \{\omega_g, \zeta_g, r_{freq}, PGV, t_d, r_{mid}\}$ , the nonlinear least square method is implemented. This method constructs a difference between the Hu model and recorded ground motion in the form of a square sum:

$$\begin{aligned}\mathbf{E}(\mathbf{X}_a) &= \sum_{k=1}^n [S_k - \mathbf{AGMG}(\omega_k, \mathbf{X}_a)]^2 \\ \mathbf{X}_a &= [p1, p2, p3, p4, p5, p6] = [\omega_g, \zeta_g, r_{freq}, PGV, t_d, r_{mid}]\end{aligned}\tag{2.24}$$

where  $n$  is the number of discrete points of the spectral series;  $S_k$  and  $\mathbf{AGMG}(\omega_k, \mathbf{X}_a)$  are the spectral amplitudes of recorded ground motion acceleration and the Hu model at  $\omega_k$  respectively;  $\mathbf{X}_a$  is the vector of parameters that input to the model.

Once the equation (2.24) is formed and the initial values of  $\mathbf{X}_a = \mathbf{X}_a^{(0)}$  are inputted, the algorithm will linearize the equation firstly by Taylor series expansion at corresponding  $\mathbf{X}_a$ . Then, it searches for the optimized  $\mathbf{X}_a$  that minimizes the square sum of difference,  $\mathbf{E}$ , iteratively, following the procedure of derivative-based optimization. The nonlinear least square method details are not mentioned here, but in [Chen et al 2022], the general formulation is shown as equation (2.25).

$$\mathbf{X}_{a,opt} = \arg \min_{\mathbf{X}_a} \mathbf{E}(\mathbf{X}_a)\tag{2.25}$$

It is worth mentioning that the nonlinear least square method depends on initial guess  $\mathbf{X}_a^{(0)}$ . An inappropriate  $\mathbf{X}_a^{(0)}$  might lead to an undesired local minimum or convergence difficulty. In this thesis, the fitting procedure is repeated three times with randomly selected initial values of  $\mathbf{X}_a^{(0)}$  for each recorded ground motion. If two of the optimization processes return two optimized candidates  $\mathbf{X}_{a,opt}^{cand.i}$ ,  $i = 1, 2$ , where  $\sum[|\mathbf{X}_{a,opt}^{cand.2} - \mathbf{X}_{a,opt}^{cand.1}|] \leq \varepsilon_P$ , the mean value of these two results will be chosen as the optimized  $\mathbf{X}_{a,opt}$ .



## 2.4 Relationship of seismic parameters and model parameters

Even though AGMG is sufficient to generate the artificial ground motion for further analysis, its input model parameters,  $\mathbf{X}_a$ , are not well linked with the seismic parameters,  $\mathbf{X}_{sb}$ . Therefore, regression analysis will be introduced to describe the relationship between  $\mathbf{X}_a$  and  $\mathbf{X}_{sb}$ . Once the regression model is built,  $\mathbf{X}_a$  can be obtained by inputting  $\mathbf{X}_{sb}$  and used to generate artificial ground motion from AGMG.

### 2.4.1 Linear mixed-effects model

Usually, there is only one  $M_L$  and  $D$ , but several  $R$  will be reported from each event due to the different distances between seismic stations and the source. This makes using a simple ordinary regression model inappropriate in this case. Hence, the LME model is proposed to address this issue, which is suitable for handling the problem with weighting observation. Unlike the ordinary regression model, the LME model assumes the record and data within each event (each earthquake cluster) are statically dependent [Rezaeian 2010]. The regression model can be stated as in equation (2.26).

$$\mathbf{X}_{a\ i,jk} = \text{LME}_i(\mathbf{\Lambda}_{jk}, \phi_i) + \hat{\eta}_{i,j} + \hat{\epsilon}_{i,jk} \quad (2.26)$$

where  $\mathbf{X}_{a\ i,jk}$  is the output of LME model used for AGMG.  $\mathbf{\Lambda}_{jk}$  is the vector of the potential input parameters e.g.  $M_L, R, D$  (Here the parameters are described as "potential" because some might be ignored if high values are shown in the p-value test; therefore, new notation  $\mathbf{\Lambda}_{jk}$  is used instead of  $\mathbf{X}_{sb\ jk}$ ).  $\phi_i$  are the coefficients of regression and  $\hat{\eta}_{i,j}, \hat{\epsilon}_{i,jk}$  are the inter-event<sup>1</sup> and intra-event<sup>2</sup> residuals, which are under the assumption that their error terms are independent, zero-mean, and follow a normal distribution with variance  $\tau_{lme}^2$  and  $\sigma_{lme}^2$ , respectively.

Additionally, the total error of  $\mathbf{X}_{a\ i,jk}$  will also be a zero-mean, normal distributed random variable with variance  $\tau_{lme}^2 + \sigma_{lme}^2$ . Regarding the indices,  $i = 1, \dots, n_i$  are for  $\mathbf{X}_a$ .  $j = 1, \dots, n_j$  are the indices of the seismic event (earthquake), and  $k = 1, \dots, n_k$  denotes the indices of the records of  $j^{th}$  seismic event.  $n_i, n_j, n_k$  are the number of them respectively.

<sup>1</sup>Inter-event: within a single event or group

<sup>2</sup>Intra-event: between different events or groups

A realization  $\mathbf{x}_a$  of LME model with uncertainty follow distribution  $\sim \mathcal{N}(0, \sqrt{\tau_{lme}^2 + \sigma_{lme}^2})$  can be generated by:

$$\mathbf{x}_a = \mathbf{LME}(\mathbf{\Lambda}, \boldsymbol{\phi}) + \mathbf{L}_{XX}^T \mathbf{y}_a \quad (2.27)$$

LME is the mean function with fitted coefficients  $\boldsymbol{\phi}_i$  at certain input parameter set  $\mathbf{\Lambda}$ , and  $\mathbf{L}_{XX}^T$  is a lower triangular matrix obtained from the Cholesky decomposition of the covariance matrix  $\boldsymbol{\Sigma}_{XX}$ , formed by  $\mathbf{X}_a$ ;  $\mathbf{y}_a$  is a realization of the vector of uncorrelated standard normal random variables following distribution  $\sim \mathcal{N}(0, \sqrt{\tau_{lme}^2 + \sigma_{lme}^2})$ . To generate such normally distributed random variables, the Ziggurat algorithm implemented by MATLAB toolbox is utilized [Marsaglia and Tsang 2000].

## 2.4.2 Fitting LME model and estimating the variance of error

To find the  $\boldsymbol{\phi}_i$  and estimate  $\tau_{lme}^2$  and  $\sigma_{lme}^2$ , maximum likelihood technique is introduced. The likelihood function could be formed as equation (2.28) [Rezaeian 2010].

$$L(\boldsymbol{\phi}_i, \tau_{lme}^2, \sigma_{lme}^2) = \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}[\mathbf{X}_{a\ i} - \mathbf{LME}_i(\mathbf{\Lambda}_i, \boldsymbol{\phi}_i)]^T \boldsymbol{\Sigma}_i^{-1} [\mathbf{X}_{a\ i} - \mathbf{LME}_i(\mathbf{\Lambda}_i, \boldsymbol{\phi}_i)]\right) \quad (2.28)$$

$\boldsymbol{\Sigma}_i$  is the covariance matrix, expressing as the function of  $\tau_{lme}^2$  and  $\sigma_{lme}^2$  in the from:

$$\boldsymbol{\Sigma}_i(\tau_{lme}^2, \sigma_{lme}^2) = \begin{bmatrix} \sigma_{lme}^2 \mathbf{I}_{n1} + \tau_{lme}^2 \mathbf{1}_{n1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sigma_{lme}^2 \mathbf{I}_{n2} + \tau_{lme}^2 \mathbf{1}_{n2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{lme}^2 \mathbf{I}_{n_j} + \tau_{lme}^2 \mathbf{1}_{n_j} \end{bmatrix}_{N \times N} \quad (2.29)$$

where  $\mathbf{I}_{n_j}$  is the identical matrix of size  $n$ , corresponding to number of record in  $j^{th}$  event.  $\mathbf{1}$  is a matrix of all one with size of  $n \times n$ .  $\mathbf{0}$  is a matrix of all zero, filled in the off-diagonal space. Once the equation (2.28) is formed,  $\boldsymbol{\phi}_i, \tau_{lme}, \sigma_{lme}$  can be found by maximizing the likelihood.

$$\{\boldsymbol{\phi}, \tau_{lme}, \sigma_{lme}\} = \arg \max_{\boldsymbol{\phi}, \tau_{lme}, \sigma_{lme}} L(\boldsymbol{\phi}, \tau_{lme}^2, \sigma_{lme}^2) \quad (2.30)$$

This thesis uses the MATLAB in-built optimization toolbox to fit the LME model [Pineiro and Bates 1996].

## 2.5 Baseline correction

When numerical integration is performed on an acceleration signal, due to the inappropriate assumption of the initial condition of velocity and displacement or truncating the signal, baseline shift or drift might occur in the computed velocity and displacement signal. In this case, baseline correction is crucial to remove the incorrect baseline for further dynamics analysis if the inconsistent acceleration signal is used.

Polynomial detrending is a widely used approach for removing the incorrect baseline in a time series. However, this approach might lead to signal inconsistency if the detrending technique is applied individually to velocity and displacement signals without considering the cross-effect. Another popular way to remove inconsistent trending is applying a high-pass filter [Trifunac 1971] to the acceleration signal since the shifted and drifted baseline is typically synthetic by the low-frequency signal. Nevertheless, [Pan et al 2016] argued that this approach might not perform well if the baseline of displacement drifts severely since insufficient attention is drawn to velocity and displacement signal, only on acceleration itself.

To obtain the well-corrected displacement signal from artificial acceleration data for further harmonic analysis, this thesis uses the baseline correction method developed by [Pan et al 2016], which could output the signal that is consistent between acceleration, velocity, and displacement after numerical integration is done. Its general concept will be stated in the following.

First, let's assume the relationships between uncorrected and corrected time series are:

$$\begin{aligned}\tilde{a}(t) &= a(t) + \hat{a}(t) \\ \tilde{v}(t) &= v(t) + \hat{v}(t) \\ \tilde{d}(t) &= d(t) + \hat{d}(t)\end{aligned}\tag{2.31}$$

Where  $a(t), v(t), d(t)$  are the original, uncorrected time series of acceleration, velocity, and displacement, respectively.  $\tilde{a}(t), \tilde{v}(t), \tilde{d}(t)$  represent the corrected time series;  $\hat{a}(t), \hat{v}(t), \hat{d}(t)$

represent the corrective time series.

Second,  $v(t)$  can be obtained by numerically integrating  $a(t)$ . Following this, the algorithm fit  $v(t)$  with polynomial  $p_v(t)$  that initial condition  $p_v(t_0) = 0$  by least squares method. The  $M^{th}$  order polynomial  $p_v(t)$  with coefficients  $c_{v,i}$  is formed as equation (2.32).

$$p_v(t) = \sum_{i=1}^{M+1} c_{v,i}(t - t_0)^i \quad (2.32)$$

After  $p_v(t)$  is obtained, corrective acceleration from velocity term  $\hat{a}_v(t)$  can be expressed as equation (2.33).

$$\hat{a}_v(t) = \frac{d}{dt}p_v(t) = - \sum_{i=1}^{M+1} [ic_{v,i}(t - t_0)^{i-1}] \quad (2.33)$$

Similarly, the algorithm fits  $p_d(t)$  with  $d(t)$  by the same fitting procedure, in which  $d(t)$  is from twice the numerical integration of  $a(t)$ . Then  $p_d(t)$  and  $\hat{a}_d(t)$  are written as

$$\begin{aligned} p_d(t) &= \sum_{i=2}^{M+2} c_{d,i}(t - t_0)^i \\ \hat{a}_d(t) &= \frac{d^2}{dt^2}p_d(t) = - \sum_{i=2}^{M+2} [i(i-1)c_{d,i}(t - t_0)^{i-2}] \end{aligned} \quad (2.34)$$

where  $c_{d,i}$  is the coefficient of polynomial  $p_d(t)$  and  $\hat{a}_d(t)$  denotes corrective acceleration from displacement term.

Finally, the corrected acceleration  $\tilde{a}_d(t)$  after baseline correction will be formed.

$$\begin{aligned} \tilde{a}_d(t) &= a(t) - \hat{a}_v(t) - \hat{a}_d(t) \\ &= a(t) - \sum_{i=1}^{M+1} [ic_{v,i}(t - t_0)^{i-1}] - \sum_{i=2}^{M+2} [i(i-1)c_{d,i}(t - t_0)^{i-2}] \end{aligned} \quad (2.35)$$

## 2.6 Model validation through response spectral analysis

Response spectral analysis is an approach to estimate the structure response under short, nondeterministic, transient dynamic events [COMSOL 2019]. It is based on the mode super-

position of the response, implying that the use of response spectral analysis also has to fulfill the assumption of the linear system. The concept of this approach assumes the structure is a mass-spring-damper system attached to a moving base where the damping ratio is fixed at a certain value, but the natural frequency of the structure is variable, excited by the input signal. During sweeping the natural frequency of structure from low to high, the peak responses at each certain frequency will be recorded continuously, eventually forming the plot of maximum response against the frequency or period.

From the result, it can easily be interpreted where the natural frequency of the input signal is located or how the structure will respond based on the specific input type. In other words, response spectral analysis will not directly represent the frequency content of the input signal as applying the Fourier transform. Instead, it shows the effect of the structure under such signals.

This thesis uses response spectral analysis [Mostafa Tazarv 2024] to validate the artificial ground motion by comparing it with the recorded events. The valid synthetic ground motion should respond similarly to the recorded one, with a close natural frequency and response pattern.

## 3 Surrogate model

To simplify the real-world challenge to a solvable problem, engineers tend to define the system and input parameters as deterministic, e.g., performing finite element analysis (FEA) with fixed material properties and density. Most of the time, this assumption is valid, and its result is representative; however, to fully understand the relationship between the system and input variables, including uncertainty propagation, conventional FEA might be unsatisfied, and a more advanced method that can capture the uncertainty introduced by whole systems is required.

In this thesis, the input parameters have the uncertainty, following certain probability distributions, inherited from the nature of recorded ground motion; plus, the ground motion generator is based on a stochastic simulator, meaning that the proposed surrogate model needs to have the ability to capture all the stochasticity not only from input variables but also from the system. The overall framework of SPCE is shown in Appendix, figure A.6.

### 3.1 Introduction of possible surrogate model

In the early 90s, the stochastic finite element method (SFEM) was developed. This approach could solve the problem with uncertain coefficients [Ghanem and Spanos 2003]. In the concept of SFEM, a crucial method called polynomial chaos expansion (PCE) is used for uncertainty propagation. This method allows one to represent a random output as a polynomial series in the input parameters [Sudret 2014], and it is one of the widely-used techniques to deal with the uncertainty introduced by input variables in engineering problems during the last decades. Although this approach takes a huge step in dealing with uncertainty quantification, it is still challenging to represent the system's stochasticity, i.e., the system that outputs different results when running the stochastic simulator repeatedly with the same input variables.

A few approaches have been proposed to capture those phenomena. Those methods can be separated into two categories based on whether they require replication and random seed manipulation.

For the first category, which required simulation replication, Moutoussamy proposed forming the response distribution at different sample points by replicating the simulation and filling the entire input space with parametric or non-parametric techniques. This surrogate method can accurately represent the result if the number of replications is large enough, but, on the other hand, it might not be suitable if the simulation is too time-consuming.

Another category, without simulation replication, utilizes the concept of statistical approaches to build the surrogate model, such as a generalized linear model or kernel density estimator (KDE). In comparison to the generalized linear model, its response distribution has to belong to the exponential family; a non-parametric estimator has an advantage in fitting arbitrary distribution types. However, it is well-known to suffer from the curse of dimensionality. To solve this problem, [Zhong et al 2023] proposed using principal components analysis (PCA) and diffusion map to reduce the dimension of the input first and form the KDE in reduced-order space, then generating the realization by Ito stochastic differential equation followed by the transformation back to the original space. This approach can build a surrogate model efficiently; however, several hyperparameters in the algorithm must be tuned manually depending on different structural systems.

Furthermore, [Zhu and Sudret 2021] proposed an approach that uses generalized lambda distribution to fit the response distribution, which takes advantage of the non-parametric approach and follows the parametric assumption. This approach, however, has the limitation that it cannot represent the multimodal distribution. Finally, the stochastic polynomial chaos expansions model (SPCE) is introduced [Zhu and Sudret 2023]. This surrogate model represents the randomness from the realizations by the latent variable  $Z$  and noise term  $\epsilon$ ; it also overcomes the limitation of reproducing the multimodal distribution. Moreover, the foundation of SCPE is based on the classical PCE method, meaning that it inherits the advantage of the PCE method. For instance, the global sensitivity analysis is easy to perform once the surrogate model is built, which makes this approach practical and appealing for the current research.

## 3.2 Stochastic polynomial chaos expansion

SPCE surrogate model includes two parts. The first part introduces the PCE model, which is the mean function of the surrogate model. The second part explains the theory and technique of extending the PCE to the SPCE model. The second section also briefly mentions the optimization methods, L-BFGS-B, and the Gaussian process. Finally, the framework of the SPCE model will be plotted at the end. For convenience, in section 3.2.1, the same notations as [Sudret 2014] are used, except for symbol related to the PCE coefficients, which denotes by  $\{\mathbf{c}, c_\alpha\}$ , instead of  $\{\mathbf{y}, y_\alpha\}$  to avoid misunderstanding. On the other hand, the mathematical expression notates in section 3.2.2 corresponds to [Zhu and Sudret 2023].

### 3.2.1 Polynomial chaos expansions

#### Mathematical setting of PCE

Let us consider the system in which the system's behavior is represented by computational model  $\mathcal{M}$  and input parameters are represented by random vector  $\mathbf{X}$ . Its associated quantity of interest (QoI)  $Y$  is formed as :

$$Y = \mathcal{M}(\mathbf{X}) \quad (3.1)$$

where  $\mathbf{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_M\}^T$ , each  $\mathbf{X}_i$  corresponds to one random vector in  $M$  input parameters. The input parameters  $\mathbf{X}$  can be described by certain joint probability density function  $f_{\mathbf{X}}$ , computed by the marginal distribution of all input parameter vectors  $\{f_{X_i}, i = 1 \dots M\}$ . Each input parameter vector  $\{\mathbf{X}_i, i = 1 \dots M\}$  has to be independent. Otherwise, the Cupula approach needs to be performed first [Nelsen 1999].

If  $Y$  assumes to have a finite variance, the Hilbert space of second-order random variables assumption is valid, and  $Y$  can be described as [Soize and Ghanem 2004]:

$$Y = \sum_{j=0}^{\infty} c_j Z_j \quad (3.2)$$

In equation (3.2),  $Y$  is cast as an infinite series, where  $\{c_j\}_{j=0}^{\infty}$  are the coefficients and  $\{Z_j\}_{j=0}^{\infty}$  is the numerable set of random variables, forming the bases of the Hilbert space. The basis terms  $\{Z_j\}_{j=0}^{\infty}$  are multivariate orthonormal polynomials in the input vector  $\mathbf{X}$ , which is interpreted as  $Z_j = \Psi_j(\mathbf{X})$ .



### The basis: Univariate orthonormal polynomials

Let's assume now the components of  $\mathbf{X}$ ,  $\{\mathbf{X}_i, i = 1 \dots M\}$  are independent. The joint distribution is the product of the  $M$  marginal distributions  $\{f_{X_i}\}_{i=1}^M$ :

$$f_{\mathbf{X}}(x) = \prod_{i=1}^M f_{X_i}(x_i), \quad x_i \in \mathcal{D}_{x_i} \quad (3.3)$$

where  $\mathcal{D}_{x_i}$  is the space of  $X_i$ . In [Sudret 2014], a functional inner product by the following integral is defined in equation (3.4). This is formed for each single variable  $X_i$  and any two functions  $\phi_1, \phi_2 : x \in D_{X_i} \mapsto \mathbb{R}$ . It can also be interpreted as the expectation  $\mathbb{E}[\phi_1(X_i), \phi_2(X_i)]$  with respect to the marginal distribution  $f_{X_i}$ .

$$\langle \phi_1, \phi_2 \rangle_i = \int_{D_{X_i}} \phi_1(x) \phi_2(x) f_{X_i} dx = \mathbb{E}[\phi_1(X_i), \phi_2(X_i)] \quad (3.4)$$

If the above two functions are orthogonal for the probability measure  $\mathbb{P}(dx) = f_{X_i}(x)dx$ , then  $\mathbb{E}[\phi_1(X_i), \phi_2(X_i)] = 0$ . Using this concept, one can build a family of orthogonal polynomials  $\{\pi_k^{(i)}, k \in \mathbb{N}\}$  that satisfy:

$$\begin{aligned} \langle \pi_j^{(i)}, \pi_k^{(i)} \rangle_i &= \int_{D_{X_i}} \pi_j^{(i)}(x) \pi_k^{(i)}(x) f_{X_i}(x) dx \\ &= \mathbb{E}[\pi_j^{(i)}(X_i), \pi_k^{(i)}(X_i)] = a_j^i \delta_{jk} \end{aligned} \quad (3.5)$$

where subscript  $k$  denotes the degree of the polynomial  $\pi_k^{(i)}$ ,  $\delta_{jk}$  is the Kronecker symbol, and  $a_j^i$  corresponds to the squared norm of  $\pi_k^{(i)}$ :

$$a_j^i = \|\pi_j^{(i)}\|_i^2 = \langle \pi_j^{(i)}, \pi_j^{(i)} \rangle_i \quad (3.6)$$

Now, the obtained family of polynomial  $\pi_k^{(i)}$  is usually orthogonal but not normalized yet. To make it orthonormal, the normalization of polynomial  $\pi_k^{(i)}$  is performed to form the orthonormal family  $\{\psi_j^{(i)}\}_{j=0}^\infty$  by:

$$\psi_j^{(i)} = \pi_k^{(i)} / \sqrt{a_j^i} \quad i = 1 \dots M, \quad j \in \mathbb{N} \quad (3.7)$$

The classical families of orthogonal polynomials and their normalizing coefficients are included in the table in [Sudret 2014].

### The basis: Multivariate polynomials

To build up the basis (equation (3.2)), let us define multi-indices (also called tuples)  $\boldsymbol{\alpha} \in \mathbb{N}^M$ , which are ordered lists of integers,  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_M)$ ,  $\alpha_i \in \mathbb{N}$ . The multivariate polynomial  $\Psi_{\boldsymbol{\alpha}}$  to any multi-index  $\boldsymbol{\alpha}$  can be formed by:

$$\Psi_{\boldsymbol{\alpha}}(x) = \prod_{i=1}^M \psi_{\alpha_i}^{(i)}(x_i) \quad (3.8)$$

where the univariate polynomials  $\{\psi_k^{(i)}, k \in \mathbb{N}\}$  are defined according to the  $i$ -th marginal distribution in equation (3.5) and (3.7). This equation implies that the multivariate polynomials in the input vector  $\mathbf{X}$  are also orthonormal, described as:

$$\mathbb{E}[\Psi_{\boldsymbol{\alpha}}(\mathbf{X})\Psi_{\boldsymbol{\beta}}(\mathbf{X})] = \int_{D_{\mathbf{X}}} \Psi_{\boldsymbol{\alpha}}(\mathbf{x})\Psi_{\boldsymbol{\beta}}(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} = \delta_{\boldsymbol{\alpha}\boldsymbol{\beta}} \quad \forall \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}^M \quad (3.9)$$

where  $\delta_{\boldsymbol{\alpha}\boldsymbol{\beta}}$  is the Kronecker symbol. Consequently,  $Y = \mathcal{M}(\mathbf{X})$  in equation (3.1) can be written as in [Soize and Ghanem 2004]. The random response  $Y$  in an abstract space through an orthonormal basis and coefficients are the coordinates of  $Y$  in the basis, which are formed by the set of all multivariate polynomials in the input random vector  $\mathbf{X}$ :

$$Y = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\mathbf{X}) \quad (3.10)$$

### Truncation scheme and curse of dimensionality

The representation in equation (3.10) involves an infinite sum of terms. However, the series needs to be truncated to finite terms in practice. Therefore, a truncation scheme needs to be introduced. Since the series is constructed using polynomials, it can be truncated by limiting the polynomial basis to a certain degree. The standard truncation scheme is defined by polynomial terms whose total degree is smaller than a given value  $p$ . First the total degree of a multivariate polynomial  $\Psi_{\boldsymbol{\alpha}}$  is formed as:

$$|\boldsymbol{\alpha}| = \sum_{i=1}^M \alpha_i \quad (3.11)$$

The standard truncation scheme following the above definition is then shown as follows:

$$\mathcal{A}^{M,p} = \{\boldsymbol{\alpha} \in \mathbb{N}^M : |\boldsymbol{\alpha}| \leq p\} \quad (3.12)$$

where the number of polynomial terms in truncated series is

$$\text{card } \mathcal{A}^{M,p} = \binom{M+p}{p} = \frac{(M+p)!}{M!p!} \quad (3.13)$$

In equation (3.13), the terms of the cardinality of  $\mathcal{A}^{M,p}$  increases with  $M$  and  $p$ , and the number of coefficients rise tremendously when  $M \geq 10$ . This phenomenon is called the curse of dimensionality, which describes the high complexity of a system due to large dimensions. Two approaches have been utilized to reduce this computational complexity in the PCE. The first approach is introducing the hyperbolic truncation scheme, replacing the standard one. The second approach will be mentioned later as an adaptive approach. Both methods can be used simultaneously to efficiently reduce the number of coefficients that must be trained. Here, the hyperbolic truncation scheme is defined as selecting all multi-indices  $\boldsymbol{\alpha}$  of  $q$ -norm,  $q$ , less than or equal to  $p$ , and the  $q$  ranges in  $(0,1]$ :

$$\mathcal{A}^{M,p,q} = \{\boldsymbol{\alpha} \in \mathbb{N}^M : \|\boldsymbol{\alpha}\|_q \leq p\} \quad (3.14)$$

where  $\|\boldsymbol{\alpha}\|_q$  is defined as:

$$\|\boldsymbol{\alpha}\|_q = \left( \sum_{i=1}^M \alpha_i^q \right)^{\frac{1}{q}} \quad (3.15)$$

It is observed that if  $q = 1$ , then the hyperbolic truncation scheme is equivalent to the standard one. Typically, the maximum polynomial degree  $p$  may be selected between 3 and 5, and the testing  $q$ -norm will be  $q = \{0.25, 0.5, 0.75, 1\}$  in practical application [Sudret 2014].

### Adaptive approach: Least square angle method

The second approach of reducing the number of coefficients to be computed is introduced in this section. In the previous section, the hyperbolic truncation scheme a priori minimizes the number of coefficients in a truncated series expansion. However, it might still be too costly if a large-dimensional, high-complex problem is faced. In [Sudret 2014], it is mentioned that it is often observed a posteriori that the non-zero coefficients in the expansion form a sparse subset of  $\mathcal{A}^{M,p,q}$ . Therefore, it is better to focus only on the sparse basis with non-zero coefficients and neglect the useless terms in the expansions. This procedure can be done using the least square angle method (LAR) algorithm [Efron et al 2004].

### Computation of the coefficients

Once the truncated basis is selected, the coefficients  $\{c_\alpha\}_{\alpha \in \mathcal{A}^{M,p,q}}$  could be computed. The advantage of the PCE algorithm is that it can be treated as a non-intrusive approach; it requires no modification of the computation model and relies on replication of the model for realizations of random vector  $\mathbf{X}$ , same as the concept of Monte-Carlo simulation. Hence, the ordinary least-square (OLS) is suitable for estimating the coefficients.

Since the polynomial basis in equation (3.10) is orthonormal, it can be formed that:

$$\mathbb{E}[Y\Psi_\alpha(\mathbf{X})] = \mathbb{E}\left[\Psi_\alpha(\mathbf{X}) \cdot \sum_{\beta \in \mathbb{N}^M} c_\beta \Psi_\beta(\mathbf{X})\right] = \sum_{\beta \in \mathbb{N}^M} c_\beta \cdot \mathbb{E}[\Psi_\alpha(\mathbf{X})\Psi_\beta(\mathbf{X})] = c_\alpha \quad (3.16)$$

Hence, each  $c_\alpha$  is treated as the orthogonal projection of the random response  $Y$  onto the corresponding basis function  $\Psi_\alpha(\mathbf{X})$ . Additionally, equation (3.16) can be written as:

$$c_\alpha = \mathbb{E}[Y\Psi_\alpha(\mathbf{X})] = \int_{D_{\mathbf{X}}} \mathcal{M}(\mathbf{x})\Psi_\alpha(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} \quad (3.17)$$

OLS is implemented once the approach to estimate each coefficient is formed (equation (3.17)). Since the truncation scheme is purposed, e.g.,  $\mathcal{A} = \mathcal{A}^{M,p,q}$ , the infinite series of equation (3.1) is recast as the sum of the truncated series and a residual  $\varepsilon$ :

$$Y = \mathcal{M}(\mathbf{X}) = \sum_{\alpha \in \mathcal{A}} c_\alpha \Psi_\alpha(\mathbf{X}) + \varepsilon \quad (3.18)$$

where  $\varepsilon$  denotes the truncated error computed by the ignoring terms. The concept of OLS approach is finding the set of  $y_\alpha$  that minimizes the mean square error  $\mathbb{E}[\varepsilon^2]$  that:

$$\mathbb{E}[\varepsilon^2] = \mathbb{E}\left[\left(Y - \sum_{\alpha \in \mathcal{A}} c_\alpha \Psi_\alpha(\mathbf{X})\right)^2\right] \quad (3.19)$$

The equation of OLS is then stated as:

$$\begin{aligned} \mathbf{c} &= \arg \min_{\mathbf{c} \in \mathbb{R}^{card\mathcal{A}}} \mathbb{E}\left[\left(Y - \sum_{\alpha \in \mathcal{A}} c_\alpha \Psi_\alpha(\mathbf{X})\right)^2\right] \\ &= \arg \min_{\mathbf{c} \in \mathbb{R}^{card\mathcal{A}}} \mathbb{E}\left[\left(\mathcal{M}(\mathbf{X}) - \sum_{\alpha \in \mathcal{A}} c_\alpha \Psi_\alpha(\mathbf{X})\right)^2\right] \end{aligned} \quad (3.20)$$

The discretized version of equation (3.20) is as follows:

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c} \in \mathbb{R}^{\text{card} \mathcal{A}}} \frac{1}{n} \sum_{i=1}^n \left( \mathcal{M}(\mathbf{x}^{(i)}) - \sum_{\alpha \in \mathcal{A}} c_{\alpha} \Psi_{\alpha}(\mathbf{x}^{(i)}) \right)^2 \quad (3.21)$$

where  $\mathcal{X} = \{\mathbf{x}^{(i)}, i = 1, \dots, n\}$  is a sample set of points from input random vector  $\mathbf{X}$ .

The analytical solution of the least-square minimization problem (equation (3.21)) is proceed as follows:

1. Compute the response  $y^{(i)}$  at each sample set of points  $\mathbf{x}^{(i)}$  in input random vector  $\mathbf{X}$  and stored as a realization vector  $\mathcal{Y}$ :

$$\mathcal{Y} = \{y^{(1)} = \mathcal{M}(\mathbf{x}^{(1)}), \dots, y^{(n)} = \mathcal{M}(\mathbf{x}^{(n)})^T\} \quad (3.22)$$

2. Forming the information matrix from the basis of polynomials onto each sample set of points  $\mathbf{x}^{(i)}$ :

$$\mathbf{A} = \{\mathbf{A}_{ij} = \Psi_j(\mathbf{x}^{(i)}), i = 1, \dots, n, j = 1, \dots, \text{card } \mathcal{A}\} \quad (3.23)$$

3. The analytical solution of the least-square minimization problem is formed:

$$\hat{\mathbf{c}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathcal{Y} \quad (3.24)$$

Once the equation (3.24) is solved, the approximation of the response  $\hat{Y}$  is presented as:

$$\hat{Y} = \mathcal{M}^{PC}(\mathbf{X}) = \sum_{\alpha \in \mathcal{A}} \hat{c}_{\alpha} \Psi_{\alpha}(\mathbf{X}) \quad (3.25)$$

The above equation can generate the prediction by inputting the sample set of points  $\mathbf{x}^{(i)}$ . This is the so-called surrogate model of the original model  $y = \mathcal{M}(\mathbf{x})$ .

$$\mathcal{M}^{PC}(\mathbf{x}^{(i)}) = \sum_{\alpha \in \mathcal{A}} \hat{c}_{\alpha} \Psi_{\alpha}(\mathbf{x}^{(i)}) \approx \mathcal{M}(\mathbf{x}^{(i)}) \quad (3.26)$$

### Leave-one-out cross-validation

As mentioned in the previous section, the maximum polynomial degree  $p$  and  $q$ -norm,  $q$ , in the hyperbolic truncation scheme (equation (3.14)) for sufficient accuracy cannot be known

in advance. Therefore, a posteriori error estimation for searching the optimal  $p$  and  $q$  is introduced in [Blatman and Sudret 2010], where leave-one-out cross-validation (LOOCV) is proposed [Allen 1971].

To perform the LOOCV, let us separate one sample set of points  $\mathbf{x}^{(i)}$  out from  $\mathcal{X}$ . This denotes as  $\mathcal{X} \setminus \mathbf{x}^{(i)} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(i-1)}, \mathbf{x}^{(i+1)}, \dots, \mathbf{x}^{(n)}\}$ . The predicted residual error at  $\mathbf{x}^{(i)}$  is:

$$\Delta_i = \mathcal{M}(\mathbf{x}^{(i)}) - \mathcal{M}^{PC \setminus i}(\mathbf{x}^{(i)}) = \frac{\mathcal{M}(\mathbf{x}^{(i)}) - \mathcal{M}^{PC}(\mathbf{x}^{(i)})}{1 - h_i} \quad (3.27)$$

where  $h_i$  is the  $i$ -th diagonal term of matrix  $\mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ . Finally, the leave-one-out error read:

$$\widehat{Err}_{LOO} = \frac{1}{n} \sum_{i=1}^n \Delta_i^2 = \frac{1}{n} \sum_{i=1}^n \left( \frac{\mathcal{M}(\mathbf{x}^{(i)}) - \mathcal{M}^{PC}(\mathbf{x}^{(i)})}{1 - h_i} \right)^2 \quad (3.28)$$

Additionally, the normalized version of the leave-one-out error  $\hat{\epsilon}_{LOO}$  is obtained and corrected [Chapelle et al 2002] as shown in follows:

$$\hat{\epsilon}_{LOO} = \frac{1}{n - P} \left( \frac{1 + \frac{1}{n} \text{tr}(\mathbf{C}_{emp}^{-1})}{Var \mathcal{Y}} \right) \sum_{i=1}^n \left( \frac{\mathcal{M}(\mathbf{x}^{(i)}) - \mathcal{M}^{PC}(\mathbf{x}^{(i)})}{1 - h_i} \right)^2 \quad (3.29)$$

where  $\text{tr}(\cdot)$  is the trace,  $P = \text{card } \mathcal{A}$  and  $\mathbf{C}_{emp} = \frac{1}{n} \mathbf{\Psi}^T \mathbf{\Psi}$ .

The optimal  $p$  and  $q$  can be obtained by searching the minimum normalized leave-one-out error  $\hat{\epsilon}_{LOO}$  in different combinations of  $p$  and  $q$  in the PCE algorithm. However, in the following SPCE algorithm,  $\widehat{Err}_{LOO}$  will eventually be used instead of  $\hat{\epsilon}_{LOO}$  for further cross-validation process.

### Advance solver: Hybrid-LAR approach for estimating coefficients $\mathbf{c}$

In the previous section, the ordinary least-square approach is used to estimate the coefficients  $\{c_\alpha\}_{\alpha \in \mathcal{A}^{M,p,q}}$ . It is noticed that the coefficients  $c_\alpha$  correspond to the polynomial basis  $\mathcal{A}^{M,p,q}$ , which can be reduced to an optimal polynomial basis  $\mathcal{A}^*$ , introduced in LAR algorithm. To do so, the efficient approach, Hybrid-LAR, that maintains satisfactory accuracy is proposed in [Blatman and Sudret 2011], which combines the LAR and ordinary least-square approach. The overall framework of Hybrid-LAR is as follows:

1. Perform the original LAR algorithm only once and obtain the sequence of less and less sparse expansion.
2. Compute the coefficients  $\mathbf{c}_\alpha$  corresponding to each sparse expansion iteratively from single basis  $\mathcal{A}^{(1)}$  with only one basis term  $\Psi_{\alpha_1}$  to the full basis  $\mathcal{A}^{(m)}$  with sequence  $(\Psi_{\alpha_1}, \dots, \Psi_{\alpha_m})$ , where  $m = \text{card}\mathcal{A}^{M,p,q}$  by ordinary least-square regression. Then, the sets of the metamodel are constructed  $\{\widehat{\mathcal{M}}_{\mathcal{A}^{(1)}}, \dots, \widehat{\mathcal{M}}_{\mathcal{A}^{(m)}}\}$ .
3. Compute the normalized leave-one-out error  $\hat{\epsilon}_{LOO}$  of each metamodel  $\widehat{\mathcal{M}}_{\mathcal{A}}$  using the a cross-validation procedure.
4. Eventually, the metamodel associated with the lowest  $\hat{\epsilon}_{LOO}$  estimated will be retained. Its corresponding optimal basis  $\mathcal{A}^*$  will also be obtained. This metamodel will be the PCE surrogate model  $\widehat{\mathcal{M}}^{PC}$  replacing the equation (3.26) as:

$$\widehat{\mathcal{M}}^{PC}(\mathbf{x}^{(i)}) = \sum_{\alpha \in \mathcal{A}^*} \hat{c}_\alpha \Psi_\alpha(\mathbf{x}^{(i)}) \quad (3.30)$$

### 3.2.2 Extension of PCE for stochastic simulators

#### Emulating stochastic simulators

Section 3.2.1 introduces the PCE surrogate model. This expansion of this model (equation (3.26)) is a deterministic function of input parameters. It cannot directly model the stochastic simulators.

First, let us define a stochastic simulator mathematically:

$$\begin{aligned} \mathcal{M}_s : \mathcal{D}_X \times \Omega &\longrightarrow \mathbb{R} \\ (\mathbf{x}, \omega) &\longrightarrow \mathcal{M}_s(\mathbf{x}, \omega) \end{aligned} \quad (3.31)$$

where  $\mathcal{D}_X$  denotes the domain of the input parameters, and  $\Omega$  is the probability space of system internal stochasticity. The latent random variables  $\Xi(\omega)$  are introduced, which implicitly affects the system and are not considered input variables. In other words, the stochastic simulator  $\mathcal{M}_s$  can be treated as a deterministic function of the input vector  $\mathbf{x}$  and the latent variable  $\Xi$ . Still, the  $\Xi$  cannot be controlled when evaluating the model. Therefore, the output remains random when  $\mathbf{x}$  is fixed and  $\Xi$  is random, following the

assumed probability distribution [Zhu and Sudret 2023]. To represent such a relationship, the model is constructed following equation (3.31):

$$\begin{aligned} Y_{\mathbf{X}} &= \mathcal{M}_s(\mathbf{x}, \omega) \\ Y &= \mathcal{M}_s(\mathbf{X}(\omega), \omega) \end{aligned} \quad (3.32)$$

The  $Y_{\mathbf{X}}$  here is the model response for input parameter set  $\mathbf{x}$ , and  $Y$  denotes the model output that includes all the uncertainties. From a probabilistic perspective,  $Y_{\mathbf{X}}$  is the expression of conditional random variable  $[Y|\mathbf{X} = \mathbf{x}]$ .

Let us denote the associated cumulative distribution function (CDF) by  $F_{Y|\mathbf{X}}(y|\mathbf{x})$ . According to the probability integral transform theory, *any* continuous random variable  $Z$  can be transformed to the desired distribution:

$$Y_{\mathbf{x}} = F_{Y|\mathbf{X}}^{-1}(F_Z(Z)|\mathbf{x}) \quad (3.33)$$

where  $F_Z$  is the CDF of  $Z$ . Both  $Y_{\mathbf{X}}$  and  $F_{Y|\mathbf{X}}^{-1}(F_Z(Z)|\mathbf{x})$  follow the same distribution. Furthermore, the right-hand side of equation (3.33) is a deterministic function of both  $\mathbf{x}$  and  $Z$ . If  $Y$  has a finite variance, the right-hand side of equation (3.33) can be formed by the sense of the PCE model in the  $(\mathbf{X}, Z)$  space.

$$\begin{aligned} Y &= F_{Y|\mathbf{X}}^{-1}(F_Z(Z)|\mathbf{X}) = \sum_{\alpha \in \mathbb{N}^{M+1}} c_{\alpha} \psi_{\alpha}(\mathbf{X}, Z) \\ Y_{\mathbf{x}} &= F_{Y|\mathbf{X}}^{-1}(F_Z(Z)|\mathbf{x}) = \sum_{\alpha \in \mathbb{N}^{M+1}} c_{\alpha} \psi_{\alpha}(\mathbf{x}, Z) \end{aligned} \quad (3.34)$$

Finally, a truncation scheme  $\mathcal{A}$  is applied to reduce equation (3.34) from infinite terms to finite terms.

$$Y_{\mathbf{x}} \approx \tilde{Y}_{\mathbf{x}} = \sum_{\alpha \in \mathcal{A}} c_{\alpha} \psi_{\alpha}(\mathbf{x}, Z) \quad (3.35)$$

This equation (3.33) can represent the response distribution by transforming a latent variable. Here, the latent variable  $Z$  cannot describe the specific prediction process of simulation but is used to reproduce the stochasticity, representing the distribution of  $Y_{\mathbf{x}}$ .

### Formulation of stochastic polynomial chaos expansions

To form the SPCE and cover the error from the truncated scheme and from the uncertainty, an additive noise variable  $\epsilon$  is introduced. The stochastic surrogate model can be defined



according to equation (3.35):

$$Y_{\mathbf{x}} \approx \tilde{Y}_{\mathbf{x}} = \sum_{\alpha \in \mathcal{A}} c_{\alpha} \psi_{\alpha}(\mathbf{x}, Z) + \epsilon = G_{\mathbf{x}} + \epsilon \quad (3.36)$$

where  $\epsilon$  is a Gaussian random variable with standard deviation  $\sigma$ , that  $\epsilon \sim \mathcal{N}(0, \sigma^2)$  and  $G_{\mathbf{x}} = \sum_{\alpha \in \mathcal{A}} c_{\alpha} \psi_{\alpha}(\mathbf{x}, Z)$  for convenience. By adding this noise variable, the PDF of  $\tilde{Y}_{\mathbf{x}}$  is:

$$f_{\tilde{Y}_{\mathbf{x}}}(y) = (f_{G_{\mathbf{x}}} * f_{\epsilon})(y) = \int_{-\infty}^{+\infty} f_{G_{\mathbf{x}}}(y - t) f_{\epsilon}(t) dt \quad (3.37)$$

where equation (3.37) is bounded by following according to Hölder's inequality.

$$\|f_{G_{\mathbf{x}}}\|_1 \|f_{\epsilon}\|_{\infty} = \|f_{\epsilon}\|_{\infty} = \frac{1}{\sigma\sqrt{2\pi}} \quad (3.38)$$

This implies that the PDF of  $\tilde{Y}_{\mathbf{x}}$  and the associated likelihood function are bounded. This standard deviation  $\sigma$  for noise variable  $\epsilon$  will be treated as an additional hyperparameter that needs to be fitted. The next section introduces the fitting procedure for coefficients  $\mathbf{c}$  and hyperparameter  $\sigma$ .

### Fitting procedure

To fit the SPCE by optimizing the coefficients  $\mathbf{c}$  and hyperparameter  $\sigma$ , the first step is the formulation for maximum likelihood estimation (MLE). Here the PDF of equation (3.36) can be computed as:

$$\begin{aligned} f_{\tilde{Y}_{\mathbf{x}}}(y) &= \int_{D_Z} f_{\tilde{Y}_{\mathbf{x}}|Z}(y|z) f_Z(z) dz \\ &= \int_{D_Z} \frac{1}{\sigma} \varphi\left(\frac{y - \sum_{\alpha \in \mathcal{A}} c_{\alpha} \psi_{\alpha}(\mathbf{x}, z)}{\sigma}\right) f_Z(z) dz \end{aligned} \quad (3.39)$$

where the  $\varphi$  denotes the corresponding PDF. Since  $[\tilde{Y}_{\mathbf{x}}|Z = z]$  is Gaussian random variable that  $\sim \mathcal{N}(G_{\mathbf{x}}, \sigma)$ , the likelihood of the coefficients  $\mathbf{c}$  conditioned on  $\sigma$  can be formed as:

$$l(\mathbf{c}; \mathbf{x}, y, \sigma) = \int_{D_Z} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y - \sum_{\alpha \in \mathcal{A}} c_{\alpha} \psi_{\alpha}(\mathbf{x}, z))^2}{2\sigma^2}\right) f_Z(z) dz \quad (3.40)$$

To make equation (3.40) computationally practical, Gaussian quadrature [Golub and Welsch 1969], the numerical integration schemes, is introduced and equation (3.40) can be approxi-

mated as:

$$l(\mathbf{c}; \mathbf{x}, y, \sigma) \approx \tilde{l}(\mathbf{c}; \mathbf{x}, y, \sigma) = \sum_{j=1}^{N_Q} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y - \sum_{\alpha \in \mathcal{A}} c_{\alpha} \psi_{\alpha}(\mathbf{x}, z_j))^2}{2\sigma^2}\right) w_j \quad (3.41)$$

where  $N_Q$  is the number of integration points,  $z_j$  is the  $j$ -th integration point, and  $w_j$  is the corresponding weight.  $z_j$  and  $w_j$  is obtained from the weight function  $f_Z$ . For instance, if  $Z$  is assumed to follow Gaussian distribution, the Hermite polynomial as a weighted function  $f_Z$  is considered, and Gauss–Hermite quadrature will be implemented. Based on equation (3.41) and the input data  $(\mathcal{X}, y)$ , the coefficients  $\mathbf{c}$  can be fitted by MLE if  $\sigma$  is known:

$$\tilde{\mathbf{c}} = \arg \max_{\mathbf{c}} \sum_i^N \log \left( \tilde{l}(\mathbf{c}; \mathbf{x}^{(i)}, y^{(i)}, \sigma) \right) \quad (3.42)$$

Since the gradients of equations (3.40) and (3.42) can be derived analytically. the derivative-based optimizer L-BFGS-B is proposed to solve for the coefficients  $\mathbf{c}$ . L-BFGS-B is an optimization algorithm in the family of quasi-Newton methods that approximates the BFGS algorithm using a limited amount of computer memory [Liu and Nocedal 1989]. It is very suitable for machine learning or the current problem that a large number of coefficients  $\mathbf{c}$  might be fitted (sometimes more than 100 coefficients). This optimizer is written in the SciPy python package [Virtanen et al 2020].

To ensure convergence when optimizing equation (3.42), an appropriate initial guess for the L-BFGS-B algorithm is required. According to the properties of the polynomial chaos basis functions, the mean of the SPCE is:

$$\tilde{m}(\mathbf{x}) = \mathbb{E} [\tilde{Y}_{\mathbf{x}}] = \mathbb{E}_{Z, \epsilon} \left[ \sum_{\alpha \in \mathcal{A}} c_{\alpha} \psi_{\alpha}(\mathbf{x}, Z) + \epsilon \right] = \sum_{\alpha \in \mathcal{A}, \alpha_z = 0} c_{\alpha} \psi_{\alpha}(\mathbf{x}) = \sum_{\alpha \in \mathcal{A}_m} c_{\alpha} \psi_{\alpha}(\mathbf{x}) \quad (3.43)$$

where  $\alpha_z$  is the degree of the univariate polynomial in  $Z$  and  $\mathcal{A}_m$  is defined as  $\{\alpha \in \mathcal{A} : \alpha_z = 0\}$ . And if Hybrid-LAR solver is used, the optimal basis  $\mathcal{A}^*$  is considered, then

$$\tilde{m}(\mathbf{x}) = \sum_{\alpha \in \mathcal{A}_m^*} c_{\alpha} \psi_{\alpha}(\mathbf{x}) \quad (3.44)$$

where  $\mathcal{A}_m^* = \{\alpha \in \mathcal{A}^* : \alpha_z = 0\}$ . For convenience, let's assume basis  $\mathcal{A}_m$  is used. An appropriate initial guess of coefficients  $\mathbf{c}_{\alpha}^0$  are consisted of two parts, which are  $\{\mathbf{c}_{\alpha}^0 : \alpha \in \mathcal{A}_m\}$  and  $\{\mathbf{c}_{\alpha}^0 : \alpha \in \mathcal{A} \setminus \mathcal{A}_m\}$ , respectively. For the first part,  $\mathbf{c}_{\alpha}^0$  are computed by OLS or Hybrid-LAR solver introduced in section (3.2.1). For the other  $\mathbf{c}_{\alpha}^0$  in the second part, the values are initialized randomly. The further "Warm-start strategy" to avoid instability or convergence

to undesired local optimum is mentioned in [Zhu and Sudret 2023].

As mentioned above, in the equation (3.42), the coefficients  $\mathbf{c}$  are only fitted after  $\sigma$  is assigned. Until now, the method estimating  $\sigma$  has not been discussed yet. Hence, in the next section, the approach for optimizing  $\sigma$  and performing cross-validation is focused.

### Cross-validation and fitting the hyperparameter $\sigma$

In [Zhu and Sudret 2023], it is mentioned that the hyperparameter  $\sigma$  cannot be estimated with the coefficients  $\mathbf{c}$  simultaneously by derivative-based optimizer since the singularity might occur for certain choices of  $\mathbf{c}$  and  $\sigma = 0$ . Hence,  $\sigma$  has to be tuned individually from estimating  $\mathbf{c}$ .

It is proposed to apply cross-validation (CV) [Hastie et al 2009] or a more advanced method, so-called  $N_{cv}$ -fold CV, to select the optimal value of  $\sigma$ .  $N_{cv}$ -fold CV first randomly equal-distributes the data  $(\mathcal{X}, \mathbf{y})$  into  $N_{cv}$  groups  $\{V_k : k = 1, \dots, N_{cv}\}$ . For  $k \in \{k = 1, \dots, N_{cv}\}$ , the  $k$ -th group  $V_k$  is chosen randomly as validation set and other  $N_{cv} - 1$  folds, denoting as  $V_{\sim k}$ , is selected as the training set. Therefore, the procedure (equation (3.42)) will be re-written as:

$$\tilde{\mathbf{c}}_k(\sigma) = \arg \max_{\mathbf{c}} \sum_{i \in V_{\sim k}}^N \log \left( \tilde{l}(\mathbf{c}; \mathbf{x}^{(i)}, \mathbf{y}^{(i)}, \sigma) \right) \quad (3.45)$$

where the coefficients  $\mathbf{c}$  depend on  $\sigma$ . Additionally, the likelihood  $\tilde{l}$  of out-of-sample performance is evaluated by inputting validation set  $V_k$  once  $\tilde{\mathbf{c}}_k(\sigma)$  is computed:

$$l_k(\sigma) = \sum_{i \in V_k}^N \log \left( \tilde{l}(\tilde{\mathbf{c}}_k(\sigma); \mathbf{x}^{(i)}, \mathbf{y}^{(i)}, \sigma) \right) \quad (3.46)$$

By repeating the procedure (equation (3.46)) for each group  $k \in \{k = 1, \dots, N_{cv}\}$ , their returned likelihood  $l_k(\sigma)$  or so-called the "score" is summed up as a performance of inputted  $\sigma$ , described as *CV score*. Finally, the optimal value of  $\sigma$  is chosen as the one that maximizes *CV score*:

$$\tilde{\sigma} = \arg \max_{\sigma} \sum_{k=1}^{N_{cv}} l_k(\sigma) \quad (3.47)$$

Here, the nested optimization is used, where  $\sigma$  needs to be optimized in equation (3.47), following with the estimation of coefficients  $\mathbf{c}$  in equation (3.45). The derivative-free Bayesian

optimizer [Snoek et al 2012] or so-called Gaussian process is introduced to tune the hyperparameter  $\sigma$  and solve equation (3.47). The search space of  $\sigma$  ranges in  $[0.1, 1.1] \times \sqrt{\bar{\varepsilon}}$  where  $\bar{\varepsilon} = \widehat{Err}_{LOO}$  came from the LOOCV of the PCE algorithm mentioned in section 3.2.1. A detailed explanation of such boundary for searching domain is written in [Zhu and Sudret 2023]. Here, the Gaussian process optimizer is implemented in the scikit-learn python package [Pedregosa et al 2011].

### 3.3 Validation and error measurement

In machine learning, separating the training set and testing set is crucial. The experimental design (training set) and testing set are generated independently by the Latin hypercube sampling (LHS) approach.

After the model is trained from the training set, the response predictions will be generated based on the data points from the testing set. A global and local comparison will be performed to validate the trained model. All the predictions in the testing set will be compared with their corresponding simulation result for the global test. On the other hand, four random data points in the testing set will be selected and used to generate predictions for the local test. When repetitively running the simulation and generating the same number of predictions at those fixed data points, their comparison can be used to evaluate the measurement's accuracy of the stochasticity in the system.

To compare the surrogate model with the simulation result quantitatively, the error measurement is defined as:

$$\varepsilon_{val} = \frac{\mathbb{E}[d_{ws}^2(Y_X, \hat{Y}_X)]}{Var[Y]} \quad (3.48)$$

Where  $Y_X$  is the simulation result of data point  $\mathbf{x}$ , and  $\hat{Y}_X$  denotes the prediction (response of surrogate model) of the same data point as  $Y_X$ ,  $Y$  is all the simulation results propagate the uncertainty from input variables and the system itself, finally,  $d_{ws}^2$  is the Wasserstein distance of order two [Villani et al 2009], which is defined as the distance between two probability distribution.

### 3.4 Global sensitivity analysis

For the proposed surrogate model, variance-based sensitivity analysis or so-called Sobol's indices is suitable for performing global sensitivity analysis, which can be computed analytically for a PCE-based surrogate model [Zhu and Sudret 2023]. In the family of Sobol's indices, the QoI-based Sobol's indices are implemented to quantify the effect of input parameters on the model response and exclude the artificial variables introduced in the surrogate.

Since the mean function of the SPCE model is a PCE surrogate model, the corresponding Sobol's indices can be computed directly based on the coefficient of each polynomial term [Sudret 2008]. The first-order index is given by:

$$S_i^m \stackrel{\text{def}}{=} \frac{\text{Var}[\mathbb{E}[\tilde{m}(\mathbf{X}) | X_i]]}{\text{Var}[\tilde{m}(\mathbf{X})]} = \frac{\sum_{\alpha \in A_i} c_\alpha^2}{\sum_{\alpha \in A_i \setminus 0} c_\alpha^2} \quad (3.49)$$

where  $A_i \stackrel{\text{def}}{=} \{\boldsymbol{\alpha} \in A : \alpha_i \neq 0, \alpha_j = 0, \forall j \neq i\}$ .

The higher-order and total Sobol's indices of the mean function are

$$S_u^m = \frac{\sum_{\alpha \in A_u} c_\alpha^2}{\sum_{\alpha \in A_m \setminus 0} c_\alpha^2} \quad (3.50)$$

and

$$S_{T_i}^m = \frac{\sum_{\alpha \in A, \alpha_i \neq 0} c_\alpha^2}{\sum_{\alpha \in A_m \setminus 0} c_\alpha^2} \quad (3.51)$$

, respectively, where  $A_m \stackrel{\text{def}}{=} \{\boldsymbol{\alpha} \in A : \alpha_z = 0\}$  and  $A_u \stackrel{\text{def}}{=} \{\boldsymbol{\alpha} \in A : \alpha_i \neq 0, \alpha_j = 0, \alpha_z = 0, \forall i \in \mathbf{u}, \forall j \in \sim \mathbf{u}\}$

## 4 Serviceability and comfortability analysis

This thesis uses the DIN standards as a reference to evaluate the effect of vibration on humans and structure design. Two standards will be considered mainly. In DIN 4150-2, the impact of human comfortability when exposed to vibration in the building is mentioned; furthermore, the effect of structural vibration is specified in DIN 4150-3.

For the research case, the assumption is made that all the generated vibration signals from the model should be treated as earthquake-induced ground motions, meaning that they belong to the infrequent, short-term vibration in the category of DIN standards.

### 4.1 DIN standards 4150-2

DIN standards 4150-2 evaluate the effect of human exposure on structure vibration, e.g., vibration generated by construction work, earthquake, and traffic.

The vibration type assumed is infrequent, short-term vibration, which fulfills the requirement of no more than three events per day. In the case study, we assume all conditions mentioned in DIN standards are also met; therefore, the evaluation can be performed by approximating evaluation parameters using the unweighted vibration signal. To perform evaluation, weighted velocity  $KB$  and r.m.s value  $KB_{Fmax}^*$  have to be calculated first as following:

$$KB = \frac{1}{\sqrt{2}} \frac{v_{max}}{\sqrt{1 + (\frac{f_0}{f})^2}} \quad (4.1)$$

$$KB_{Fmax}^* = KB \cdot c_F$$

where  $f$  is the frequency, in Hz;  $f_0 = 5.6\text{Hz}$  is the cutoff frequency of high pass filter.  $v_{max}$  is the maximum velocity, in mm/s and  $c_F$  is the constant taken from table 4.1. Then,  $KB_{Fmax}^*$  will be used to compare with the criteria, shown in table 4.2.

**Table 4.1:** Empirical values for constant for different types of vibration

	Description of vibration	$c_F$
1	Sinusoidal vibration with little distortion (e.g., as generated by sawmills at a great distance from the receiver or where there is considerable resonance)	0.9
2	As in line 1, but with greater distortion (i.e., more than about 20%)	0.8
3	Random and periodic vibration with beats: a) with resonance (e.g., as generated by weaving mills, pile driving, or as measured on the vibrating floor of a dwelling)	0.8
	b) without resonance (e.g., as measured on the floor of a dwelling without a basement)	0.7
4	Single events of short duration: a) with resonance	0.8
	b) without resonance	0.6
1.The $c_F$ value should be selected based on the shape of the recorded signal. 2.The $c_F$ values are empirical average values; a deviation of about +15% can be expected.		

**Table 4.2:** Guideline values for evaluating human exposure to vibration in dwellings and similar spaces  
( $A_u$  = lower limit,  $A_o$  = upper limit,  $A_r$  = value for comparison with  $KB$  values)

	Location of building	Day			Night		
		$A_u$	$A_o$	$A_r$	$A_u$	$A_o$	$A_r$
1	Buildings in purely industrial areas, where the only dwellings are intended for plant owners or managers, superintendents, stand-by service personnel, etc.	0.4	6	0.2	0.3	0.6	0.15
2	Buildings in predominantly commercial areas	0.3	6	0.15	0.2	0.4	0.1
3	Buildings in areas which are neither predominantly commercial nor predominantly residential	0.2	5	0.1	0.15	0.3	0.07
4	Buildings in areas which are predominantly or purely residential	0.15	3	0.07	0.1	0.2	0.05
5	Buildings in specially protected areas (such as hospitals) or health resorts	0.1	3	0.05	0.1	0.15	0.05

## 4.2 DIN standards 4150-3

DIN standards 4150-3 evaluate the effect of vibration on building for predominantly static loading. The short-term vibration is assumed in the research case, which belongs to Clause 5. The effect of vibration on structure is evaluated in two parts, which are:

1. Horizontal vibration levels:

The horizontal vibration velocities occurring at the topmost floor are considered when assessing the vibration levels. The maximum velocities in both horizontal directions must follow the requirements in the second column of table 4.3.

2. Vertical vibration levels:

Vertical vibration is measured to assess the vibration on the floor slab. The recorded point should be located where maximum vibration might occur, usually is assumed at the center of the floor slab. The maximum value must also follow the requirements in the third column of table 4.3.

**Table 4.3:** Guideline values for vibration velocity, for evaluating the effect of short-term vibration on structures

	Type of building	Guideline values for $v_i$ in mm/s	
		Topmost floor, horizontal direction, all frequencies, i = x,y	Floor slab, vertical direction, all frequencies
Column Line	1	2	3
1	The buildings used for commercial purposes, industrial buildings, and buildings of similar design	40	20
2	Residential buildings and buildings of similar design or occupancy	15	20
3	Structures that, because of their particular sensitivity to vibration, cannot be classified under Lines 1 and 2 and are of great intrinsic value(e.g., listed buildings )	8	20



### 4.3 Fragility curve

The fragility curve shows the conditional probability that the specified limit state will be exceeded by the engineering demand parameter (EDP). In [Nazri 2018], several different equations for generating the fragility curves are summarized, and its general concept can be expressed as what [Muntasir Billah and Shahria Alam 2015] shows :

$$\text{Fragility} = P[\text{LS} \mid \text{IM} = \mathbf{x}_i] \quad (4.2)$$

where,

- LS** is the limit-state or damage state.
- IM** is the intensity measurement, e.g., Arias intensity, PGA, PGV, and PGD.
- $\mathbf{x}_i$  is the realized condition of **IM**

This thesis will conduct a global sensitivity analysis and a serviceability and comfortability analysis. The **IM** we are interested in are the top two sensitive parameters from the input variables instead of the Arias intensity or PGA mentioned above. Furthermore, the serviceability and comfortability criteria represent **LS** by weighted vibration severity (*KB*) and maximum velocity for DIN 4150-2 and DIN 4150-3, respectively.

## 5 Overall framework of seismic assessment

This thesis aims to analyze whether the geothermally induced earthquake affects residents' daily lives and to study its corresponding significant parameters. To achieve this, the proposed framework is introduced, adapted from [Zhu et al 2023].

### 5.1 Extension from existing framework

In [Zhu et al 2023], seismic fragility assessment is interpreted by the fragility model. Its concept is introduced in section 4.3. The engineering demand parameter (EDP) for the fragility model is the maximum response of the structure under the excitation of artificial ground motion generated from the stochastic ground motion model (SGMM)[Zhu et al 2023], which targets the broad-band excitation that typically relates to far-field ground motion. Furthermore, the relationship between EDP and inputted parameter sets is described by the SPCE surrogate model to accelerate the seismic fragility assessment.

To fulfill the goals of this thesis, extensions and modifications of this existing framework are done, which can be concluded by:

- Since this thesis focuses on serviceability and comfortability assessment, the modifications on the fragility model are first done. Instead of the building's maximum draft ratio or maximum velocity, the EDP selected is based on serviceability and comfortability assessment, e.g., r.m.s weighted velocity. Furthermore, the **IM** are the top two significant parameters.
- This thesis introduces the linear mixed-effect model mentioned in section 2.4 to describe the relationship between seismic parameters and model parameters extracted by our database. In [Zhu et al 2023], the inputted parameter of its proposed SGMM is the model parameter reported by [Broccardo and Dabaghi 2017], which is obtained by

fitting SGMM with recorded ground motion from their database. However, its relationship with the seismic parameters, which can be obtained from seismic stations such as  $M_L$ ,  $R$ , are unknown. In this thesis, the proposed SBGMG can generate ground motion using seismic parameters instead of model parameters. This modification makes the seismic assessment more straightforward since those seismic parameters are commonly used and easy to obtain.

- In [Zhu et al 2023], model parameters for SGMM are fitted from PEER NGA-West2 database, which typically focuses on the moderate-scaled to strong earthquake, i.e.,  $3 \leq M_L \leq 7.9$  [Ancheta et al 2014]. It is not suitable for modeling the geothermally-induced ground motion. Therefore, in the thesis, the database from the Insheim area, Germany, is utilized to better describe the characteristics of such earthquakes. The Insheim area is where one of the geothermal power plants is located in Germany, and its corresponding induced ground motions are monitored and reported in [Küperkoch et al 2018]. Since the database is imported, the thesis will also include extracting model parameters by fitting the recorded ground motion from the database.

## 5.2 Flowchart of the proposed framework

This section will present the overall framework of this thesis for seismic assessment. This framework will be used for a further numerical example and the case study in chapter 6 and 7. The flowchart is also shown in figure 5.1.

1. Extracting the model parameters  $X_a$  by fitting AGMG with recorded ground motions using the method mentioned in section 2.3.5.
2. Training the LME model. The LME model describes the relationship of seismic parameters  $X_{sb}$  obtained from the database and the model parameters  $X_a$  extracted from step 1. This technique is discussed in section 2.4.
3. The LME model trained by step 2 is used to construct SBGMG. Then, the artificial ground motion is generated when input seismic parameters  $X_{sb}$ . Chapter 2 explains the whole concept and detail of SBGMG, where the general procedure is plotted in figure 2.1.
4. The FRFs of the system are prepared based on the inputted building parameters  $X_b$ . In the following numerical example in Chapter 6, the SDOF system is assumed, and its corresponding FRFs can be formed analytically. Additionally, in Chapter 7, the

FRFs can be simulated by performing the harmonic analysis on the FEM model of the buildings excited by unit loading.

5. Once the artificial ground motions from step 3 and FRFs of the building from step 4 are prepared, the responses of the building can be simulated.
6. Simplification and idealization can be considered to reduce the system's complexity. Furthermore, the SPCE is a parametric surrogate model. Its inputted datasets must follow the required probability distribution mentioned in [Sudret 2014].
7. Training the SPCE model. Since the SPCE surrogate model is a non-intrusive method, the preparation of inputted datasets and corresponding responses obtained from step 5 is sufficient for training. The detail of SPCE is introduced in Chapter 3, and its framework is plotted in figure A.6.
8. When the SPCE is trained in step 7, the global sensitivity analysis is conducted using the method in section 3.4. At the same time, the predicted results corresponding to varied significant parameters will be generated after validating the SPCE model.
9. Performing the seismic assessment based on the predicted results and reported significant parameters obtained in step 8. The concept of the fragility model is mentioned in section 4.3. The guideline values for the limit state of serviceability and comfortability analysis are introduced in section 4.1 and 4.2.

## 6 Numerical example

### 6.1 Scenario-based ground motion generator

#### 6.1.1 Analysis of recorded ground motion

Analysis of the ground motions from the database is conducted. In the database, 240 ground motions from 49 events were selected. Detailed information on recorded ground motion is presented in table A.5 in the Appendix A.7. In figure 6.1, the recorded ground motion in the frequency domain are shown, plus the mean function of all recorded ground motion and smoothing versions (window size = 250) are presented in figure 6.2 to give us the first impression of how ground motions look like in the frequency domain and their possible seismic characteristic. As figure 6.2 shows, the shape of the average smoothing curve looks similar to the FRF of the single degree of system applied to the high-pass filter, which matches the physical interpretation of the Hu model. The observation indicates that the natural frequencies of the average result in three directions are located at approximately 10 to 16 Hz, and the cut-off frequency of the high pass filter is lower than the natural frequency, implying that  $r_h \leq 1.0$ . These results have similar patterns as them in [Chen et al 2022].

Following this, the ground motion data is further investigated by characterizing the  $M_L$ ,  $R$ , and  $D$  values. Figure 6.3 shows the histograms of  $M_L$ ,  $R$ ,  $D$  of the 240 recorded ground motions, respectively. It is also worth mentioning that  $v_s$  is treated as a potential parameter describing the site in [Rezaeian 2010; Khansefid et al 2022]; however, research focuses on the ground motions recording in Insheim area only, and the soil in this investigated region assumed to be the same. Therefore,  $v_s$  in this area reported is fixed and will not be considered as one parameter used to fit the modal parameters. To remedy the lack of a parameter to describe the site,  $\omega_g$  is considered the new seismic parameter. In [Chen et al 2022], the comparison of different site characteristic,  $v_s$  and their corresponding fitted  $\omega_g$  are presented, implying the direct relationship between  $v_s$  and  $\omega_g$ . Furthermore,  $\omega_g$  or the parameter with similar physical meaning is also treated as a significant parameter for generating ground

Additionally, the upper limit of searching space for the nonlinear least square method is defined as  $[\omega_g^{upper}, \zeta_g^{upper}, r_h^{upper}] = [300, 1.0, 1.0]$  and the *OptimalityTolerance* is set up as  $10^{-6}$ . After the fitting, the modal parameters  $\omega_g$ ,  $\zeta_g$ ,  $r_h$ ,  $PGA$ ,  $t_d$ ,  $r_{mid}$  are all identified. Due to the Hu model's spectral stationarity assumption and the complete decoupling of time and spectral domain in the proposed AGMG, any segment in the time series should inherit the same spectral property. This feature makes the duration of ground motion  $t_d$  and time information  $r_{mid}$  less important. Hence, only modal parameters  $\omega_g$ ,  $\zeta_g$ ,  $r_h$ ,  $PGA$  are considered to be uncertain, and  $t_d, r_{mid}$  are fixed to reduce the dimension. Figures 6.4 shows the histograms of model parameters in  $z$ -directions, where the histograms for horizontal directions are in Appendix A.5.1. The mean value of  $\omega_g/2\pi$  in the  $z$ -direction is approximately 16.83 Hz,

In the next step, the quantile plot of the data for each parameter versus the corresponding normal quantiles in the  $z$ -direction is shown (Figure 6.5). This process helps us understand whether the normality assumption underlying the regression analysis is fulfilled, which is critical for fitting the LME model presented in the next section. According to the observation, the data within the first and third quartiles (marked by hollow circles) closely follow a straight line in most cases, indicating that the data, to a certain degree, fulfill the normality assumption [Rezaeian and Der Kiureghian 2010]. The worst fit is the damping ratio  $\zeta_g$ , which the curve instead of the line can be observed from the histogram plot in figure 6.4. This implies that the residuals of  $\zeta_g$  after fitting with the LME model might not follow the normal distribution as the regression analysis assumed. The quantile plots and histogram plots for both horizontal directions are shown in the Appendix, where figure A.8, A.9 are histogram plots, and figure A.10, A.11 are quantile plots,

the high p-value for the null hypothesis test. This idealization can be done by removing parameters with less significant coefficients or considering them as constant to reduce the dimension of the problem. Inspired by [Rezaeian and Der Kiureghian 2010], the first trial of predictive equations is formed in equation (6.1).

$$\begin{aligned}
 \ln(PGA) &= \phi_{1,0} + \phi_{1,1} \cdot M_L + \phi_{1,2} \cdot \ln(D) + \phi_{1,3} \cdot \ln(R) + \phi_{1,4} \cdot \omega_g + \hat{\eta}_1 + \hat{\epsilon}_1 \\
 r_h &= \phi_{2,0} + \phi_{2,1} \cdot M_L + \phi_{2,2} \cdot D + \phi_{2,3} \cdot R + \phi_{2,4} \cdot \omega_g + \hat{\eta}_2 + \hat{\epsilon}_2 \\
 \zeta_g &= \phi_{3,0} + \phi_{3,1} \cdot M_L + \phi_{3,2} \cdot D + \phi_{3,3} \cdot R + \phi_{3,4} \cdot \omega_g + \hat{\eta}_3 + \hat{\epsilon}_3
 \end{aligned} \tag{6.1}$$

,where the  $\phi_{i,j}$  is the regression coefficients, and  $\hat{\eta}$ ,  $\hat{\epsilon}$  are the inter-event and intra-event residuals, respectively.  $i = 1 \cdots 3$  denotes the index of modal parameters;  $j = 0 \cdots 3$  denotes the index of seismic parameters.

After training, the estimated coefficients with a 95 percent confidence interval for the model in the  $z$ -direction as the first trial are obtained and shown in the table 6.1. The results for both horizontal directions are presented in Appendix (table A.1 and table A.3).  $\tau$  and  $\sigma$  are the standard deviations of  $\hat{\eta}$  and  $\hat{\epsilon}$ . Additionally, the p-value for the t-test is mentioned. The coefficient is significant if the low p-value, i.e.,  $p\text{-value} \leq 5 \cdot 10^{-2}$ , is shown. On the other hand, a high value implies that the coefficient and its corresponding term might not strongly influence the model [The MathWorks Inc. 2022].

To simplify the regression model, the seismic parameters with less significance from the t-test are filtered out. Table 6.1 shows that the coefficients for  $D$  and  $\omega_g$  report high p-values for fitting  $\ln(PGA)$ ; therefore, they are ignored in the second trial. Furthermore,  $M_L$  and  $D$  are removed from fitting  $r_h$ . For  $\zeta_g$ , it is observed that none of the coefficients provide a sufficient small p-value to pass the t-test. We assume it has the same predictive equation as  $r_h$  for the second trial in this case. Finally, since  $M_L$  is removed from the predictive equation for  $r_h$  and  $\zeta_g$ , inter-event residuals  $\hat{\epsilon}$  and the intra-event residuals  $\hat{\eta}$  will not be considered separately; therefore they are replaced by the overall residuals, re-used the symbol  $\hat{\eta}$ . The predictive equations for the second trial are formed in equation (6.2).

$$\begin{aligned}
 \ln(PGA) &= \phi_{1,0} + \phi_{1,1} \cdot M_L + \phi_{1,2} \cdot \ln(R) + \hat{\eta}_1 + \hat{\epsilon}_1 \\
 r_h &= \phi_{2,0} + \phi_{2,1} \cdot R + \phi_{2,2} \cdot \omega_g + \hat{\eta}_2 + \hat{\epsilon}_2 \\
 \zeta_g &= \phi_{3,0} + \phi_{3,1} \cdot R + \phi_{3,2} \cdot \omega_g + \hat{\eta}_3 + \hat{\epsilon}_3
 \end{aligned} \tag{6.2}$$

**Table 6.1:** Result of fitting Linear Mixed-Effects Models in the  $z$ -direction as first trial

i	$\ln(PGA)$				$r_h$				$\zeta_g$			
	Estimate	Lower	Upper	P-value	Estimate	Lower	Upper	P-value	Estimate	Lower	Upper	P-value
0	-7.4019	-8.4524	-6.3514	2.67E-32	0.88157	0.71994	1.0432	4.0186E-22	0.29097	0.11635	0.46558	1.19E-03
1	1.8931	1.5616	2.2247	1.03E-23	0.008324	-0.05452	0.071163	0.79433	0.034045	-0.03603	0.10412	0.33946
2	0.39488	-0.11395	0.90372	0.12762	0.010351	-0.00853	0.029229	0.28111	0.005704	-0.01502	0.026425	0.58811
3	-1.4736	-1.7934	-1.1538	4.87E-17	-0.02448	-0.03576	-0.01319	0.0000281	-0.00683	-0.01927	0.005602	2.80E-01
4	0.022402	-0.00153	0.046335	0.06642	-0.01792	-0.02275	-0.01309	4.2912E-12	0.000893	-0.00436	0.006147	7.38E-01
$\tau$	0.30669				0.082374				0.091359			
$\sigma$	0.66712				0.16994				0.19919			



Table 6.2: Result of fitting Linear Mixed-Effects Models in  $z$ -direction as second trial

	$\ln(PGA)$				$\tau_h$				$\zeta_g$			
i	Estimate	Lower	Upper	P-value	Estimate	Lower	Upper	P-value	Estimate	Lower	Upper	P-value
0	-6.3491	-7.1384	-5.5598	6.45E-39	0.93243	0.78982	1.075	4.70E-29	0.32943	0.17624	0.48262	3.26E-05
1	2.0144	1.6972	2.3315	7.59E-28	-0.02432	-0.03579	-0.01286	4.14E-05	-0.00617	-0.01854	0.006195	0.32647
2	-1.5591	-1.9142	-1.204	8.32E-16	-0.01772	-0.02253	-0.01291	5.90E-12	0.001475	-0.00382	0.006766	0.58327
$\tau$	0.28932				$\sigma = 0.17555$				$\sigma = 0.20352$			
$\sigma$	0.68444											

### 6.1.3 Response spectral analysis - SBGMG Validation

The comparison of recorded ground motion and the one generated by SBGMG is conducted in this section. First, one recorded ground motion is selected randomly, where its seismic parameters, i.e.,  $M_L, R, \omega_g, \zeta_g$ , are known. Those seismic parameters are inputted into the trained SBGMG to generate the artificial ground motion. Once the artificial ground motion data and its corresponding record are prepared, response spectral analysis is performed. In the numerical example, the damping ratio of the system is fixed at 5 percent, and the response period sweeps from 0.01 sec to 10 sec. Additionally, SBGMG is a stochastic simulator; the realization shows different amplitudes but a similar pattern when inputting the same seismic parameter. In figure 6.9, It is observed that the recorded and artificial ground motion have a similar pattern. Both of them have peaks located at approximately 0.1 sec, and the amplitude of the recorded one is in between the upper and lower bound of the artificial ground motions. This result is expected since SBGMG includes the LME model, where the recorded ground motion can be treated as the mean function of this regression model, and the uncertainty of amplitude is propagated from the total error terms of LME, computed from inter and intra-events residuals. In the Appendix A.5.5, another two examples of response spectral analysis are shown in figure A.18 and figure A.19

first type of idealization is that those input data might not fit well with the assumed distribution (the high p-value fails to reject the null hypothesis that the data comes from a population with the assumed distribution at the default 5 percent significance level) but have similar shape compared to the corresponding histogram. For this type of idealization, once the distribution type is chosen, the mean and standard deviation will be computed by the distribution-fitting function in MATLAB. The idealization of parameters  $M_L$  and  $\omega_g$  belong to this type.

The second type of idealization is assumed because it allows us to investigate how uncertainty propagates in the system better. For example, it makes sense to assume that the rupture distance in both horizontal directions  $R_x$  and  $R_y$  follow uniform distribution since the distance between the source and the target building is unknown. Plus, assuming the source occurs at the center and the wave propagates outward as a ripple fits the engineering perspective of wave propagation. The histogram of the idealized rupture distance  $R = \sqrt{R_x^2 + R_y^2}$  is shown in figure 6.10, which give us an impression of the synthetic probability distribution will be inputted for training surrogate model.

The last type of idealization is for dimension reduction. The damping ratio of ground motion  $\zeta_g$  is usually assumed to be constant. This simplification has been done in [Zhu et al 2023]. Moreover, the low significance in global sensitivity analysis for the damping ratio is reported in [Abbiati et al 2021]. After simplification, the experimental design will generate the sample points following those idealized distributions and assumptions. Figure 6.11 and 6.12 show the fitting results of  $M_L$  and  $\omega_g$ , respectively. In addition,  $\omega_{sdo}$  and  $\zeta_{sdo}$  in the numerical example are assumed to follow log-normal distributions, which fulfill the constraint that both parameters have to be positive. Their distribution information is shown in table 6.4. Finally, the idealized input parameters for SBGMG are summarized in table 6.5.

**Table 6.4:** Parameters of SDOF system

SDOF system	
Natural frequency of system, $\omega_{sdo f}/2\pi$ [Hz]	$\sim \mathcal{N}(12, 1)$
Damping ratio of system, $\zeta_{sdo f}$ [-]	$\sim \mathcal{LN}(-3.2, 0.1)$

**Table 6.5:** Parameters of scenario-based ground motion generator

SBGMG	
Moment magnitude, $M_L$ [-]	$\sim \mathcal{LN}(-0.39, 0.56)$
Rupture distance in $x$ -direction, $R_x$ [km]	$\sim \mathcal{U}(8, 7.5)$
Rupture distance in $y$ -direction, $R_y$ [km]	$\sim \mathcal{U}(8, 7.5)$
Predominant frequency of GM in $x$ -direction, $\omega_{g,x}/2\pi$ [Hz]	$\sim \mathcal{LN}(2.47, 0.33)$
Predominant frequency of GM in $y$ -direction, $\omega_{g,y}/2\pi$ [Hz]	$\sim \mathcal{LN}(2.48, 0.35)$
Predominant frequency of GM in $z$ -direction, $\omega_{g,z}/2\pi$ [Hz]	$\sim \mathcal{LN}(2.76, 0.39)$
Damping ratio of soil, $\zeta_g$ [-]	0.31

## 6.2.2 Training result and verification

### The effect of baseline correction and simulation result

To expedite data generation, we will not directly perform numerical integration, such as the Newmark-beta method, in the time domain. Instead, we analytically compute the Frequency Response Function (FRF) of a single Degree of Freedom (DOF) system,  $\text{FRF}_{sdo f}$ , which is formed in equation (6.3).

$$\begin{aligned}
 \ddot{u}_{resp}(\Omega) &= \frac{-i2\Omega\zeta_{sdo f}\omega_{sdo f} - \omega_{sdo f}^2}{(\Omega^2 - \omega_{sdo f}^2) - i2\Omega\zeta_{sdo f}\omega_{sdo f}} \cdot \ddot{U}_g(\Omega) \\
 &= \text{FRF}_{sdo f}(\Omega) \cdot \ddot{U}_g(\Omega)
 \end{aligned} \tag{6.3}$$

where  $\ddot{u}_{resp}$  is the response of the building, and  $\ddot{U}_g$  is the excitation.

The FRF will be multiplied by the ground motion in the frequency domain, and the result will then be transformed back to the time domain. This approach is equivalent to conducting a transient analysis of the structure under the same excitation, assuming the system is linear. Once the ground motion is generated, the baseline correction introduced in section 2.5 must then be applied to remove the possible unwanted trend mentioned in section 2.5 (figure

**Table 6.6:** Global and local  $\varepsilon_{val}$ , where  $\mathbb{E}[\varepsilon_{val}]$  are the mean value of validation error and  $\sigma$  are the standard deviation. The size of prediction  $N = 500$  and the validation error are collected by repeating 30 runs

	$\mathbb{E}[\varepsilon_{val}]$	$\sigma$
$\varepsilon_{val}$ , where $\mathbf{X} = \mathbf{X}_{sb}$	0.02279	0.0068
$\varepsilon_{val}$ , where $\mathbf{X} = X_{sb,1}$	0.00455	0.00212
$\varepsilon_{val}$ , where $\mathbf{X} = X_{sb,2}$	0.03136	0.00943
$\varepsilon_{val}$ , where $\mathbf{X} = X_{sb,3}$	0.00772	0.00342
$\varepsilon_{val}$ , where $\mathbf{X} = X_{sb,4}$	0.04297	0.01409

### 6.2.3 Convergence study of error

For the convergence study, we test the accuracy of SPCE by varying the training data size from 100 to 1600, shown in figure 6.18. Three surrogate models are trained independently with respect to each training data size. The black-hollow circles represent the average result of 90 repetitions from running each of the three models 30 times. The upper and lower dots denote the maximum and minimum error measured, respectively. Additionally, the edges of the boxes are the 25 and 75 percentiles of the results. The median is shown as the horizontal red line inside the box. It is observed that the error decreases as long as the number of training data increases, which matches the common sense that better prediction can be assumed when finer exploitation is performed in the design space.

**Table 6.7:** Computer specification for training SCPE model

Computer specification for training SCPE model	
CPU	AMD Ryzen 9 3900XT 12-Core Processor
CPU speed	3.80 GHz
RAM	32.0 GB
OS version	Windows 10 Professional 64-bit

### Result of time comparison

In this section, the time comparisons focus only on the training process. The data-generating process won't be counted due to its dependency on the problem. The training time will be presented in the logarithm scale, shown in blue on the left y-label. On the other hand, the number of SPCE coefficients selected by the solver is shown in orange on the right y-label. The time comparison using different numbers of training data is shown in figure 6.19. In this case,  $4^{th}$  order of polynomial and  $q$ -norms = 1 are set. Following this, the time consumed based on different combinations of polynomial order and  $q$ -norms is tested in figure 6.20, where the number of training data is fixed at 800. The dots in the figure denote the average value of 3 repetitions of training time, where the upper and lower bound of the error bar represented the maximum and minimum time selected from repetitions, respectively. The results show that the fundamental reasons affecting the computation time are the number of coefficients and the training data size. The number of coefficients is also controlled by the polynomial degrees and  $q$ -norm selected. The computational time increases when the higher-order polynomial is chosen, or high  $q$ -norms are set, even though the hybrid-LARS solver filters out insignificant terms of a polynomial.

Further, it is also interesting to observe in figure 6.19 that the number of coefficients rises slightly from approximately 130 to 150 when the training size increases,  $N = \{100, \dots, 1600\}$ , despite the polynomial degree and  $q$ -norm being fixed. This implies that the LAR algorithm might treat more polynomial terms as significant in the hybrid-LARS solver. Figure 6.21 shows the time comparison of running in a single core and six cores, where the number of training data is 400. Running the algorithm in multi-cores indeed significantly improves training time, which drops from approximately 6.8 hours to 1.8 hours when six cores instead of 1 core is used in the example. Experiments were run on a computer with specifications listed in table 6.7.

## 7 Case study: Geothermal energy site in Taufkirchen

In this chapter, a case study is conducted. The global sensitivity analysis, serviceability, and comfortability assessment are the following. The ground motions are generated by SBGMG, and the example building model is the geothermal energy site in Taufkirchen, Munich, Germany.

### 7.1 Building description

#### 7.1.1 Building parameters

The geometry of the building (figure 7.1) and the parameters (table 7.1, 7.2) are shown in this section. The building images from different perspectives are in Appendix (figure A.20 to A.22).

The maximum length and width of the building are 18.78 m and 19.18 m, respectively. The height of each floor is fixed at 3 m. The cross sections of both beams and columns are set as  $0.3 \times 0.3$  m, where each floor slab and wall thickness is 0.2 m and 0.1 m. Additionally, the size of the foundation  $l_f$  and  $b_f$  are assumed to be 0.75 m. The soil-structure intersection effect is modeled by the lumped-parameter model (LPM). For the element choices, both beams and columns are modeled by the BEAM189. Slabs and walls are built by SHELL181. Those elements are named by the ANSYS®.

In the case study, we assumed the stiffness of the beam and the modal damping ratio of the whole system are uncertain. The building and soil properties are all stated in table 7.1 and table 7.2, respectively. Additionally, the input variables for SBGMG follow the same distribution as table 6.4 in section 6.2.1.

### 7.1.2 Soil-structure interaction

The frequency-independent LPM is introduced to model the interaction between soil and structure. In the case study, we assumed the foundation is on the surface, and the soil is a homogeneous half-space, where the shear wave velocity and soil density are fixed. The static stiffness and dimensionless coefficients of the LPM for homogeneous half-space are stated in the table 7.3, where  $G$ ,  $\nu$ ,  $\rho$  are the shear modulus, Poisson's ratio, and mass density of the soil.  $r_0 = \sqrt{\frac{4l_fb_f}{\pi}}$  is the radius of mass if the rectangle foundation with length  $l_f$  and width  $b_f$  is assumed. The dashpot coefficient  $C$  and mass  $M$  are formed as following [Wolf 1994]:

$$\begin{aligned} C &= \frac{r_0}{v_s} \gamma K \\ M &= \frac{r_0^2}{v_s^2} \mu K \end{aligned} \tag{7.1}$$

where  $v_s = \sqrt{\frac{G}{\rho}}$  is the shear-wave velocity.

Plus, their coupled effects must be considered to build not only the LPM model in horizontal and vertical directions but also the one for rocking and torsional. Conventionally, when writing the APDL in ANSYS®, the COMBIN14 element is practical to model the LPM in the general case for a decoupled system, but it is complex to present the coupling effect in our research case; therefore, the MATRIX27 element is introduced, which is formed by a matrix of two nodes with 6 DOF each [ANSYS 2010]. The detail of LPM implementation is mentioned in [Kao and Pflieger 2023]

**Table 7.3:** Static stiffness and dimensionless coefficients of the LPM [Wolf 1994]

	Static Stiffness $K$	Dashpot $\gamma$	Mass $\mu$
Horizontal	$\frac{8Gr_0}{2-\nu}$	0.58	0.095
Vertical	$\frac{4Gr_0}{1-\nu}$	0.85	0.27
Rocking	$\frac{8Gr_0^3}{3(1-\nu)}$	$\frac{0.3}{1 + \frac{3(1-\nu)I_x}{8r_0^5\rho}}$	0.24
Torsional	$\frac{16Gr_0^3}{3}$	$\frac{0.433}{1 + \frac{2I_z}{r_0^5\rho}} \sqrt{\frac{I_z}{r_0^5\rho}}$	0.045



## 7.2 Simulation and training result

This section presents the simulation result and the SPCE model built for further assessment. The building's FRFs are simulated using the Harmonic analysis solver in ANSYS®. The coordination of the target points is located at  $[8.75, 4.8, Z]$ , where  $Z$  corresponds to the height of each floor. The  $x$  and  $y$  coordinates are shown in figure 7.2. After this, the same procedure mentioned in Chapter 5 is performed to generate the data for training. The overall flowchart (figure A.7) is presented in Appendix A.4. In our research case, the maximum velocities of the structure under excitation in all three directions must be considered according to the DIN 4150-3 (table 4.3). The surrogate model for each direction will be trained and evaluated individually. Each simulation required 8 minutes to run. In the research case, 1000 realizations are generated, splitting into 500 training and 500 validation datasets. The total simulation time is approximately 8000 minutes based on the computer specification in table 7.4. Following this, the 3D and 2D predictive responses in the  $z$ -direction (figure 7.5 and 7.6) against  $M_L$  and  $R$  are presented, while the predictive responses for horizontal directions are in Appendix (figure A.23 to A.26). Finally, the global PDFs of SPCE prediction compared to the validation set, similar to figure 6.16, are shown again for both three directions, and the global validation error will also be reported, where the  $\varepsilon_{val}$  should less than 0.05.

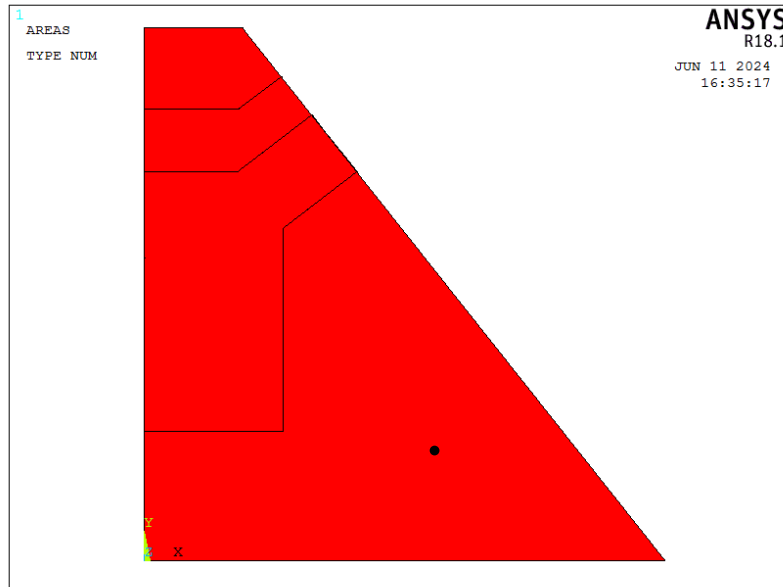
According to DIN 4150-3, the following evaluation is focused on the topmost floor for both the horizontal direction and the floor slab on each floor for the vertical direction. Therefore, their FRFs simulated by ANSYS® are presented here. In figure 7.3, The building's FRFs in the  $z$ -direction show that the first natural frequency of the building is approximately 24 Hz; other peaks are located at a higher frequency range, which is less significant to earthquake-like excitation. The maximum amplification of FRFs corresponds to the topmost floor in the vertical direction; the same patterns are also found on each floor. Hence, the assessments are performed on the topmost floor for all three directions.

Furthermore, in figure 7.4, the FRFs in  $y$ -directions also have only one significant peak at the lower frequency range. The natural frequency of FRFs in the  $y$  direction is approximately 24 Hz, similar to FRFs in the  $z$  direction. On the other hand, two peaks are shown in the lower frequency range of FRFs in the  $x$ -direction, in which the first and second modes occur at approximately 10 and 24 Hz, respectively. It can also be observed that the peaks at 24 Hz are presented in all three directions, possibly due to the coupling effect between each direction.

Additionally, The response surfaces in figure 7.5 and 7.6 give us the apparent relationship between  $v_{max}$  and seismic parameters  $M_L$  and  $R$ . The velocity in the logarithm scale is

positively proportional to  $M_L$  while negatively proportional to  $R$ . This result is expected as it inherits the property from the predictive equation (equation (6.2)) in SBGMG. In figure 7.5, it is also interesting to observe that the gradient along  $M_L$  is larger than the one along  $R$ , meaning that the  $v_{max}$  highly depends on the magnitude of  $M_L$  compared to  $R$ . This phenomenon can also be confirmed by the global sensitivity analysis in the next section, which shows that the significance of  $M_L$  should be relatively more considerable than  $R$ . Furthermore, the black-mesh surface are formed by the standard deviation of realization generated by 100 replications at each mesh point. It is expected that the result of replication at each mesh point (fixed  $M_L$  and  $R$ ) are different but following the Gaussian distribution with a specific standard deviation (from each upper mesh point to the lower one) since the system is a stochastic simulator and the normality assumption is made for SPCE model.

Finally, figure 7.7 provides us with a comparison of prediction and simulation results for validation, where the PDFs of SPCE prediction in all three directions are pretty matched with the histograms of simulation results, plus the mean validation error shown in table 7.5 are all smaller than 0.05, which indicate the trained models have satisfactory accuracy.



**Figure 7.2:** Top view of the building, the black lines inside the red area are the placement of walls. The black dot is the target point recording the FRFs, whose coordinate is [8.75, 4.8] (m) for x and y directions; the z coordinates correspond to the height of each floor.

## 7.3 Sensitivity analysis

A GSA is conducted to understand which parameter might strongly influence the system as introduced in section 3.4. The variance-based technique is used, which determines each parameter's sensitivity by examining how large the variance will be obtained from the system while fixing other parameters. Namely, the highly sensitive parameter dominates the system, and the outcome might vary severely depending on it. Inheriting the property of PCE, the SPCE can also determine the sensitivities by the coefficients of each polynomial term. Figure 7.8 shows the results of the top 5 significant parameters in three directions, and the high-order index will also be considered.

It is evident that  $M_L$  is the most significant parameter to dominate the system since it is directly related to  $PGA$  in  $AGMG$ . For  $R_x$  and  $R_y$ , it is observed that they have similar significance; this makes sense because  $R$  in  $AGMG$  is obtained by  $R = \sqrt{R_x^2 + R_y^2}$  and if the distribution of both  $R_x$  and  $R_y$  are the same, their contribution to the system should also be the same. It is interesting to observe this result, as it implies that the SPCE algorithm can capture the uncertainty propagated from individual parameters and the synthesization of two parameters.

Regarding the parameter  $\omega_g$ , it has a similar significance compared to  $R_x$  and  $R_y$  in both horizontal directions. It even exceeds them, becoming the second important one in the  $y$ -direction.  $\omega_g$  controls the predominant frequency of ground motion and can considerably impact vibration if it is close to the natural frequency of the building; therefore, its importance cannot be ignored. This outcome also matches the result presented in [Zhu et al 2023] that  $\omega_g$  is treated as the second important parameter if  $R$  is not considered.

Next, the potential reasons why the significance of  $\omega_g$  is higher than  $R$  in the  $y$ -direction and relatively low in the  $z$ -direction are discussed. As mentioned earlier, the large significance means large differences when varying that specific parameter. However, this change might also be influenced by the system's amplification scale of the FRF in each direction.

Although the patterns of FRFs in  $y$ -direction (figure 7.4) and in  $z$ -directions (figure 7.3) are similar in the low-frequency range, their amplitudes at natural frequency differ significantly, approximately 15 and 6, respectively. This implies that the gradient of response  $\frac{\partial v_{max}}{\partial \omega_g}$  is different in these two directions when sweeping  $\omega_g$  from less important frequency to near the building's natural frequency. In other words, the significance of  $\omega_g$  in  $y$ -direction should be approximately double than it in the  $z$ -direction assuming similar distributions of  $\omega_{g,y}$  and  $\omega_{g,z}$  in this case study due to the ratio of FRF in both directions being two. This phenomenon

can also be observed in figure 7.8. Hence, the potential reason for the high significance of  $\omega_g$  in the  $x$  and  $y$  directions but not in the  $z$ -direction compared to  $R$ , is due to the different amplification scales. Nevertheless, in the predictive equation of SBGMG (equation (6.2)),  $\omega_g$  has no contribution to  $PGA$ , but  $R$  has, meaning the change of  $R$  should influence the amplitude of  $v_{max}$  more directly, while the change of  $\omega_g$  can only affect the amplification ratio implicitly. Plus, the significance of  $R_x$  and  $R_y$  reported in  $y$ -directions are still high and cannot be ignored. Considering the above points, importance of  $R$  should still be considered before  $\omega_g$ . Therefore, further serviceability and comfortability assessments are still based on  $M_L$  and  $R$  for all three directions.

the horizontal and vertical direction in the frequency range 0 to 100 Hz for DIN 4150-3, respectively. The highest probability for both assessments among all three directions will then be reported. To present the serviceability and comfortability assessment by fragility curve, we use the concept introduced in 4.3. The  $\mathbf{x}_i, i = 1 \dots 2$  in equation (4.2) are the top two significant seismic parameters from  $\mathbf{X}_{sb}$ , and the probability of whether exceeding the limit state **LS** is shown in the z-axis. 100 replications of SCPE prediction compute the probability at each specific  $x_1$  and  $x_2$ , that  $p_f = P[Y \leq LS \mid X_{sb} = x_1, x_2]$ . For instance, in figure 7.9, let's repeat the prediction procedure at  $M_L = 2.4, R = 0.5$  100 times; if there are 80 realizations over the limit state,  $A_0 = 3$ , then the probability  $p_f$  is 0.8, which is the value at z-axis.

As a result (figure 7.9, 7.10), for DIN-4150-2, if  $M_L$  is less than 1.5, there is little chance of violating the standard requirement in the daytime. Over 50 percent of the probability of violation occurs only when  $M_L$  is larger than 2, and the target building is close to the epicenter,  $R \leq 6$  km. However, it shows different results at night (figure 7.11, 7.12). Unless the target is far from the epicenter,  $R \geq 8$  km and  $M_L \leq 0.5$ , there is a chance that it will fail to pass the DIN-4150-2 criteria. Most ground motion events at night will make humans uncomfortable, especially if  $M_L \geq 1.5$ . This makes sense since the strict criteria that  $A_0 = 0.2$  for the purely residential area at night is defined. Fortunately, a high proportion (approximately 75 percent) of the events recorded in the Insheim area have  $M_L$  less than 1 (figure 6.3); therefore, it might not as frequently as expected that residents feel uncomfortable due to the ground motion at night. Finally, the requirement of DIN-4150-3 is evaluated. As observed, the results in horizontal directions are more likely to violate the requirement probability due to the lower value of the limit state compared to the one in vertical direction being defined. The evaluation of the floor slab (vertical direction) shows the same pattern as a horizontal direction but less probability. In figure 7.13 and 7.14, the fragility curve shows that there is only a low probability that the response will violate the requirement. Even the highest probability, located at  $M_L = 2.4$  and  $R = 0.5$  km, has only a 30 percent chance of exceeding the serviceability criteria, implying that this type of ground motion might not severely influence the structure.

## 8 Conclusion and Outlook

### 8.1 Conclusion

This thesis presents a framework for performing seismic risk assessment for the structure excited by the minor to moderate scaled geothermal-induced ground motion, i.e.,  $M_L \leq 3.0$ . The approach of generating the ground motion with stochastic behavior, based on the target ground motion database by the seismic parameters, is also mentioned. Additionally, a computationally efficient surrogate model that can capture the uncertainty from a stochastic ground motion simulator is introduced to better interpret the relationship between the damage state and seismic parameter, which avoids the large number of high computational cost simulations.

In the numeral example, the recorded ground motion is thoroughly investigated and characterized by model parameters first, and those parameters are linked with the seismic parameter by the LME model. Following this, response spectral analysis is performed to validate the proposed ground motion generator. Plus, the performance of the surrogate model is also tested in terms of validation error and required training time.

Once the validations are complete, the case study is conducted on the target building structure with uncertain material properties. In the case study, the global sensitivity analysis reflects the top two significant seismic parameters that might cause adverse effects on the structure or residents. Finally, the serviceability and comfortability analysis of the structure is presented in the manner of fragility curves.

As the result of the thesis, the conclusions are stated:

- The SBGMB can generate artificial ground motion with characteristics and patterns similar to recorded data, even when accounting for stochasticity, provided that it is trained with a sufficient number of ground motions.

- The accuracy of the SPCE model is positively proportional to  $p$  and  $N$ . In this thesis, it is observed that satisfactory accuracy, i.e.,  $\mathbb{E}[\epsilon_{val}] \leq 0.05$ , could be reached when  $p \geq 3$ , if  $q_{norm} = 1$  and  $N \geq 400$  for the current problem.
- The computational time for training the SPCE model is negatively proportional to  $p$  and  $N$ , which is in contrast to the accuracy of the SPCE model. An appropriate  $q_{norm}$ , i.e.,  $0.7 \leq q_{norm} \leq 1.0$  can be selected to shrink the computational time without sacrificing too much accuracy. Another approach for reducing computational time is training the model with parallel programming techniques. According to the computer specifications stated in table 6.7, the time for training in 6 cores parallelly is only approximately 30 percent of using a single core.
- In the research case,  $M_L$  and  $R$  should be the top two most significant seismic parameters influencing the response. In the horizontal directions, the significance of the predominant frequencies  $\omega_{g,i}$ , where  $i = x, y$  are also noticeable. However, since these frequencies have less impact on the amplitude of ground motion, the primary focus will remain on  $M_L$  and  $R$ .
- The result shows that the weak to moderate-scaled earthquake might affect comfortability, mainly when the ground motion occurs at night near a purely residential area, according to DIN 4150-2 criteria.
- The result shows that the weak to moderate-scaled earthquake has less impact on serviceability. The highest possibility of a 30 percent chance of exceeding the serviceability criteria occurs only when larger  $M_L$  and closer  $R$ , i.e.,  $M_L \geq 2.4$  and  $R \leq 0.5$  according to DIN 4150-3 criteria.

## 8.2 Outlook

Although the proposed framework and approaches allow us to perform the seismic risk analysis efficiently and with satisfactory accuracy, a few points could be improved to present a more full-scale risk assessment.

- Currently, spectral stationarity is assumed in the SBGMG, which is valid for evaluating weak to moderate-scaled ground motion; however, this idealization can be improved by the spectral non-stationary ground motion model to have a more realistic representation.

- To avoid large amounts of input parameters, which might cause the curse of dimensionality, several simplifications on SBGMG and SPCE are done. To have a more realistic evaluation, one might input the whole seismic and building parameters considered, but applying PCA or diffusion map as in [Zhong et al 2023] first to reduce the dimension for forming the surrogate model and transfer it back to total dimension space. With this approach, less simplification has to be done, and a more complex relationship in the system can be presented.
- Several idealizations are done when assuming the probability distribution of seismic parameters. The potential improvement is using the non-parametric probability-based surrogate model instead of the parametric one.
- Since the focus is on performing the serviceability and comfortability analysis for weak to moderate-scaled ground motion, it is valid to assume a linear vibration system for simulation. However, several conventional seismic risk indices are evaluated by non-linear techniques, e.g., pushover analysis, non-linear time history analysis, and incremental dynamic analysis. Hence, a better, full-scaled assessment should include them.
- To make the simulation more flexible, the FRFs of the building are obtained first by harmonic analysis. Then, the structure's response is received by exciting the ground motions on FRF using the convolution technique. However, the location selection for extracting FRFs is based on one's understanding of structure and experience. In practice, the locations are usually chosen in the middle of each floor slab for vertical directions or at the topmost floor for horizontal directions. However, if one wants to have a precise analysis, it is recommended to evaluate the vibration of the whole structure instead.
- Finally, the computational cost of training the SPCE model is still too high. Compared to the original SPCE model [Zhu and Sudret 2023], there are already some improvements to reduce the training time done in the thesis; for instance, in the self-written SPCE algorithm, the parallel programming technique is implemented for the Gaussian process; plus, instead of conventionally BFGS optimization algorithm, the efficient substitution, L-BFGS-B algorithm, is utilized to accelerate the procedure estimating the coefficients of the SPCE model; however there is still the room for improving the training time.



# A Appendix

## A.1 Software

### A.1.1 MATLAB

*Matlab* is a commercial software of Mathworks used to solve mathematical problems and graphically display solutions. Several toolboxes for data analysis, optimization, and machine learning are included in MATLAB.

Internet: <https://www.mathworks.com/products/matlab.html>

### A.1.2 PYTHON

*Python* is a high-level programming language. Several advanced packages written in Python can be utilized for data analysis, optimization, and machine learning, such as Numpy, Scikit-optimize, and Scikit-learn.

Internet: <https://www.python.org/>

### A.1.3 ANSYS®

*ANSYS®* is a finite element analysis software from Ansys Inc. It can perform structural analysis using different solvers, including linear dynamics, nonlinearities, thermal analysis, materials, composites, hydrodynamic, explicit, harmonic, and more. Here, the main solver used for the thesis is the Harmonic analysis technique provided in ANSYS mechanical.

Internet: <https://www.ansys.com/>

Table A.1 : Result of fitting Linear Mixed-Effects Models in x-direction as first trial

	$ln(PGA)$				$r_h$				$\zeta_g$			
i	Estimate	Lower	Upper	P-value	Estimate	Lower	Upper	P-value	Estimate	Lower	Upper	P-value
0	-7.772	-8.7229	-6.8211	3.50E-40	0.61835	0.49445	0.74225	2.04E-19	0.47799	0.32173	0.63426	6.21E-09
1	1.811	1.4883	2.1337	2.82E-23	-0.02849	-0.08516	0.028184	0.32306	-0.01951	-0.084254	0.045242	0.55345
2	0.45008	-0.05665	0.95682	0.081456	0.001633	-0.01358	0.018849	0.85195	0.000981	-0.01856	0.020521	0.92135
3	-1.5546	-1.7748	-1.3344	9.89E-33	-0.00352	-0.01136	0.0043216	0.37761	-0.00493	-0.014543	0.004685	0.31354
4	0.025757	-0.00917	0.060687	0.14766	-0.01804	-0.0244	-0.011676	6.29E-08	-0.00558	-0.014278	0.003114	0.2073
$\tau$	0.26926				0.068242				0.075759			
$\sigma$	0.67067				0.16477				0.17426			

**Table A.2:** Result of fitting Linear Mixed-Effects Models in y-direction as first trial

i	$\ln(PGA)$				$r_h$				$\zeta_g$			
	Estimate	Lower	Upper	P-value	Estimate	Lower	Upper	P-value	Estimate	Lower	Upper	P-value
0	-8.1509	-9.047	-7.2548	8.69E-46	0.71381	0.55983	0.86779	3.24E-17	0.3725	0.21943	0.52557	2.90E-06
1	1.8922	1.598	2.1864	2.29E-28	-0.0309	-0.09663	0.034817	0.35518	0.003303	-0.05082	0.057431	0.9044
2	0.36224	-0.11198	0.83646	0.13369	0.006697	-0.01458	0.027976	0.53585	0.025055	0.007375	0.042735	0.005672
3	-1.3023	-1.5644	-1.0403	3.43E-19	-0.01351	-0.02304	-0.00399	0.005624	-0.01034	-0.01948	-0.0012	0.026841
4	0.029193	0.00362	0.054766	0.025442	-0.01794	-0.02429	-0.01159	7.05E-08	-0.00671	-0.01479	0.001374	0.10334
$\tau$			0.27207				0.092712				0.078913	
$\sigma$			0.69972				0.18564				0.15723	

Table A.3: Result of fitting Linear Mixed-Effects Models in x-direction as second trial

	$\ln(PGA)$				$r_h$				$\zeta_g$			
	Estimate	Lower	Upper	P-value	Estimate	Lower	Upper	P-value	Estimate	Lower	Upper	P-value
i												
0	-6.3491	-7.4325	-6.3488	1.63E-69	0.60956	0.50904	0.71007	3.20E-26	0.47206	0.17624	0.33754	4.11E-11
1	2.0144	1.6755	2.2927	1.31E-28	-0.0042	-0.0118	0.0034022	0.27756	-0.00544	-0.018541	-0.01481	0.25447
2	-1.5591	-1.8132	-1.3216	2.95E-28	-0.01808	-0.02443	-0.011733	5.42E-08	-0.00566	-0.0038154	-0.0144	0.20262
$\tau$	0.29263				$\sigma = 0.16962$				$\sigma = 0.18544$			
$\sigma$	0.72878											

**Table A.4:** Result of fitting Linear Mixed-Effects Models in y-direction as second trial

i	$\ln(PGA)$				$r_h$				$\zeta_g$			
	Estimate	Lower	Upper	P-value	Estimate	Lower	Upper	P-value	Estimate	Lower	Upper	P-value
0	-7.1432	-7.7302	-6.5562	3.51E-65	0.72623	0.60103	0.85144	2.33E-24	0.47865	0.3375	0.61979	1.69E-10
1	1.9389	1.6526	2.2253	1.27E-30	-0.01375	-0.02321	-0.00428	0.004598	-0.00873	-0.01826	0.000793	0.07218
2	-1.3672	-1.6432	-1.0912	4.04E-19	-0.01826	-0.02455	-0.01198	3.10E-08	-0.00648	-0.01483	0.001877	0.12802
$\tau$	0.27542											
$\sigma$	0.70552				$\sigma = 0.19033$				$\sigma = 0.16633$			

## A.7 Database of recorded ground motion

**Table A.5:** Table of recorded ground motion for database

Index	Event	Station	Time	$M_L$	D	R
1	1	TMO55	2023.01.11-13:55:29	0.7	5	7.3035
2	1	TMO22	2023.01.11-13:55:29	0.7	5	5.1721
3	1	TMO20	2023.01.11-13:55:29	0.7	5	8.2915
4	1	INS8	2023.01.11-13:55:29	0.7	5	10.3534
5	1	INS1	2023.01.11-13:55:29	0.7	5	9.3079
6	2	TMO66	2023.01.10-19:02:01	0.7	7.9	10.3922
7	2	INSH	2023.01.10-19:02:01	0.7	7.9	2.342
8	2	INS6B	2023.01.10-19:02:01	0.7	7.9	7.8568
9	2	INS3	2023.01.10-19:02:01	0.7	7.9	6.4882
10	2	INS1	2023.01.10-19:02:01	0.7	7.9	6.4037
11	3	TMO66	2023.01.10-10:45:58	2	8.9	10.13
12	3	TMO55	2023.01.10-10:45:58	2	8.9	8.3737
13	3	TMO20	2023.01.10-10:45:58	2	8.9	7.5393
14	3	INS8	2023.01.10-10:45:58	2	8.9	11.9023
15	3	INS3	2023.01.10-10:45:58	2	8.9	8.3417
16	3	INS1	2023.01.10-10:45:58	2	8.9	7.6114
17	4	TMO55	2022.12.22-06:05:04	1.7	5	9.5429
18	4	TMO20	2022.12.22-06:05:04	1.7	5	8.3645
19	4	INS8	2022.12.22-06:05:04	1.7	5	13.058
20	4	INS3	2022.12.22-06:05:04	1.7	5	7.7209
21	4	INS1	2022.12.22-06:05:04	1.7	5	7.9147
22	5	TMO66	2022.12.17-07:24:35	1.9	5	10.5496
23	5	TMO55	2022.12.17-07:24:35	1.9	5	8.7127
24	5	TMO22	2022.12.17-07:24:35	1.9	5	7.46
25	5	TMO20	2022.12.17-07:24:35	1.9	5	7.9618
26	5	INS8	2022.12.17-07:24:35	1.9	5	12.1816
27	5	INS6B	2022.12.17-07:24:35	1.9	5	8.2586
28	5	INS1	2022.12.17-07:24:35	1.9	5	7.9483
29	6	TMO66	2022.11.28-17:38:35	0.5	7	9.1177
30	6	TMO55	2022.11.28-17:38:35	0.5	7	7.0516
31	6	TMO20	2022.11.28-17:38:35	0.5	7	6.8976

32	6	INS8	2022.11.28-17:38:35	0.5	7	10.5504
33	6	INS6B	2022.11.28-17:38:35	0.5	7	9.9205
34	6	INS1	2022.11.28-17:38:35	0.5	7	7.6845
35	7	TMO66	2022.11.19-03:59:04	0.8	6.1	11.3142
36	7	TMO55	2022.11.19-03:59:04	0.8	6.1	9.019
37	7	TMO22	2022.11.19-03:59:04	0.8	6.1	6.9887
38	7	TMO20	2022.11.19-03:59:04	0.8	6.1	8.9588
39	7	INS6B	2022.11.19-03:59:04	0.8	6.1	8.1523
40	7	INS3	2022.11.19-03:59:04	0.8	6.1	9.3287
41	7	INS1	2022.11.19-03:59:04	0.8	6.1	9.1283
42	8	INS8	2022.09.22-02:36:06	0.5	3.8	11.0712
43	8	INS6B	2022.09.22-02:36:06	0.5	3.8	9.482
44	8	INS4B	2022.09.22-02:36:06	0.5	3.8	3.9831
45	8	INS3	2022.09.22-02:36:06	0.5	3.8	8.7477
46	9	TMO66	2022.09.20-20:45:41	0.7	6	9.3803
47	9	TMO22	2022.09.20-20:45:41	0.7	6	7.0485
48	9	INSH	2022.09.20-20:45:41	0.7	6	2.3283
49	9	INS8	2022.09.20-20:45:41	0.7	6	11.2834
50	9	INS1	2022.09.20-20:45:41	0.7	6	7.2037
51	10	TMO66	2022.07.21-23:47:40	0.6	5	9.968
52	10	TMO55	2022.07.21-23:47:40	0.6	5	8.5975
53	10	TMO22	2022.07.21-23:47:40	0.6	5	8.0199
54	10	TMO20	2022.07.21-23:47:40	0.6	5	7.1499
55	10	INSH	2022.07.21-23:47:40	0.6	5	1.9118
56	10	INS8	2022.07.21-23:47:40	0.6	5	12.2635
57	10	INS6B	2022.07.21-23:47:40	0.6	5	8.4628
58	10	INS3	2022.07.21-23:47:40	0.6	5	7.6001
59	11	TMO66	2022.04.27-22:54:46	0.5	4	10.7869
60	11	TMO55	2022.04.27-22:54:46	0.5	4	8.8992
61	11	TMO22	2022.04.27-22:54:46	0.5	4	7.507
62	11	INS3	2022.04.27-22:54:46	0.5	4	8.4444
63	12	TMO55	2022.04.25-00:58:55	0.5	3.7	9.8801
64	12	TMO22	2022.04.25-00:58:55	0.5	3.7	8.4453
65	12	INS8	2022.04.25-00:58:55	0.5	3.7	13.3415
66	12	INS4B	2022.04.25-00:58:55	0.5	3.7	6.3821

67	12	INS3	2022.04.25-00:58:55	0.5	3.7	7.8207
68	13	TMO66	2022.03.16-14:53:57	1	3	10.5911
69	13	TMO55	2022.03.16-14:53:57	1	3	9.1464
70	13	TMO22	2022.03.16-14:53:57	1	3	8.2752
71	13	TMO20	2022.03.16-14:53:57	1	3	7.7494
72	13	INS3	2022.03.16-14:53:57	1	3	7.5043
73	13	INS1	2022.03.16-14:53:57	1	3	7.3319
74	14	TMO66	2022.03.16-01:11:52	0.3	3.8	10.1202
75	14	TMO22	2022.03.16-01:11:52	0.3	3.8	7.9359
76	14	TMO20	2022.03.16-01:11:52	0.3	3.8	7.3414
77	14	INS8	2022.03.16-01:11:52	0.3	3.8	12.2859
78	14	INS3	2022.03.16-01:11:52	0.3	3.8	7.7253
79	14	INS1	2022.03.16-01:11:52	0.3	3.8	7.1325
80	15	TMO55	2022.03.07-03:05:03	0.7	3.1	8.5826
81	15	TMO20	2022.03.07-03:05:03	0.7	3.1	7.6313
82	15	INS8	2022.03.07-03:05:03	0.7	3.1	12.1222
83	15	INS6B	2022.03.07-03:05:03	0.7	3.1	8.3905
84	15	INS3	2022.03.07-03:05:03	0.7	3.1	8.1637
85	15	INS1	2022.03.07-03:05:03	0.7	3.1	7.5864
86	16	TMO66	2022.03.01-22:39:16	0.4	4	11.1576
87	16	TMO55	2022.03.01-22:39:16	0.4	4	9.5429
88	16	INSH	2022.03.01-22:39:16	0.4	4	0.78624
89	16	INS8	2022.03.01-22:39:16	0.4	4	13.058
90	16	INS6B	2022.03.01-22:39:16	0.4	4	7.4295
91	16	INS3	2022.03.01-22:39:16	0.4	4	7.7209
92	17	TMO55	2022.02.28-19:32:17	0.6	3.8	9.425
93	17	INS8	2022.02.28-19:32:17	0.6	3.8	13.1375
94	17	INS6B	2022.02.28-19:32:17	0.6	3.8	7.7557
95	18	TMO66	2022.02.01-00:48:12	1.2	5.2	10.5911
96	18	TMO55	2022.02.01-00:48:12	1.2	5.2	9.1464
97	18	TMO22	2022.02.01-00:48:12	1.2	5.2	8.2752
98	18	TMO20	2022.02.01-00:48:12	1.2	5.2	7.7494
99	18	INS1	2022.02.01-00:48:12	1.2	5.2	7.3319
100	19	TMO66	2022.01.31-04:06:10	1.2	5.2	10.5112
101	19	TMO55	2022.01.31-04:06:10	1.2	5.2	9.0402



102	19	TMO22	2022.01.31-04:06:10	1.2	5.2	8.1704
103	19	INSH	2022.01.31-04:06:10	1.2	5.2	1.3999
104	19	INS3	2022.01.31-04:06:10	1.2	5.2	7.5909
105	19	INS1	2022.01.31-04:06:10	1.2	5.2	7.3339
106	20	TMO55	2022.01.25-15:27:49	1.1	5	10.9669
107	20	TMO22	2022.01.25-15:27:49	1.1	5	10.078
108	20	TMO20	2022.01.25-15:27:49	1.1	5	8.8564
109	20	INS8	2022.01.25-15:27:49	1.1	5	14.6308
110	20	INS6B	2022.01.25-15:27:49	1.1	5	6.2221
111	20	INS1	2022.01.25-15:27:49	1.1	5	7.5521
112	21	TMO55	2022.01.11-22:36:30	0.4	4	10.4212
113	21	TMO22	2022.01.11-22:36:30	0.4	4	7.8093
114	21	INSH	2022.01.11-22:36:30	0.4	4	1.6257
115	21	INS6B	2022.01.11-22:36:30	0.4	4	7.103
116	21	INS4B	2022.01.11-22:36:30	0.4	4	7.3956
117	21	INS3	2022.01.11-22:36:30	0.4	4	9.6649
118	22	TMO66	2022.01.08-07:28:29	0.6	4	9.5887
119	22	TMO55	2022.01.08-07:28:29	0.6	4	8.3848
120	22	TMO22	2022.01.08-07:28:29	0.6	4	8.1031
121	22	TMO20	2022.01.08-07:28:29	0.6	4	6.7155
122	22	INS8	2022.01.08-07:28:29	0.6	4	12.1192
123	22	INS4B	2022.01.08-07:28:29	0.6	4	4.6444
124	22	INS3	2022.01.08-07:28:29	0.6	4	7.4575
125	22	INS1	2022.01.08-07:28:29	0.6	4	6.4899
126	23	TMO55	2021.12.30-18:31:50	0.8	3.7	8.5158
127	23	TMO22	2021.12.30-18:31:50	0.8	3.7	6.896
128	23	TMO20	2021.12.30-18:31:50	0.8	3.7	8.2677
129	23	INS4B	2021.12.30-18:31:50	0.8	3.7	5.2569
130	24	TMO55	2021.12.23-13:54:18	0.8	4	9.3938
131	24	TMO22	2021.12.23-13:54:18	0.8	4	7.1673
132	24	TMO20	2021.12.23-13:54:18	0.8	4	9.3766
133	24	INS6B	2021.12.23-13:54:18	0.8	4	7.8662
134	24	INS3	2021.12.23-13:54:18	0.8	4	9.4302
135	24	INS1	2021.12.23-13:54:18	0.8	4	9.4699
136	25	TMO55	2021.12.22-01:21:05	0.3	4	8.523

137	25	INS6B	2021.12.22-01:21:05	0.3	4	9.5429
138	26	TMO55	2021.12.15-23:02:26	0.3	4	7.844
139	26	TMO22	2021.12.15-23:02:26	0.3	4	5.6846
140	26	TMO20	2021.12.15-23:02:26	0.3	4	8.5284
141	26	INS8	2021.12.15-23:02:26	0.3	4	10.9274
142	26	INS6B	2021.12.15-23:02:26	0.3	4	9.4753
143	26	INS1	2021.12.15-23:02:26	0.3	4	9.2934
144	27	TMO55	2021.12.15-18:17:05	0.3	4	8.5291
145	27	TMO22	2021.12.15-18:17:05	0.3	4	7.4208
146	27	TMO20	2021.12.15-18:17:05	0.3	4	7.7173
147	27	INSH	2021.12.15-18:17:05	0.3	4	1.394
148	27	INS6B	2021.12.15-18:17:05	0.3	4	8.4397
149	28	TMO55	2021.12.13-21:01:08	0.6	5.2	10.2142
150	28	TMO22	2021.12.13-21:01:08	0.6	5.2	9.3315
151	28	INS8	2021.12.13-21:01:08	0.6	5.2	13.8573
152	28	INS4B	2021.12.13-21:01:08	0.6	5.2	6.5194
153	29	TMO55	2021.12.11-04:13:21	0.6	4	9.2527
154	29	TMO22	2021.12.11-04:13:21	0.6	4	8.3802
155	29	TMO20	2021.12.11-04:13:21	0.6	4	7.8062
156	29	INS8	2021.12.11-04:13:21	0.6	4	12.8637
157	29	INS3	2021.12.11-04:13:21	0.6	4	7.4182
158	30	TMO55	2021.12.06-16:50:11	0.9	4	9.1264
159	30	TMO22	2021.12.06-16:50:11	0.9	4	7.8867
160	30	TMO20	2021.12.06-16:50:11	0.9	4	8.1535
161	31	TMO55	2021.11.25-02:12:54	0.2	2.9	9.5081
162	31	INS3	2021.11.25-02:12:54	0.2	2.9	7.3439
163	31	INS1	2021.11.25-02:12:54	0.2	2.9	7.4813
164	32	TMO55	2021.11.25-00:08:56	0.4	4	9.2302
165	32	TMO20	2021.11.25-00:08:56	0.4	4	8.2045
166	32	INS8	2021.11.25-00:08:56	0.4	4	12.7291
167	32	INS3	2021.11.25-00:08:56	0.4	4	7.9663
168	33	TMO55	2021.11.19-22:51:54	0.5	4	9.2961
169	33	TMO22	2021.11.19-22:51:54	0.5	4	8.3337
170	33	TMO20	2021.11.19-22:51:54	0.5	4	7.9312
171	33	INS8	2021.11.19-22:51:54	0.5	4	12.8824

172	33	INS6B	2021.11.19-22:51:54	0.5	4	7.7073
173	34	TMO55	2021.11.18-06:58:45	0.5	2.4	9.5429
174	34	TMO54	2021.11.18-06:58:45	0.5	2.4	2.0834
175	34	TMO22	2021.11.18-06:58:45	0.5	2.4	8.3153
176	34	TMO20	2021.11.18-06:58:45	0.5	2.4	8.3645
177	34	INS8	2021.11.18-06:58:45	0.5	2.4	13.058
178	34	INS4B	2021.11.18-06:58:45	0.5	2.4	5.9986
179	34	INS3	2021.11.18-06:58:45	0.5	2.4	7.7209
180	35	TMO55	2021.11.03-20:40:15	0.9	3.9	9.0402
181	35	TMO22	2021.11.03-20:40:15	0.9	3.9	8.1704
182	35	TMO20	2021.11.03-20:40:15	0.9	3.9	7.6938
183	35	INSH	2021.11.03-20:40:15	0.9	3.9	1.3999
184	35	INS8	2021.11.03-20:40:15	0.9	3.9	12.643
185	35	INS3	2021.11.03-20:40:15	0.9	3.9	7.5909
186	35	INS1	2021.11.03-20:40:15	0.9	3.9	7.3339
187	36	INS8	2021.09.30-23:15:16	0.5	4	12.4134
188	36	INS1	2021.09.30-23:15:16	0.5	4	8.0031
189	37	TMO66	2021.08.21-17:00:29	1.2	4	10.5409
190	37	TMO55	2021.08.21-17:00:29	1.2	4	9.1252
191	37	TMO22	2021.08.21-17:00:29	1.2	4	8.2997
192	37	TMO20	2021.08.21-17:00:29	1.2	4	7.6867
193	37	INS8	2021.08.21-17:00:29	1.2	4	12.7445
194	37	INS1	2021.08.21-17:00:29	1.2	4	7.2592
195	38	TMO66	2021.08.21-16:25:38	1.1	4	10.5409
196	38	TMO55	2021.08.21-16:25:38	1.1	4	9.1252
197	38	TMO20	2021.08.21-16:25:38	1.1	4	7.6867
198	38	INS8	2021.08.21-16:25:38	1.1	4	12.7445
199	38	INS1	2021.08.21-16:25:38	1.1	4	7.2592
200	39	TMO22	2021.08.18-06:18:09	1.2	4	4.0921
201	39	TMO20	2021.08.18-06:18:09	1.2	4	7.9226
202	39	INS8	2021.08.18-06:18:09	1.2	4	9.0999
203	39	INS1	2021.08.18-06:18:09	1.2	4	9.4896
204	40	TMO55	2021.08.05-23:33:26	0.4	4	9.3532
205	40	TMO54	2021.08.05-23:33:26	0.4	4	8.2729
206	40	INS8	2021.08.05-23:33:26	0.4	4	11.1465

207	41	TMO66	2021.07.30-19:46:16	0.7	5.2	10.3269
208	41	INS8	2021.07.30-19:46:16	0.7	5.2	12.3217
209	41	INS3	2021.07.30-19:46:16	0.7	5.2	7.8987
210	41	INS1	2021.07.30-19:46:16	0.7	5.2	7.4227
211	42	TMO55	2021.06.28-22:30:14	0.2	4	10.3428
212	49	TMO55	2021.04.29-01:38:05	0.3	6.3	7.7766
213	49	INSH	2021.04.29-01:38:05	0.3	6.3	2.131
214	49	INS6B	2021.04.29-01:38:05	0.3	6.3	9.1937
215	50	TMO66	2021.04.21-04:22:48	0.3	4	9.6519
216	50	TMO22	2021.04.21-04:22:48	0.3	4	7.6137
217	50	TMO20	2021.04.21-04:22:48	0.3	4	6.9449
218	50	INS1	2021.04.21-04:22:48	0.3	4	6.9628
219	51	TMO55	2021.03.27-14:18:43	1.3	5	8.0624
220	51	TMO20	2021.03.27-14:18:43	1.3	5	6.8971
221	51	INSH	2021.03.27-14:18:43	1.3	5	2.1607
222	51	INS8	2021.03.27-14:18:43	1.3	5	11.7101
223	51	INS4B	2021.03.27-14:18:43	1.3	5	4.428
224	52	INS6B	2021.03.16-21:54:23	0.4	3.9	8.5294
225	53	TMO22	2021.02.26-16:28:45	1.2	5.3	6.4841
226	53	TMO20	2021.02.26-16:28:45	1.2	5.3	8.0579
227	53	INS8	2021.02.26-16:28:45	1.2	5.3	11.399
228	53	INS3	2021.02.26-16:28:45	1.2	5.3	9.3565
229	53	INS1	2021.02.26-16:28:45	1.2	5.3	8.508
230	54	TMO66	2021.01.08-05:14:16	0.7	7.7	11.4767
231	54	TMO55	2021.01.08-05:14:16	0.7	7.7	10.1674
232	54	TMO54	2021.01.08-05:14:16	0.7	7.7	1.5529
233	54	TMO20	2021.01.08-05:14:16	0.7	7.7	8.4771
234	54	INS8	2021.01.08-05:14:16	0.7	7.7	13.7738
235	54	INS4B	2021.01.08-05:14:16	0.7	7.7	6.5083
236	54	INS3	2021.01.08-05:14:16	0.7	7.7	6.911
237	54	INS1	2021.01.08-05:14:16	0.7	7.7	7.6059
238	55	TMO66	2020.11.09-21:47:07	1.1	7	9.5887
239	55	TMO55	2020.11.09-21:47:07	1.1	7	8.3848
240	55	INS8	2020.11.09-21:47:07	1.1	7	12.1192

## Bibliography

- [Abbiati et al 2021] ABBIATI, Giuseppe ; BROCCARDO, Marco ; ABDALLAH, Imad ; MARELLI, Stefano ; PAOLACCI, Fabrizio: Seismic fragility analysis based on artificial ground motions and surrogate modeling of validated structural simulators. In: *Earthquake Engineering & Structural Dynamics* 50 (2021), Nr. 9, p. 2314–2333
- [Allen 1971] ALLEN, David M.: *The prediction sum of squares as a criterion for selecting predictor variables*. University of Kentucky, 1971
- [Ancheta et al 2014] ANCHETA, Timothy D. ; DARRAGH, Robert B. ; STEWART, Jonathan P. ; SEYHAN, Emel ; SILVA, Walter J. ; CHIOU, Brian S-J ; WOODDELL, Katie E. ; GRAVES, Robert W. ; KOTTKE, Albert R. ; BOORE, David M. et al: NGA-West2 database. In: *Earthquake Spectra* 30 (2014), Nr. 3, p. 989–1005
- [ANSYS 2010] ANSYS, Inc.: *ANSYS ®Mechanical APDL Command Reference, Release 13.0*. 2010
- [Blatman and Sudret 2010] BLATMAN, Géraud ; SUDRET, Bruno: An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. In: *Probabilistic Engineering Mechanics* 25 (2010), Nr. 2, p. 183–197
- [Blatman and Sudret 2011] BLATMAN, Géraud ; SUDRET, Bruno: Adaptive sparse polynomial chaos expansion based on least angle regression. In: *Journal of computational Physics* 230 (2011), Nr. 6, p. 2345–2367
- [Broccardo and Dabaghi 2017] BROCCARDO, Marco ; DABAGHI, Mayssa: A spectral-based stochastic ground motion model with a non-parametric time-modulating function. In: *12th International Conference on Structural Safety and Reliability; Vienna* Volume 2017, 2017, p. 1–10
- [Chapelle et al 2002] CHAPELLE, Olivier ; VAPNIK, Vladimir ; BENGIO, Yoshua: Model selection for small sample regression. In: *Machine Learning* 48 (2002), p. 9–23
- [Chen et al 2022] CHEN, Bo ; SUN, Guangjun ; LI, Hongjing: Power spectral models of stationary earthquake-induced ground motion process considering site characteristics. In: *Emergency Management Science and Technology* 2 (2022), Nr. 1, p. 1–12
- [Clough 1975] CLOUGH, RW: *Penzien, j.(1975), Dynamics of structures*. 1975
- [COMSOL 2019] COMSOL: Response Spectrum Analysis. (2019). – URL <https://www.comsol.com/multiphysics/response-spectrum-analysis>. – date visited: April 24th, 2024

- [Efron et al 2004] EFRON, Bradley ; HASTIE, Trevor ; JOHNSTONE, Iain ; TIBSHIRANI, Robert: Least angle regression. (2004)
- [Ghanem and Spanos 2003] GHANEM, Roger G. ; SPANOS, Pol D.: *Stochastic finite elements: a spectral approach*. Courier Corporation, 2003
- [Golub and Welsch 1969] GOLUB, Gene H. ; WELSCH, John H.: Calculation of Gauss quadrature rules. In: *Mathematics of computation* 23 (1969), Nr. 106, p. 221–230
- [Hastie et al 2009] HASTIE, Trevor ; TIBSHIRANI, Robert ; FRIEDMAN, Jerome H. ; FRIEDMAN, Jerome H.: *The elements of statistical learning: data mining, inference, and prediction*. Volume 2. Springer, 2009
- [Housner et al 1953] HOUSNER, George W. ; MARTEL, RR ; ALFORD, JL: Spectrum analysis of strong-motion earthquakes. In: *Bulletin of the Seismological Society of America* 43 (1953), Nr. 2, p. 97–119
- [Housner and Jennings 1964] HOUSNER, GW ; JENNINGS, Paul C.: Generation of artificial earthquakes. In: *Journal of the Engineering Mechanics Division* 90 (1964), Nr. 1, p. 113–150
- [Hu and Zhou 1962] HU, Y ; ZHOU, X: The response of the elastic system under the stationary and nonstationary ground motions. In: *Earthquake Engineering Research Report* (1962), Nr. 1, p. 33–55
- [Kanai 1957] KANAI, Kiyoshi: Semi-empirical formula for the seismic characteristics of the ground. 35 (1957), p. 309–325
- [Kao and Pfleger 2023] KAO, Wei-Teng ; PFLEGER, Stefen: Frequency Dependent LPM for Soil-Structure Interaction under Induced Seismicity. (2023). – URL <https://collab.dvb.bayern/pages/viewpage.action?pageId=71122813>. – date visited: February 17th, 2024
- [Khansefid et al 2021] KHANSEFID, Ali ; YADOLLAHI, Mahmoudreza ; MÜLLER, Gerhard ; TADDEI, Francesca: *Seismic Risk Assessment of a Masonry Building Due to the Geothermal Power Plant Earthquakes*. Lehrstuhl für Baumechanik, 2021
- [Khansefid et al 2022] KHANSEFID, Ali ; YADOLLAHI, Seyed M. ; MÜLLER, Gerhard ; TADDEI, Francesca: Induced Earthquake Hazard by Geothermal Power Plants: Statistical Evaluation and Probabilistic Modeling. In: *International Journal of Disaster Risk Science* 13 (2022), Nr. 5, p. 758–777
- [Kraft et al 2009] KRAFT, Toni ; MAI, Paul M. ; WIEMER, Stefan ; DEICHMANN, Nicholas ; RIPPERGER, Johannes ; KÄSTLI, Philipp ; BACHMANN, Corinne ; FÄH, Donat ; WÖSSNER, Jochen ; GIARDINI, Domenico: Enhanced geothermal systems: Mitigating risk in urban areas. In: *Eos, Transactions American Geophysical Union* 90 (2009), Nr. 32, p. 273–274
- [Küperkoch et al 2018] KÜPERKOCH, L ; OLBERT, K ; MEIER, T: Long-term monitoring of induced seismicity at the Insheim geothermal site, Germany. In: *Bulletin of the Seismological Society of America* 108 (2018), Nr. 6, p. 3668–3683

- [Lagaros et al 2013] LAGAROS, Nikos D. ; MITROPOULOU, Chara C. ; PAPADRAKAKIS, Manolis: Time history seismic analysis. In: *Encyclopedia of Earthquake Engineering* (2013), p. 1–19
- [Liu and Nocedal 1989] LIU, Dong C. ; NOCEDAL, Jorge: On the limited memory BFGS method for large scale optimization. In: *Mathematical programming* 45 (1989), Nr. 1, p. 503–528
- [Liu 1970] LIU, Shin-Chi: Evolutionary power spectral density of strong-motion earthquakes. In: *Bulletin of the Seismological Society of America* 60 (1970), Nr. 3, p. 891–900
- [Marsaglia and Tsang 2000] MARSAGLIA, George ; TSANG, Wai W.: The ziggurat method for generating random variables. In: *Journal of statistical software* 5 (2000), p. 1–7
- [Mostafa Tazarv 2024] MOSTAFA TAZARV: Elastic Response Spectrum. (2024).  
– URL <https://www.mathworks.com/matlabcentral/fileexchange/31254-elastic-response-spectrum>
- [Muntasir Billah and Shahria Alam 2015] MUNTASIR BILLAH, AHM ; SHAHRIA ALAM, M: Seismic fragility assessment of highway bridges: a state-of-the-art review. In: *Structure and infrastructure engineering* 11 (2015), Nr. 6, p. 804–832
- [Nazri 2018] NAZRI, Fadzli M.: *Seismic fragility assessment for buildings due to earthquake excitation*. Springer, 2018
- [Nelsen 1999] NELSEN, Roger B.: An Introduction to Copulas, volume 139 of. In: *Lecture Notes in Statistics* (1999)
- [Pan et al 2016] PAN, Chao ; ZHANG, Ruifu ; LUO, Hao ; SHEN, Hua: Baseline correction of vibration acceleration signals with inconsistent initial velocity and displacement. In: *Advances in Mechanical Engineering* 8 (2016), Nr. 10, p. 1687814016675534
- [Pedregosa et al 2011] PEDREGOSA, F. ; VAROQUAUX, G. ; GRAMFORT, A. ; MICHEL, V. ; THIRION, B. ; GRISEL, O. ; BLONDEL, M. ; PRETTENHOFER, P. ; WEISS, R. ; DUBOURG, V. ; VANDERPLAS, J. ; PASSOS, A. ; COURNAPEAU, D. ; BRUCHER, M. ; PERROT, M. ; DUCHESNAY, E.: Scikit-learn: Machine Learning in Python. In: *Journal of Machine Learning Research* 12 (2011), p. 2825–2830
- [Pineiro and Bates 1996] PINHEIRO, José C ; BATES, Douglas M.: Unconstrained parametrizations for variance-covariance matrices. In: *Statistics and computing* 6 (1996), p. 289–296
- [Priestley 1967] PRIESTLEY, MB: Power spectral analysis of non-stationary random processes. In: *Journal of Sound and Vibration* 6 (1967), Nr. 1, p. 86–97
- [Rezaeian 2010] REZAEIAN, Sanaz: *Stochastic modeling and simulation of ground motions for performance-based earthquake engineering*. University of California, Berkeley, 2010
- [Rezaeian and Der Kiureghian 2010] REZAEIAN, Sanaz ; DER KIUREGHIAN, Armen: Simulation of synthetic ground motions for specified earthquake and site characteristics. In: *Earthquake Engineering & Structural Dynamics* 39 (2010), Nr. 10, p. 1155–1180

- [Snoek et al 2012] SNOEK, Jasper ; LAROCHELLE, Hugo ; ADAMS, Ryan P.: Practical bayesian optimization of machine learning algorithms. In: *Advances in neural information processing systems* 25 (2012)
- [Soize and Ghanem 2004] SOIZE, Christian ; GHANEM, Roger: Physical systems with random uncertainties: chaos representations with arbitrary probability measure. In: *SIAM Journal on Scientific Computing* 26 (2004), Nr. 2, p. 395–410
- [Sudret 2008] SUDRET, Bruno: Global sensitivity analysis using polynomial chaos expansions. In: *Reliability engineering & system safety* 93 (2008), Nr. 7, p. 964–979
- [Sudret 2014] SUDRET, Bruno: Polynomial chaos expansions and stochastic finite element methods. In: *Risk and reliability in geotechnical engineering* (2014), p. 265–300
- [Tajimi 1960] TAJIMI, Hiroshi: A statistical method of determining the maximum response of a building structure during an earthquake. In: *Proc. of the 2nd World Conference on Earthquake Engineering, Tokyo, Japan, 1960*, 1960, p. 781–797
- [The MathWorks Inc. 2022] THE MATHWORKS INC.: linear mixed-effects model, MATLAB version: 9.13.0 (R2022b). (2022). – URL <https://www.mathworks.com/help/stats/linearmixedmodel.html>
- [Trifunac 1971] TRIFUNAC, Mihailo D.: Zero baseline correction of strong-motion accelerograms. In: *Bulletin of the Seismological Society of America* 61 (1971), Nr. 5, p. 1201–1211
- [Villani et al 2009] VILLANI, Cédric et al: *Optimal transport: old and new*. Volume 338. Springer, 2009
- [Virtanen et al 2020] VIRTANEN, Pauli ; GOMMERS, Ralf ; OLIPHANT, Travis E. ; HABERLAND, Matt ; REDDY, Tyler ; COURNAPEAU, David ; BUROVSKI, Evgeni ; PETERSON, Pearu ; WECKESSER, Warren ; BRIGHT, Jonathan ; VAN DER WALT, Stéfan J. ; BRETT, Matthew ; WILSON, Joshua ; MILLMAN, K. J. ; MAYOROV, Nikolay ; NELSON, Andrew R. J. ; JONES, Eric ; KERN, Robert ; LARSON, Eric ; CAREY, C J. ; POLAT, İlhan ; FENG, Yu ; MOORE, Eric W. ; VANDERPLAS, Jake ; LAXALDE, Denis ; PERKTOLD, Josef ; CIMRMAN, Robert ; HENRIKSEN, Ian ; QUINTERO, E. A. ; HARRIS, Charles R. ; ARCHIBALD, Anne M. ; RIBEIRO, Antônio H. ; PEDREGOSA, Fabian ; VAN MULBREGT, Paul ; SCI-PY 1.0 CONTRIBUTORS: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. In: *Nature Methods* 17 (2020), p. 261–272
- [Wiener 1938] WIENER, Norbert: The homogeneous chaos. In: *American Journal of Mathematics* 60 (1938), Nr. 4, p. 897–936
- [Wolf 1994] WOLF, John P.: *Foundation vibration analysis using simple physical models*. Pearson Education, 1994
- [Xiu and Karniadakis 2002] XIU, Dongbin ; KARNIADAKIS, George E.: The Wiener–Askey polynomial chaos for stochastic differential equations. In: *SIAM journal on scientific computing* 24 (2002), Nr. 2, p. 619–644



- [Zhong et al 2023] ZHONG, Kuanshi ; NAVARRO, Javier G. ; GOVINDJEE, Sanjay ; DEIERLEIN, Gregory G.: Surrogate modeling of structural seismic response using probabilistic learning on manifolds. In: *Earthquake Engineering & Structural Dynamics* 52 (2023), Nr. 8, p. 2407–2428
- [Zhu et al 2023] ZHU, Xujia ; BROCCARDO, Marco ; SUDRET, Bruno: Seismic fragility analysis using stochastic polynomial chaos expansions. In: *Probabilistic Engineering Mechanics* 72 (2023), p. 103413
- [Zhu and Sudret 2021] ZHU, Xujia ; SUDRET, Bruno: Emulation of stochastic simulators using generalized lambda models. In: *SIAM/ASA Journal on Uncertainty Quantification* 9 (2021), Nr. 4, p. 1345–1380
- [Zhu and Sudret 2023] ZHU, Xujia ; SUDRET, Bruno: Stochastic polynomial chaos expansions to emulate stochastic simulators. In: *International Journal for Uncertainty Quantification* 13 (2023), Nr. 2