

---

Chair of Structural Mechanics  
Technical University of Munich

Integrated Parameter Study and Serviceability Assessment of  
Building Structures under Geothermally Induced Seismicity  
using Stochastic Polynomial Chaos Expansion

B.Sc. Wei-Teng, Kao

Master thesis in the study program civil engineering  
Focus area Structural Dynamics

Referent : Univ.-Prof. Dr.-Ing. Gerhard Müller  
Dr.-Ing. Francesca Taddei

Betreuer : Dr. Aditi Kumawat  
M.Sc. Felix Schneider

Eingereicht : August 13, 2024

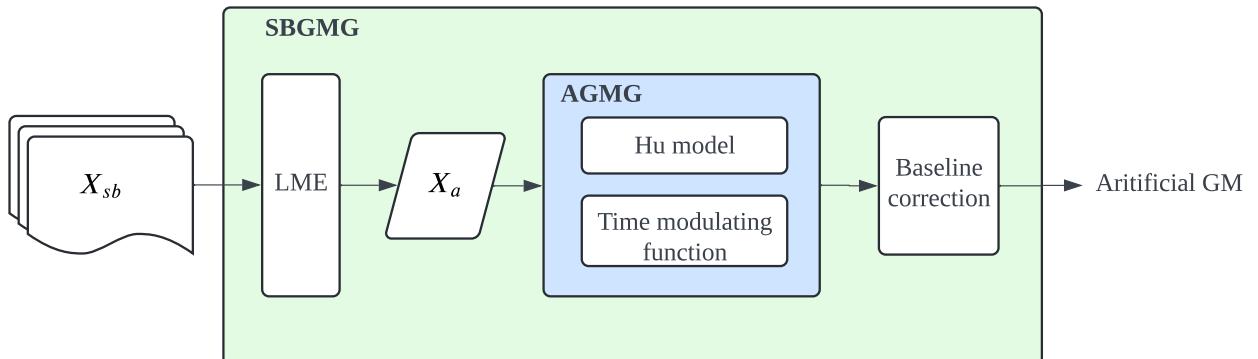
where the  $GM_{vel}$  is the artificial ground motion in velocity. The example flowchart of step 1 (figure A.1) to step 5 (figure A.5) is presented in section A.2.

4. The baseline correction is applied to ensure the baseline of ground motion will not be shifted after integration.
5. Build the linear mixed effect model (LME) for the SBGMG. The seismic parameter  $\mathbf{X}_{sb}$  is reported from the database. By training the LME model, the relationship between  $\mathbf{X}_a$  and  $\mathbf{X}_{sb}$  will be constructed:

$$\mathbf{X}_a = \text{LME}(\mathbf{X}_{sb}) + \boldsymbol{\epsilon}_{res} \quad (2.2)$$

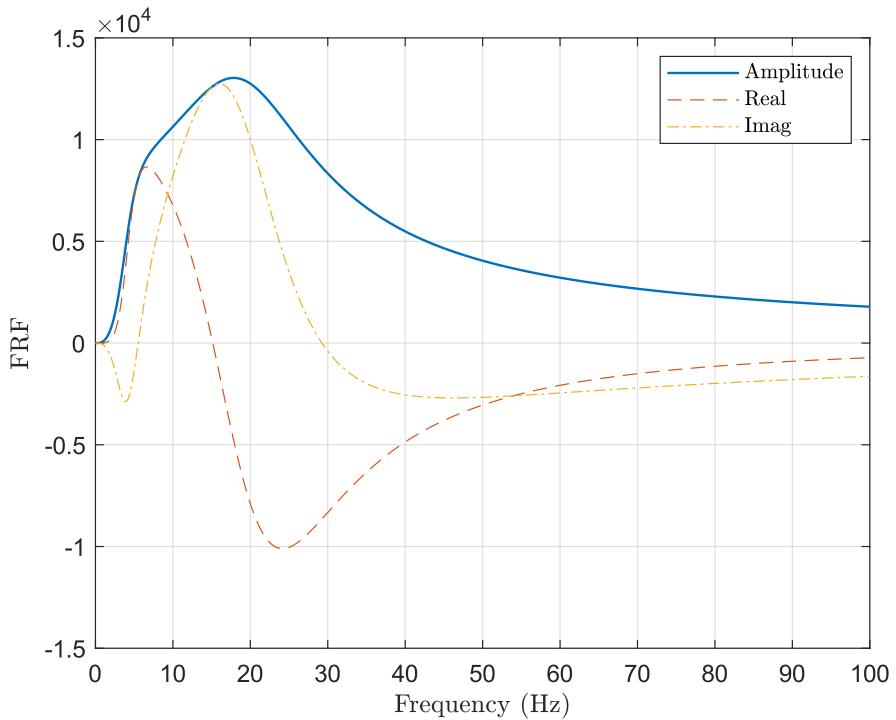
where the potential seismic parameters of  $\mathbf{X}_{sb}$  are  $\mathbf{X}_{sb} = \{M_L, R_x, R_y, D, \omega_g\}$  and  $\boldsymbol{\epsilon}_{res}$  is the residual of fitting result.

6. Generate the ground motion by SBGMG. When  $\mathbf{X}_{sb}$  is input, the model will first output the  $\mathbf{X}_a$  based on the LME model. Following this, AGMG will read those returned  $\mathbf{X}_a$  and generate the artificial ground motion. This artificial ground motion will finally be validated by response spectral analysis and used to excite the building structure. The framework of SBGMG is shown in figure 2.1.



**Figure 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.

It is worth mentioning that the self-built database is required to train the ground motion model. Since the research focuses on the geothermal-induced ground motions, which generally have relatively small moment magnitudes, i.e.,  $M_L \leq 3.0$ , and have different characteristics compared to the strong one recorded in the PEER Ground Motion Database, the



**Figure 2.3:** The FRF of the Hu model, formulated using equation (2.16).

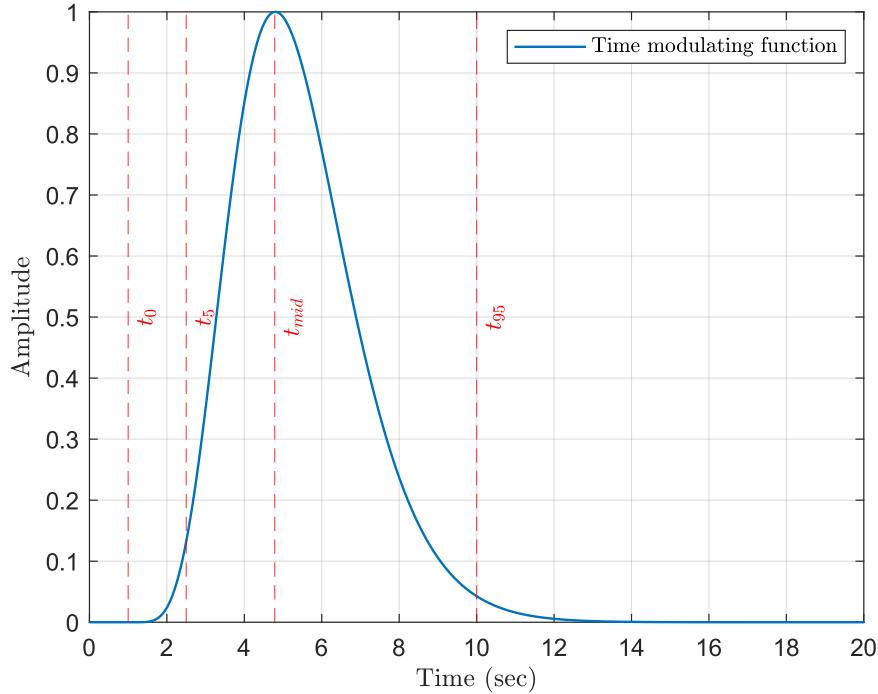
### 2.3.3 Time modulating function

The time-modulating function introduces the temporal non-stationarity of the ground motion. The temporal and spectral stationarities in the model are entirely decoupled [Rezaeian 2010]. The decoupling is done by normalizing the Hu model, equation (2.13), and controlling the intensity or amplitude of the ground motion by the time-modulating function only.

Two types of time-modulating functions are widely mentioned: the piecewise modulating function, modified from [Housner and Jennings 1964] and the Gamma PDF-like modulating function. In this thesis, the gamma PDF-like modulating function is used which can be defined as follows:

$$\begin{aligned}
 q(\tau, a, b) &= 0 && \text{if } \tau \leq T_0 \\
 &= \frac{1}{b^a \Gamma(a)} \int_0^\tau t^{a-1} e^{-\frac{\tau}{b}} dt && \text{if } \tau \geq T_0
 \end{aligned} \tag{2.19}$$

where  $\tau$  is each time step in a time series,  $T_0$  is the start time of ground motion,  $\Gamma$  denotes the gamma function and  $a, b$  is the parameter controlling the shape and scale, respectively.



**Figure 2.4:** The time-modulating function, formulated using equation (2.19).

### 2.3.4 Artifical ground motion generator

AGMG includes the time-modulating function,  $q(t, a, b)$  and Hu model,  $\mathbf{M}_{hu}(t, \mathbf{X}_{hu})$ , which can be formed as equation (2.21)).

$$GM_{vel} = PGV \frac{q(\tau, a, b)}{\max(|q(\tau, a, b) \mathbf{M}_{hu}(t, \mathbf{X}_{hu})|)} \mathbf{M}_{hu}(t, \mathbf{X}_{hu}) \quad (2.21)$$

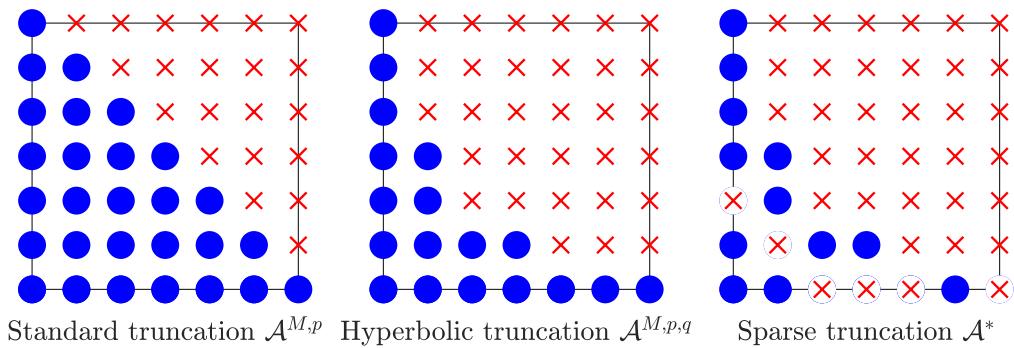
where parameters  $a$  and  $b$  in  $q$  can be derived from time information i.e. the duration,  $t_d$  and ratio  $r_{mid} = \frac{t_{mid}}{t_d}$  [Rezaeian 2010]. Therefore, the  $q(t, a, b)$  can be written as  $\hat{q}(t, t_d, r_{mid})$ , where  $\hat{q}$  is the new form of time-modulating function. Finally, the AGMG is formed as equation (2.22).

$$\begin{aligned} GM_{vel} &= PGV \frac{\hat{q}(t, t_d, r_{mid})}{\max(|\hat{q}(t, t_d, r_{mid}) \mathbf{M}_{hu}(t, \mathbf{X}_{hu})|)} \mathbf{M}_{hu}(t, \mathbf{X}_{hu}) \\ &= \mathbf{AGMG}(\mathbf{X}_{gmf}, \mathbf{X}_{hu}) \\ &= \mathbf{AGMG}(\mathbf{X}_a) \end{aligned} \quad (2.22)$$

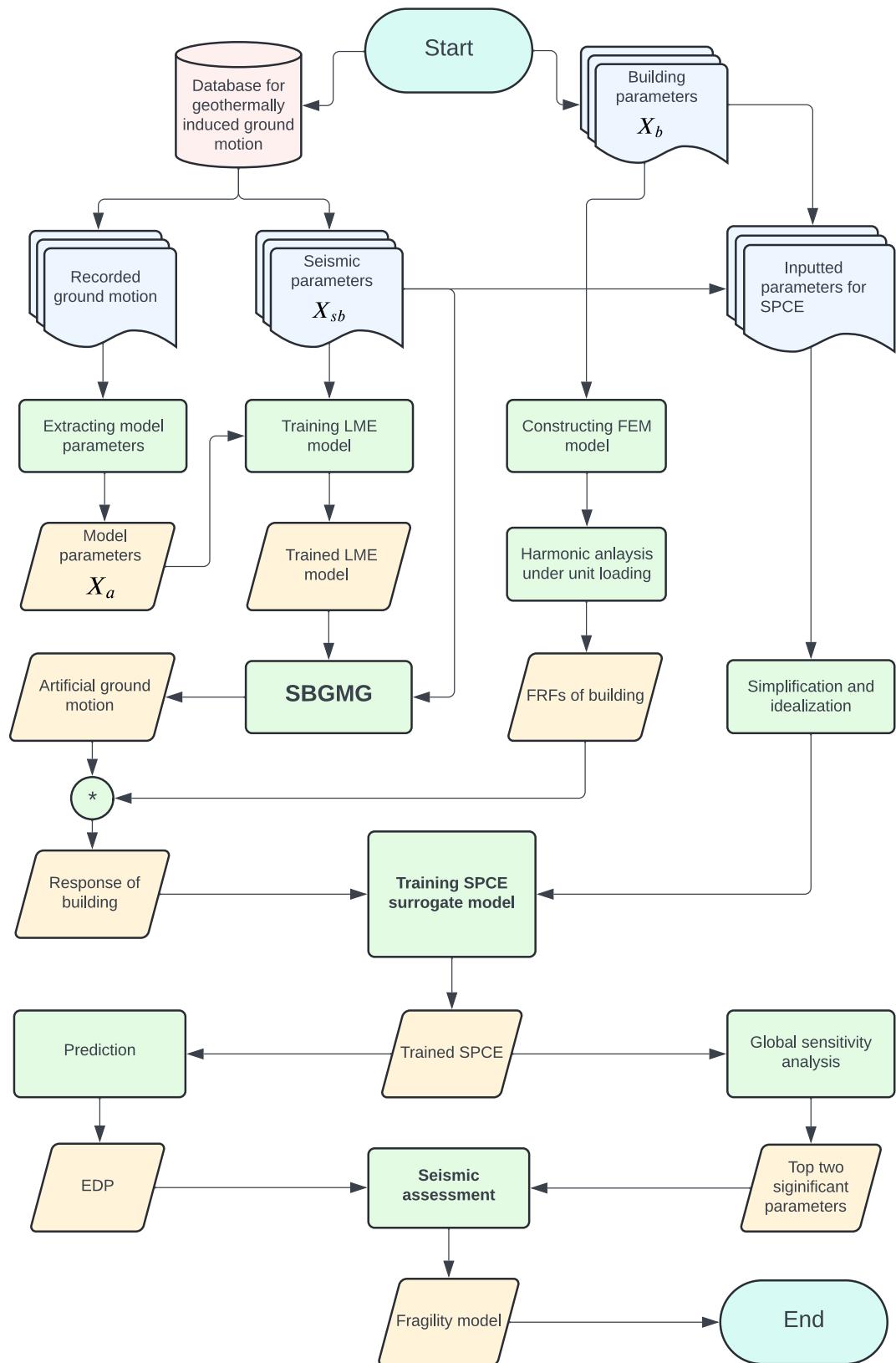
The workflow of LAR is as follows:

1. Selecting a candidate set of polynomials  $\mathcal{A}$ . The hyperbolic truncation set  $\mathcal{A}^{M,p,q}$  is assumed in this case.
2. Constructing a sequence of sparse basis with  $1, 2 \dots, \text{card } \mathcal{A}$  terms. The algorithm will first search the basis term  $\Psi_{\alpha_1}$  which is the most correlated with  $Y$ . Once it is identified, the corresponding coefficient is computed, which makes the residual  $Y - c_{\alpha_1}^{(1)} \Psi_{\alpha_1}(\mathbf{X})$  equally-correlated with two basis terms  $(\Psi_{\alpha_1}, \Psi_{\alpha_2})$ . Then, the current best approximation is updated by moving along the direction computed by the vector of  $(\Psi_{\alpha_1}, \Psi_{\alpha_2})$  to the next point where the residual becomes equally-correlated again with  $\Psi_{\alpha_3}$ . The searching and fitting procedure will iteratively run until it reaches the last polynomial terms at  $\text{card } \mathcal{A}^{M,p,q}$ .
3. Finally, the LAR algorithm will return a less and less sparse expansion sequence. One can select how many terms of the sequence are used for building the polynomial basis  $\mathcal{A}$ . To choose the number of terms for optimal polynomial basis  $\mathcal{A}^*$ , one can compare the leave-one-out error for each expansion and find the one with the minor error. The computation of the leave-one-out error will be introduced in the following section, which discusses leave-one-out cross-validation.

After performing the LAR algorithm, the optimal or, at least, the reduced polynomial basis is built up. With this basis, fewer coefficients of polynomial terms need to be computed, which makes the PCE algorithm more efficient. The detailed implementation of the LAR algorithm on the PCE application is mentioned in [Blatman and Sudret 2011]. The comparison of different truncated schemes is presented in figure 3.1.

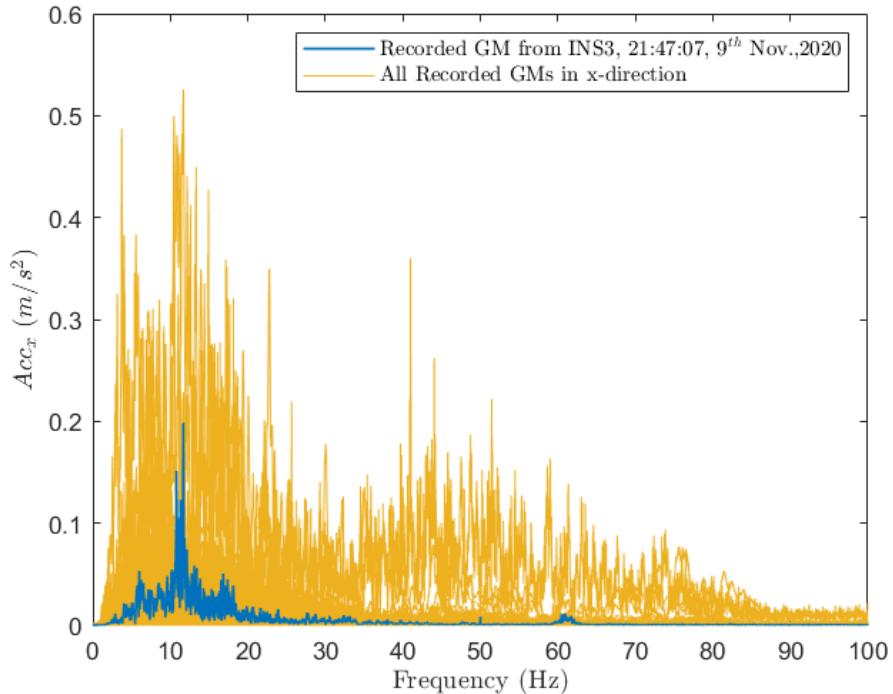


**Figure 3.1:** Sketch of the different truncation sets, modified from [Zhu and Sudret 2023].

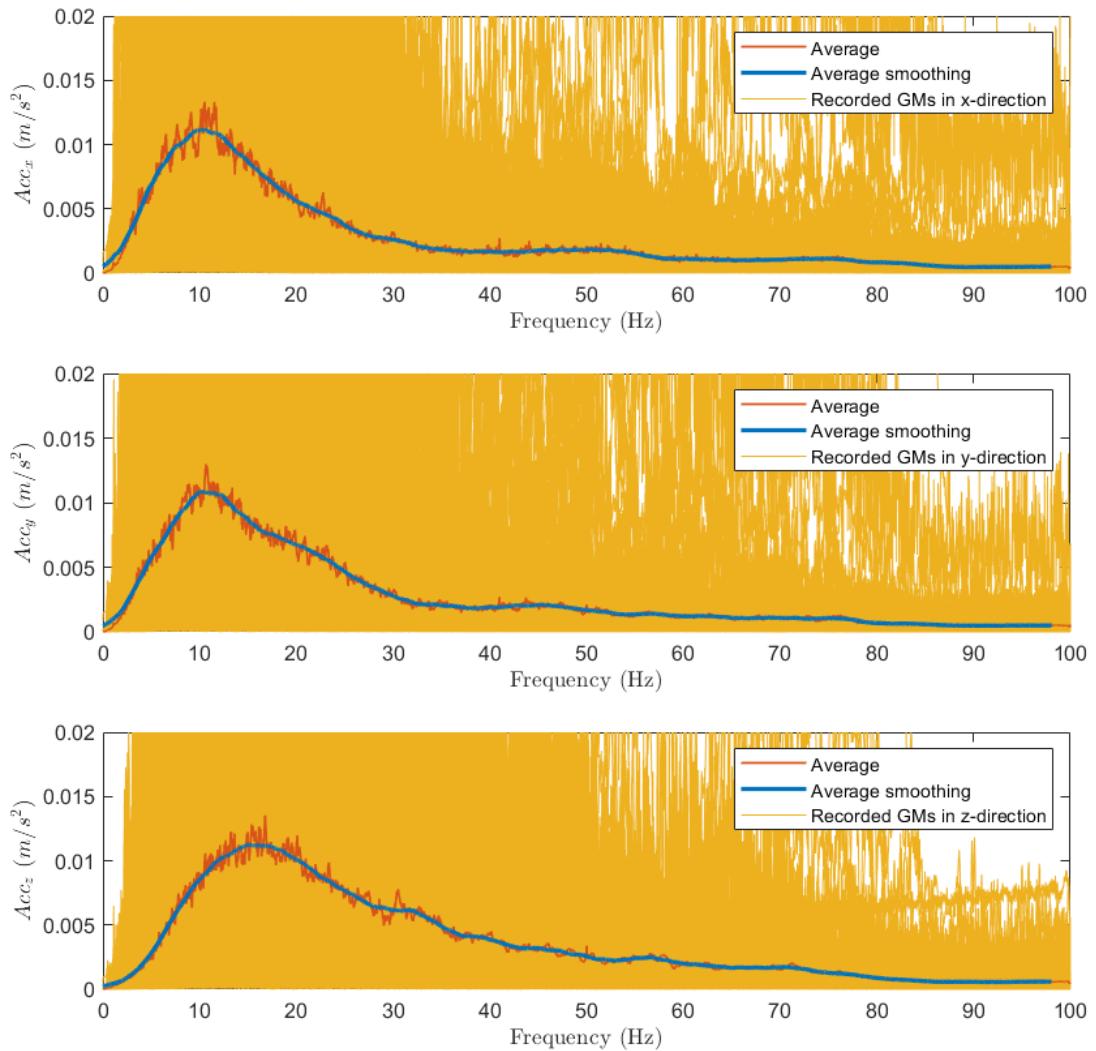


**Figure 5.1:** Overall framework used for this thesis, extend and modify from the existing framework in [Zhu et al 2023].

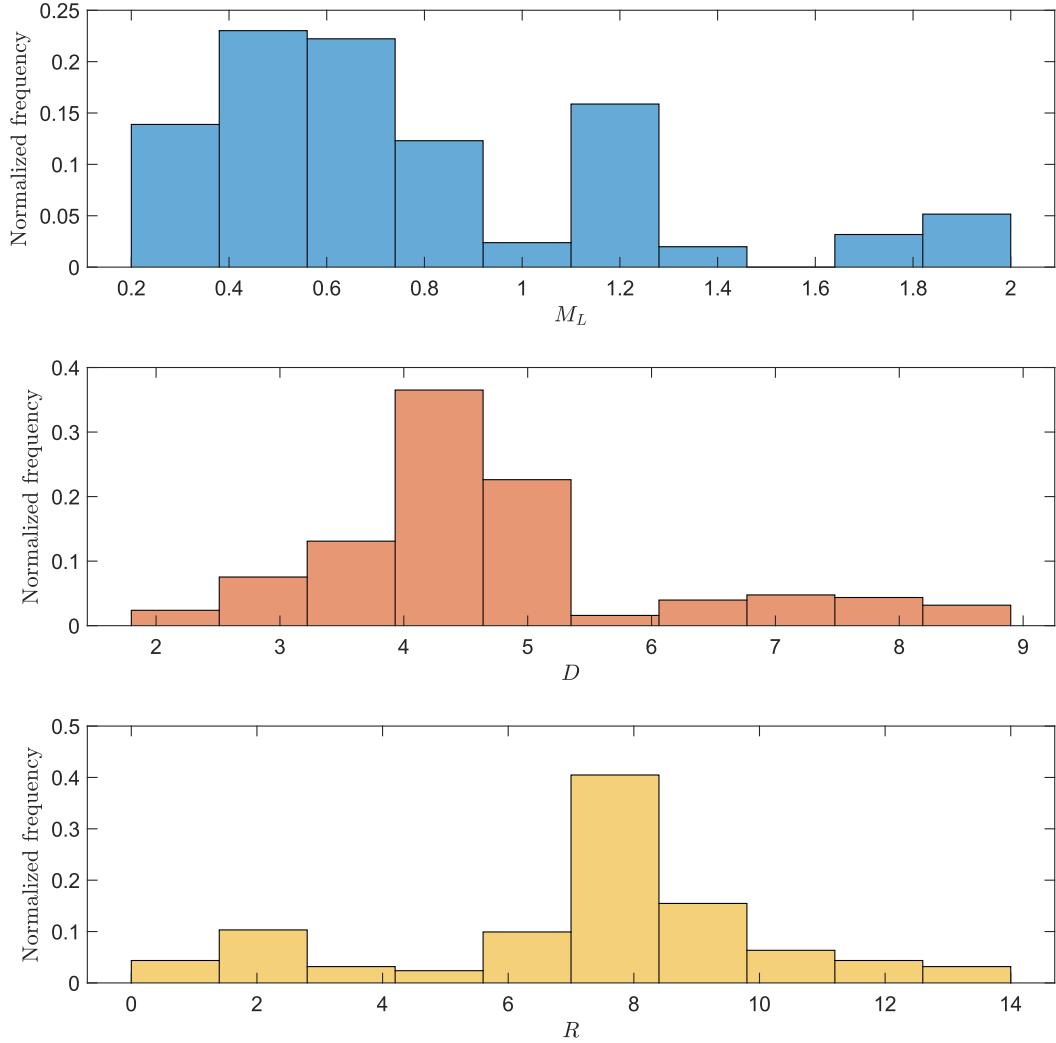
motion in [Zhu and Sudret 2023; Abbiati et al 2021]. These make  $\omega_g$  an appropriate candidate to describe the site characteristic used for SBGMG.



**Figure 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.



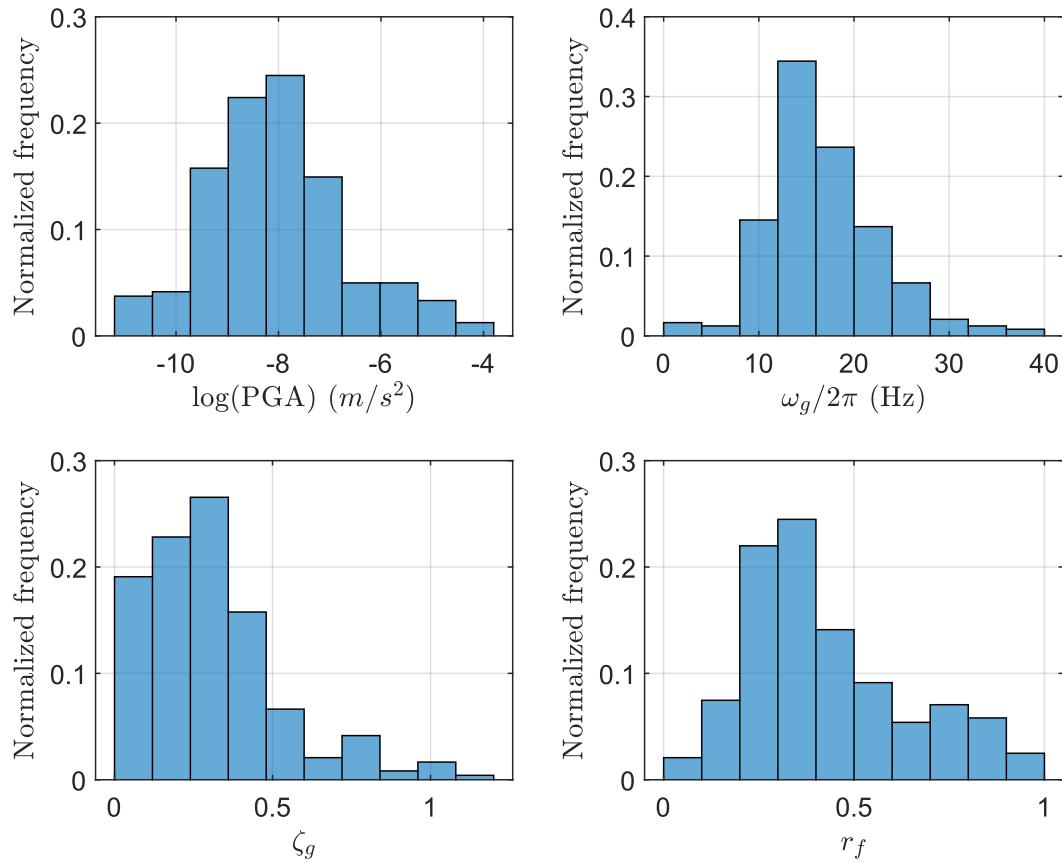
**Figure 6.2:** Collection of recorded ground motions from the database in the frequency domain. The yellow lines represent 240 ground motions used for identifying the model parameters. The orange line shows the average of ground motions, and the blue line represents the smoothing version of the average one, which is calculated by the moving average function with a window size of 250.



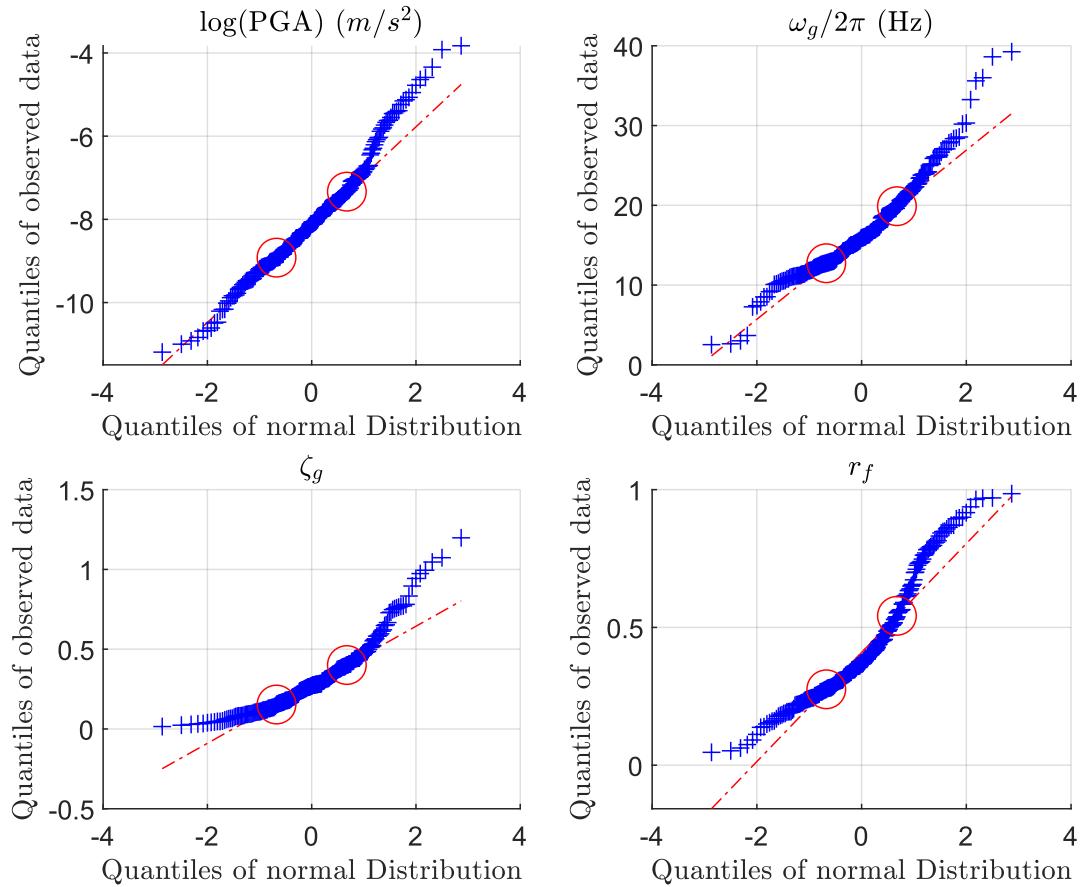
**Figure 6.3:** Histogram of  $M_L$ ,  $D$  and  $R$ . The frequency in the histogram is normalized by the relative probability and the height of bin value  $v_i$  is calculated as  $v_i = \frac{c_i}{N}$ , where  $c_i$  is the number of elements in the bin, and  $N$  is the total number of input data,  $N = 240$ .

### Model parameters fitting from the database

For each recorded ground motion, model parameters,  $\omega_g$ ,  $\zeta_g$ ,  $r_h$  are identified based on the nonlinear least square method introduced in the section, and the parameters  $PGA$ ,  $t_d$ ,  $r_{mid}$  can be extracted directly from each record. The fitting processes are repeated three times, where the initial guess,  $\mathbf{X}_a^{(0)}$  is randomly selected in the range  $50 \leq \omega_g^{(0)} \leq 300$ ,  $0.4 \leq \zeta_g^{(0)} \leq 1.0$ ,  $0.1 \leq r_h^{(0)} \leq 1.0$ . The threshold of selected result sets is settled as  $\varepsilon_P \leq 0.1$ .



**Figure 6.4:** Histograms of model parameters  $\omega_g/2\pi$ ,  $\zeta_g$ ,  $r_h$ ,  $\text{PGA}$  in  $z$ -direction. The frequency in the histogram is normalized by the relative probability, where the  $\text{PGA}$  is presented on the logarithm scale.



**Figure 6.5:** Quantile plots of the model parameter in  $z$ -direction. Hollow circles indicate the first and the third quartiles.

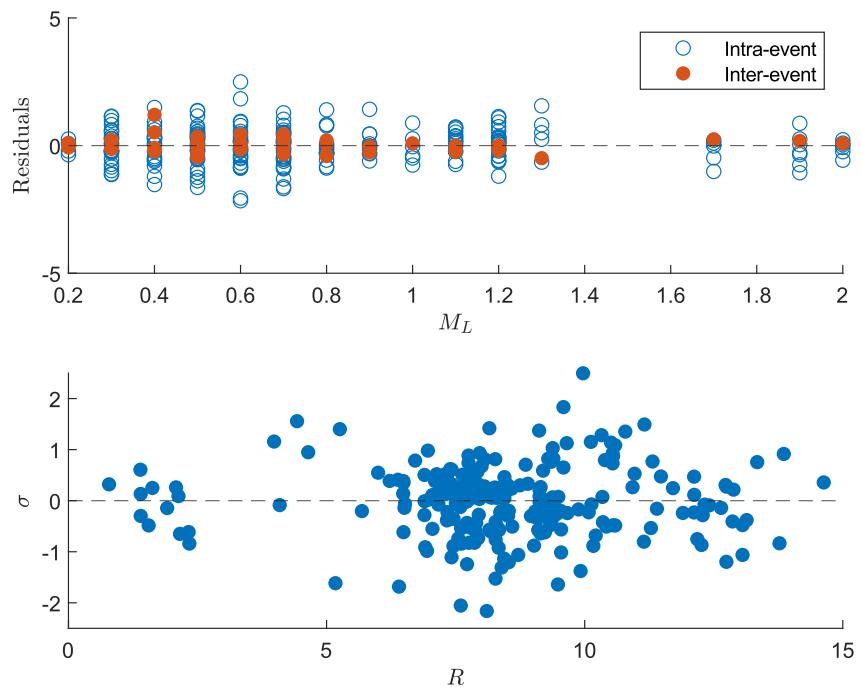
### 6.1.2 Fitted LME results of seismic parameters

A LME model for each modal parameter is built to represent the effect of seismic parameters. Various linear forms of regression models are tested, inspired by [Rezaeian and Der Kiureghian 2010; Khansefid et al 2021], and the current best-fitted LME model for each parameter is presented. As a reminder, the LME model is the regression model that explains each modal parameter  $PGA$ ,  $r_h$  and  $\zeta_g$  for AGMG by the combination of seismic parameter  $M_L$ ,  $R$ ,  $D$ , and  $\omega_g$ . Some seismic parameters might not explain the modal parameter well, or, inversely, some modal parameters might have a weak relationship with the seismic parameter.

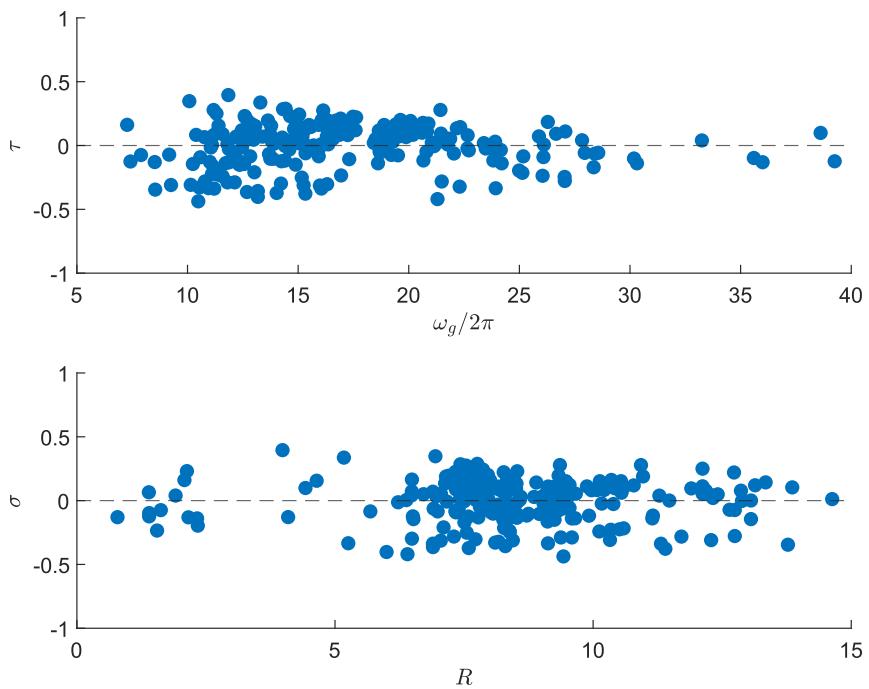
To illustrate the fitting result, standard significance tests and the residuals from intra-event and inter-event will be computed. Additionally, the idealization might be considered if the results show low significance on fitting coefficients from standard significance tests, i.e.,

The fitting results for the second trial are now shown in table 6.2 for  $z$ -direction and tables A.2, A.4 for other directions. As expected, the p-values indicate that the coefficients of the proposed seismic parameters are significant for fitting  $\ln(PGA)$  and  $r_h$ . Nevertheless,  $\zeta_g$  still cannot be explained well by seismic parameters since it fails to pass the t-test.

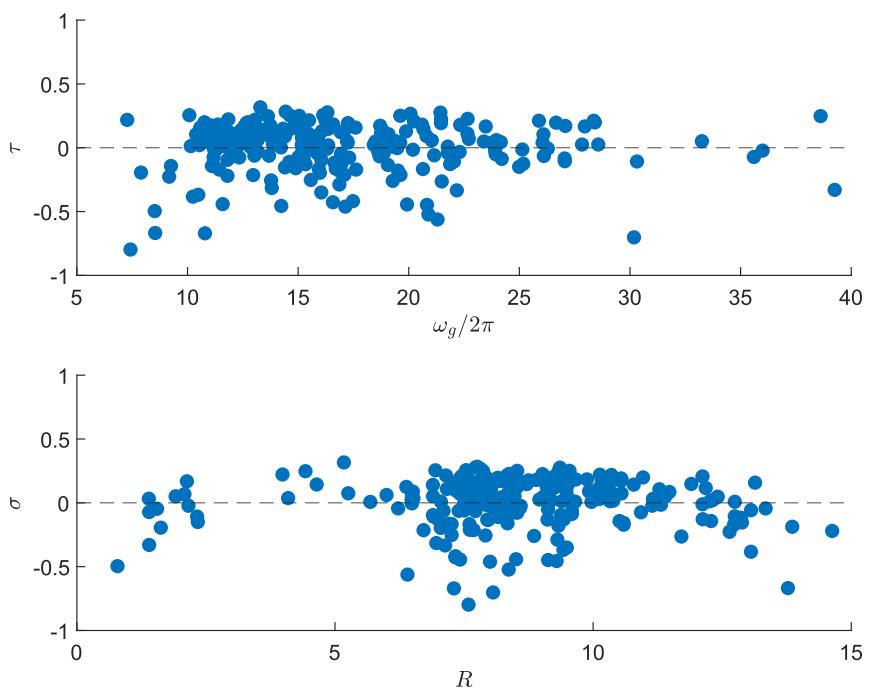
The  $\ln(PGA)$  coefficients reported in table 6.2 provide an interesting insight that this value will grow if  $M_L$  increases or  $R$  decreases. This phenomenon matches the common sense that strong vibration will be detected when a large earthquake occurs or stays near the epicenter. Following this, the model's residuals in  $z$  directions are presented in figure 6.8. Others are presented in Appendix (figure A.12 to A.17). For  $\ln(PGA)$  and  $r_h$ , the residuals are evenly distributed above and below the zero line with no obvious systematic trends. This result implies a lack of bias and a good fit of the regression models to the data [Rezaeian and Der Kiureghian 2010]; Conversely, the residuals of  $\zeta_g$  are not well distributed that negative residuals are comparably larger, especially in the  $z$ -direction (figure 6.8). This phenomenon matches the conclusion made in the section that  $\zeta_g$  might not fulfill the normality assumption for the regression model and has the worst fit to the quantile plot (figure 6.5).



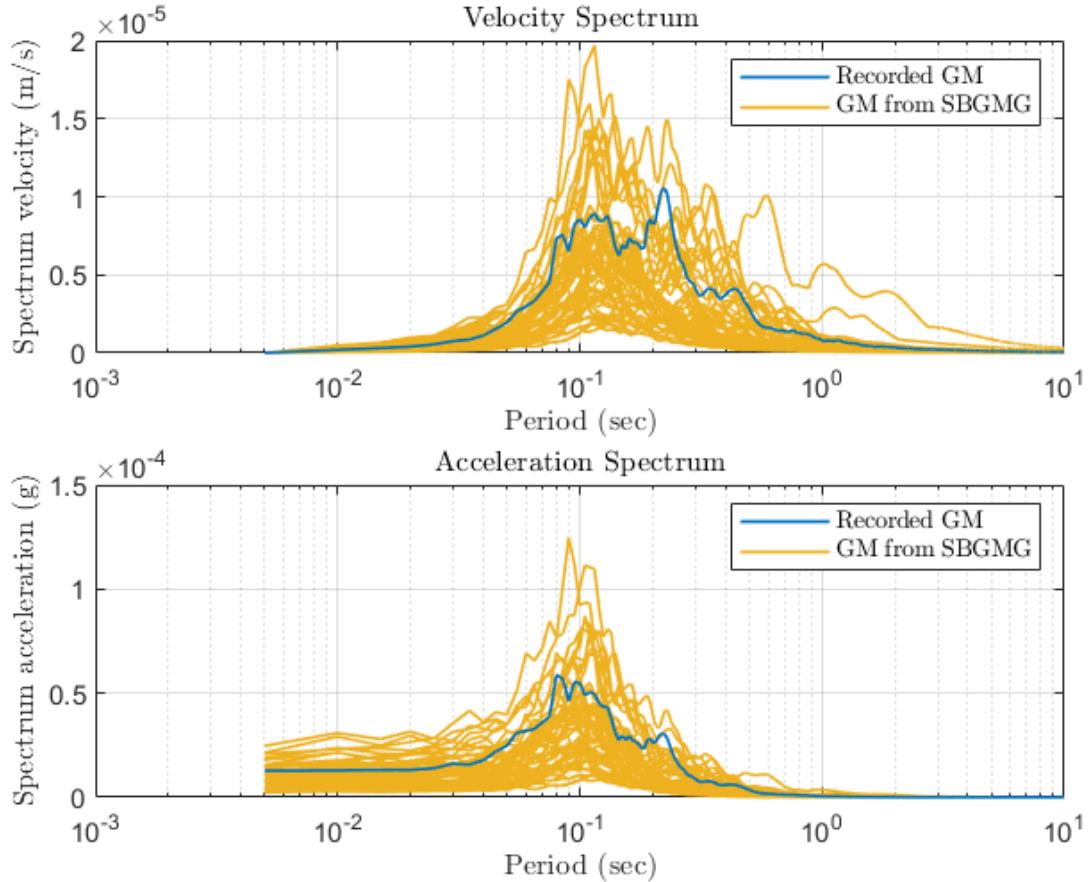
**Figure 6.6:** Scatter plots of residuals of  $\ln(PGA)$  against  $M_L$  and  $R$  in  $z$ -direction



**Figure 6.7:** Scatter plots of residuals of  $\omega_f$  against  $\omega_g$  and  $R$  in  $z$ -direction



**Figure 6.8:** Scatter plots of residuals of  $\zeta_g$  against  $\omega_g$  and  $R$  in  $z$ -direction



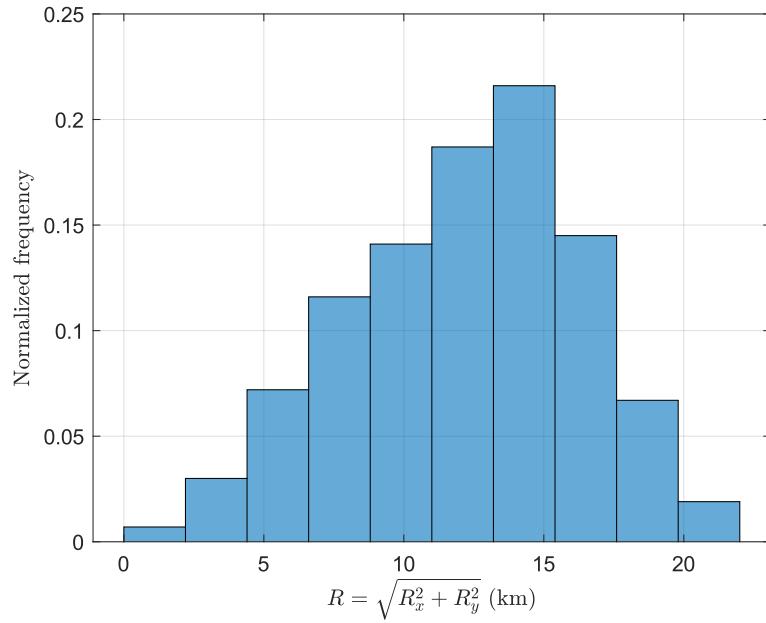
**Figure 6.9:** Response spectral analysis. The top figure shows the velocity response, and the bottom shows the acceleration response. The blue line is the spectral response of the ground motion, recorded at 20:40:15 on November 3<sup>rd</sup>, 2021, from the seismic station INS8, Insheim, Germany. 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.

## 6.2 Numerical example: SDOF system

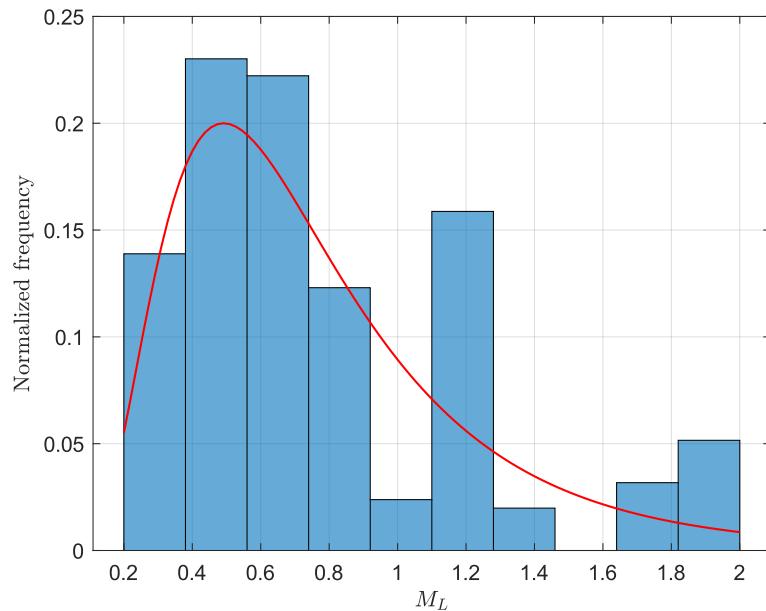
### 6.2.1 SBGMG input parameters

#### Idealization and simplification of input parameters

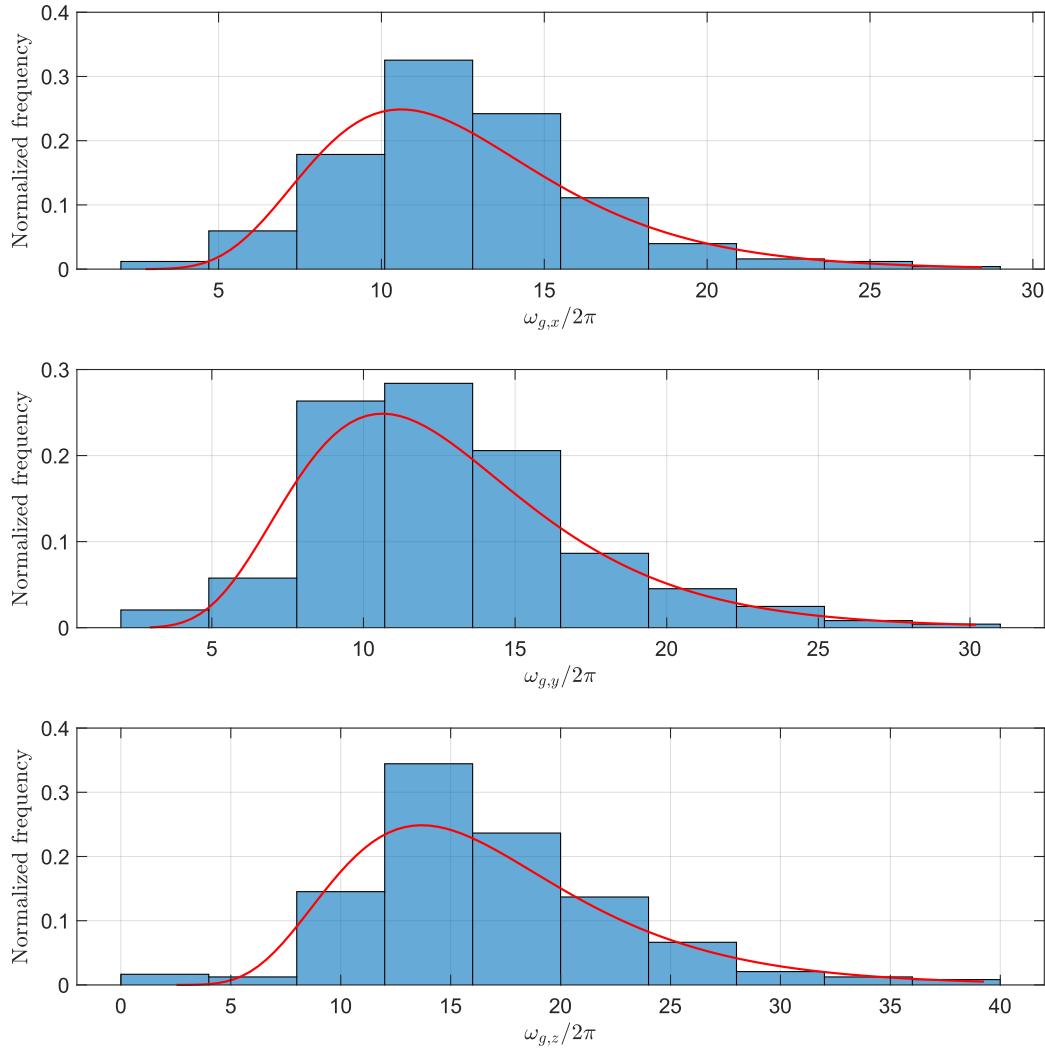
In the numerical example, we assumed the input seismic parameters for SBGMG,  $M_L, R_x, R_y, \omega_g$  and the parameters for the single DOF Mass-spring-damper model,  $\omega_{sdoF}$  and  $\zeta_{sdoF}$  are uncertain. Furthermore, the idealization and simplification of the input parameters are made in this section as the bridge to connect the practical data and the theoretical model. The



**Figure 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.



**Figure 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

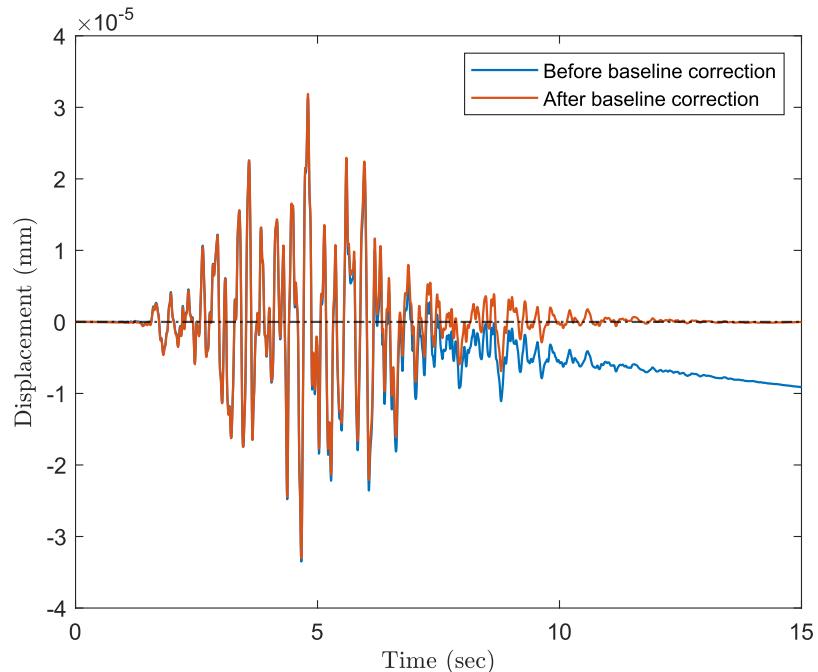


**Figure 6.12:** Histogram of predominant frequency of GM in three directions  $\omega_{g,x}$ ,  $\omega_{g,y}$  and  $\omega_{g,z}$ . The log-normal distributions are assumed for  $\omega_g$  in all directions, where the mean values, standard deviations with their 95 percent confidence interval, and p-value of a null hypothesis test, testing by Chi-square goodness-of-fit test will be presented in table 6.3

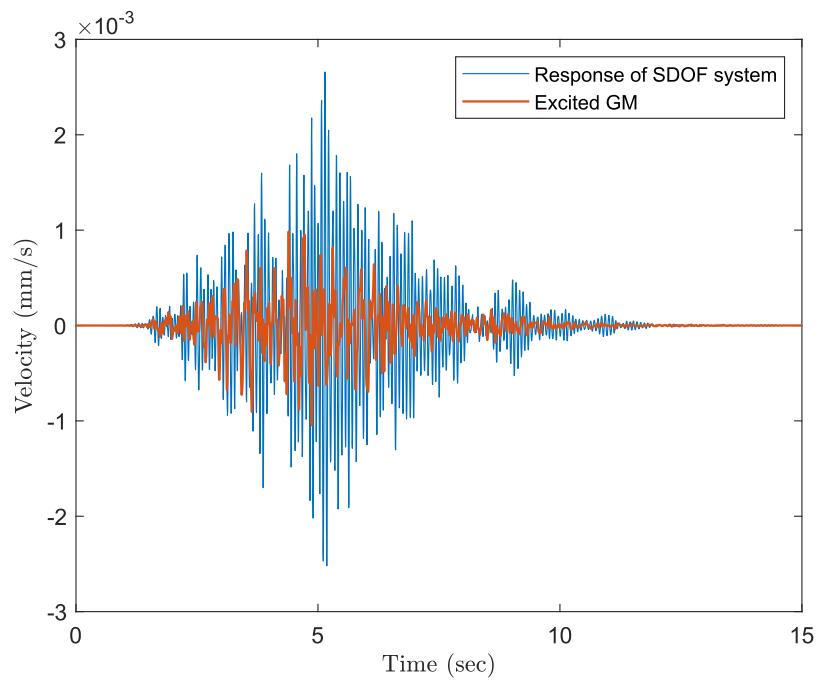
**Table 6.3:** Fitting results of  $\omega_{g,x}$ ,  $\omega_{g,y}$ ,  $\omega_{g,z}$

	Mean value $\mu$	Standard deviation $\sigma$	P-value
$\mathcal{LN}(\mu,\sigma)$ for $\omega_{g,x}/2\pi$	2.47 [2.43, 2.51]	0.33 [0.30, 0.36]	0.2050
$\mathcal{LN}(\mu,\sigma)$ for $\omega_{g,y}/2\pi$	2.49 [2.44, 2.53]	0.36 [0.33, 0.39]	0.2739
$\mathcal{LN}(\mu,\sigma)$ for $\omega_{g,z}/2\pi$	2.76 [2.71, 2.81]	0.39 [0.35, 0.42]	0.0107

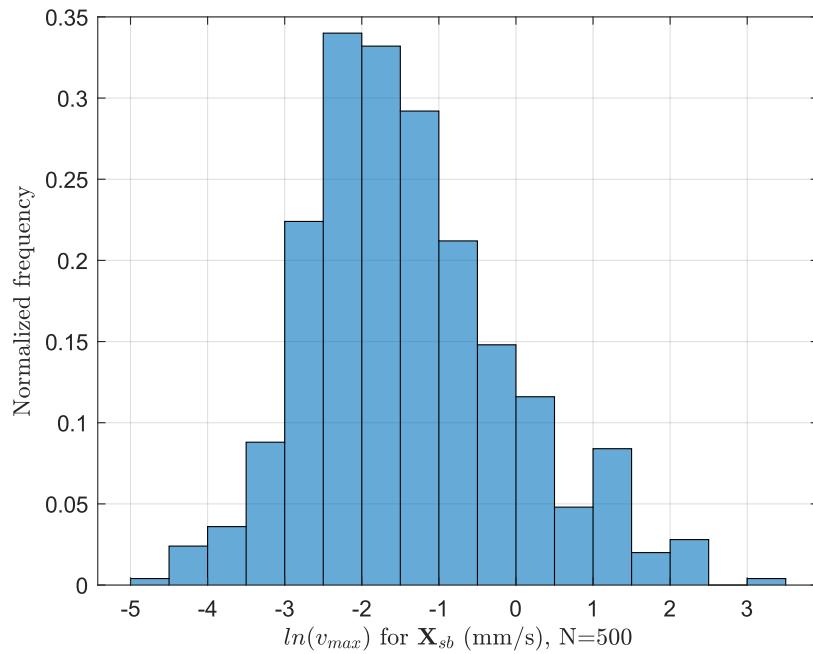
6.13). After correction, it will excite the SDOF system, and the maximum velocity in the time domain will be recorded. Figure 6.14 demonstrates the result of the system's response, where the maximum velocity is 0.0027 (mm/s). Once a sufficient number of realizations is recorded (In the numerical example, 500 realizations are recorded), the dataset of all maximum velocities in the logarithm scale and their corresponding input parameters are imported to the SPCE model for the training process. The histogram of maximum velocity in the  $z$ -direction is presented in figure 6.15, where their input parameters are generated by the LHS method following the distribution mentioned in table 6.4 and 6.5. Finally, the SPCE model can be built when all the input training data and their corresponding realization are prepared.



**Figure 6.13:** Baseline correction. The realization of ground motion is generated by SBGMG, where the input parameters  $[M_L, R_x, R_y, \omega_g/2\pi, \omega_{sdof}/2\pi, \zeta_{sdof}] = [0.74, 0.75, 14.89, 17.78, 13.47, 0.04]$ . 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.



**Figure 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.



**Figure 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.

## SPCE model configuration

The pseudo-code of configuration for the SPCE model is written in JSON format. This configuration file is crucial since it specifies all the setups of how the SPCE model should be formed and trained. In the configuration, `X_path` specifies the path for training data  $\mathcal{X}$ . Following that, the idealized distribution types of generating  $\mathcal{X}$  are includes in `X_type`, where in this case,  $x_i, i = 1 \dots 6$  are corresponding to  $M_L, R_x, R_y, \omega_g/2\pi, \omega_{sdof}/2\pi, \zeta_{sdof}$ . Additionally, `Z_type` specifies the assumed distribution of the latent variables Z. If "Gaussian" and "Uniform" are set up, the algorithm will run individually based on each assumed distribution type and choose the best one that maximizes the likelihood. In SPCE, the hybrid-LAR solver is commonly chosen for `Solver` due to its efficiency. `max_poly_degree` in the configuration defines the maximum polynomial degree of the SCPE model. Commonly, the higher polynomial degree can generate a more accurate model, but it is more time-consuming. In [Zhu and Sudret 2023], "`max_poly_degree`: 4" is normally be defined. Finally,  $q_{norm}$  is the index for coefficient reduction, which the valid  $q_{norm}$  ranges in  $(0, 1]$ . The detailed function of  $q_{norm}$  is introduced in chapter 3.

**Listing 6.1:** Pseudocode of SCPE configuration

```
"X_path": "./StochasticPCE/InputData/Numerical_Example_X.mat",
"X_type": {"x1": "Gaussian",
           "x2": "Uniform",
           "x3": "Uniform",
           "x4": "Gaussian",
           "x5": "Gaussian",
           "x6": "Gaussian"},

"OrthPolyParams": {},

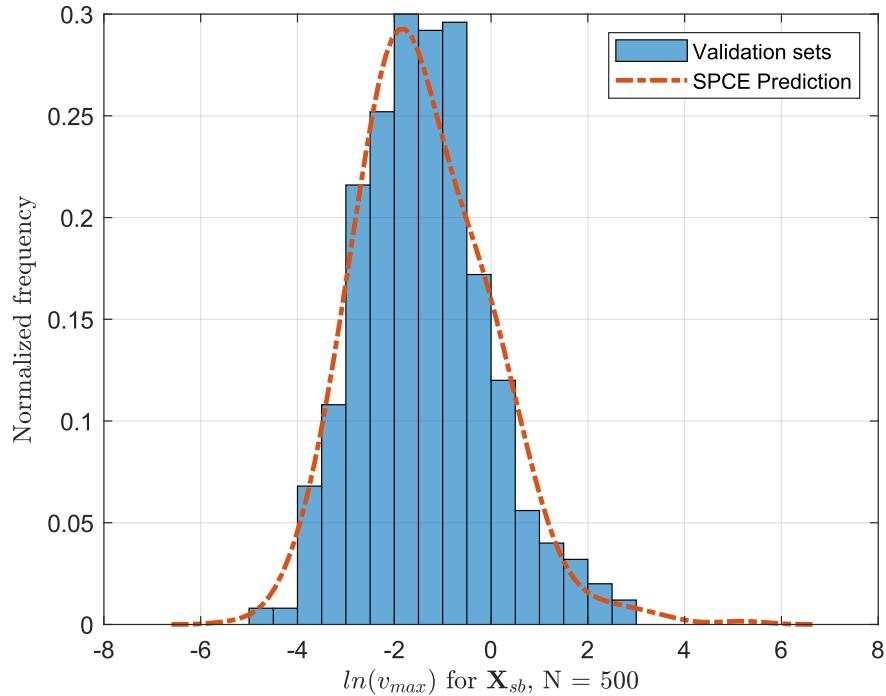
"Z_type": ["Gaussian", "Uniform"],

"Solver": "HybridLAR",

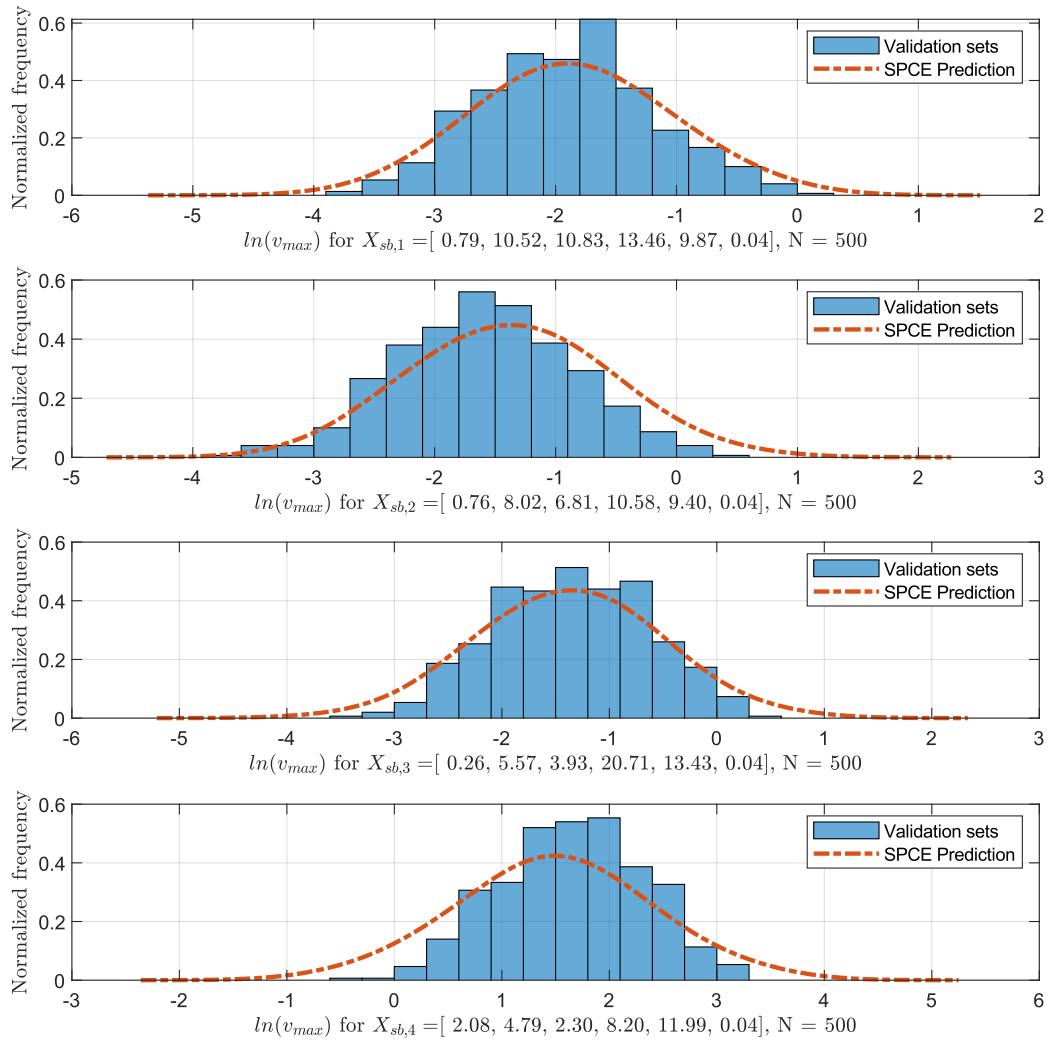
"max_poly_degree": 4,
"q_norm": [0.5, 0.75, 1],
"save_result_path": "./StochasticPCE/Training_result"
```

## Validation

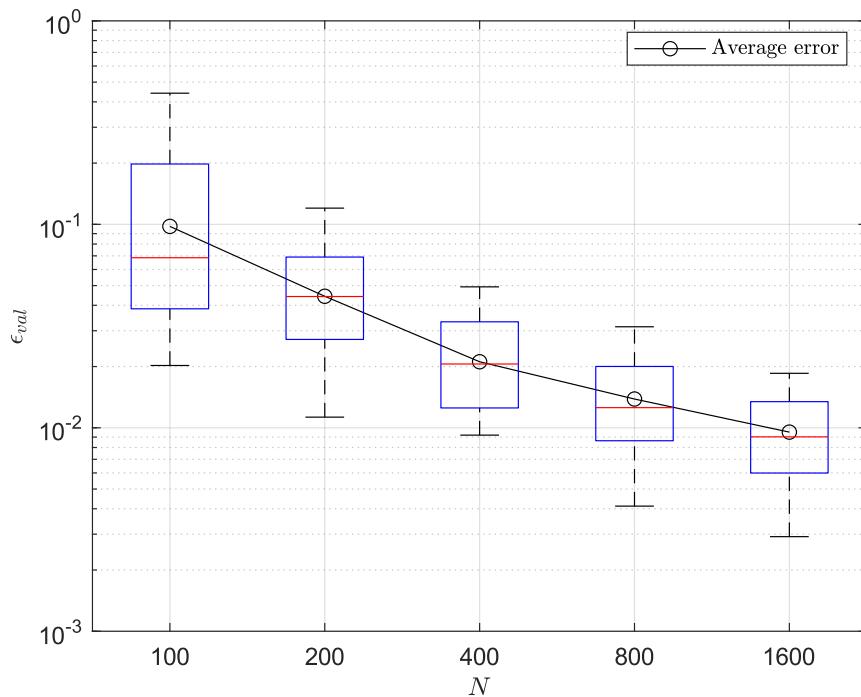
The prediction and validation data are compared to validate the SPCE model and understand its prediction ability. Two types of comparison are done for validation, where the input parameters are  $\mathbf{X}_{sb}$  following the distribution shown in table 6.5 and 6.4 in  $z$ -direction. To test SPCE performance, using different input datasets for training  $\mathcal{X}_{tr}$  and validation  $\mathcal{X}_{val}$  is essential. First, global validation is performed by inputting whole  $\mathcal{X}_{val}$  (figure 6.16), then its local accuracy is validated on randomly selecting four sample points  $\mathcal{X}_{val_i}, i = 1, \dots, 4$  in the design space. The predictive results that respond to those specific sample points are examined (figure 6.17). Once the prediction is made, their response PDFs are computed and compared with the validation data. Subsequently, the error will be calculated by the manner in equation (3.48). The validation process repeats the prediction 30 times and collects the error in each run. The mean and standard deviation of the validation error recorded from 30 runs are shown in table 6.6. It is assumed that the sufficient small  $\varepsilon_{val}$  is less than 0.05 to get enough accurate results.



**Figure 6.16:** Comparison of validation data and SPCE prediction of maximum velocities,  $\mathbf{X} = X_{sb}$ .



**Figure 6.17:** Comparison of validation data and SPCE prediction of maximum velocities at sample points  $X_{sb,i}, i = 1 \cdots 4$ .



**Figure 6.18:** Comparison of the convergence of the surrogate model with different training size  $N$ .

#### 6.2.4 Runtime comparison of SPCE

As mentioned in [Zhu and Sudret 2023], the training process is time-consuming. This might be due to four reasons:

1. The order of the polynomial degrees
2. The number of training data
3. The number of coefficients of the SPCE surrogate model
4. The usage of the nested optimization process

To accelerate it, the parallel computing technique is utilized. The Scikit-optimize toolbox has an in-built parallel computing tool for the Gaussian process optimizer, which is suitable for the current algorithm.

This chapter compares the time affected by these four potential reasons. The hyperparameters testing are the number of training data, polynomial degrees, and running in single-core / multi-cores. On the other hand, the number of coefficients will be computed automatically

depending on the hyperparameters and the hybrid-LARS solver. Furthermore, the latent variable Z is assumed to follow the Gaussian distribution when comparing the runtime.

### Implementation of parallel computing

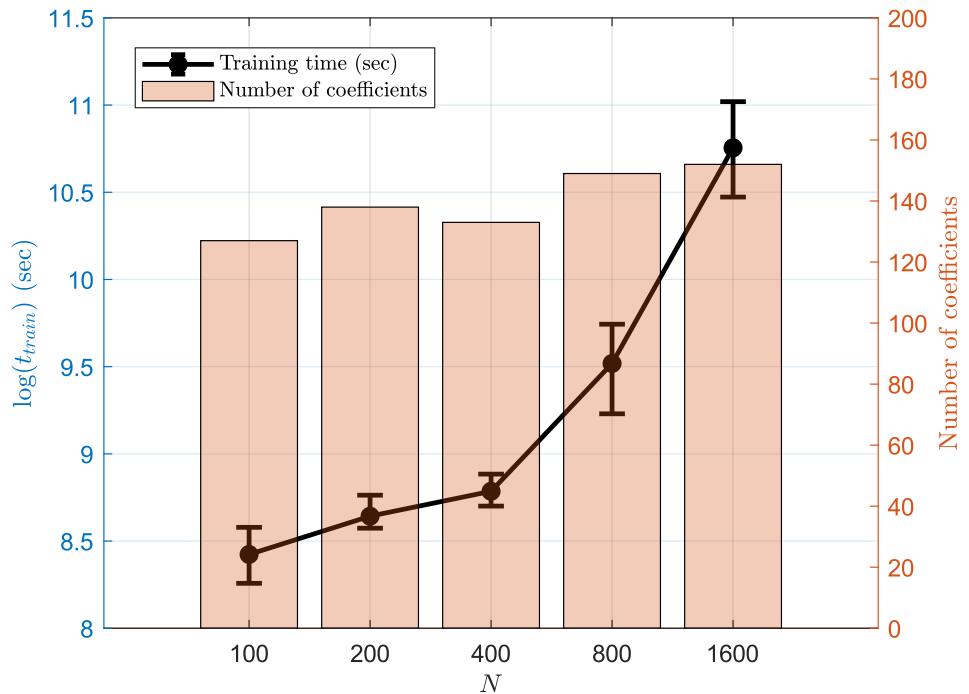
When performing the Gaussian process optimization to fit the maximum likelihood, the algorithm will explore and exploit several data points in the searching domain, and the observation computed at corresponding data points will then be utilized to form the acquisition function and for further optimization process. This process is time-consuming and repetitive, which could be accelerated by computing parallelly in several cores instead of running sequentially. The pseudo-code of the implementation, written in Python, is as follows:

```

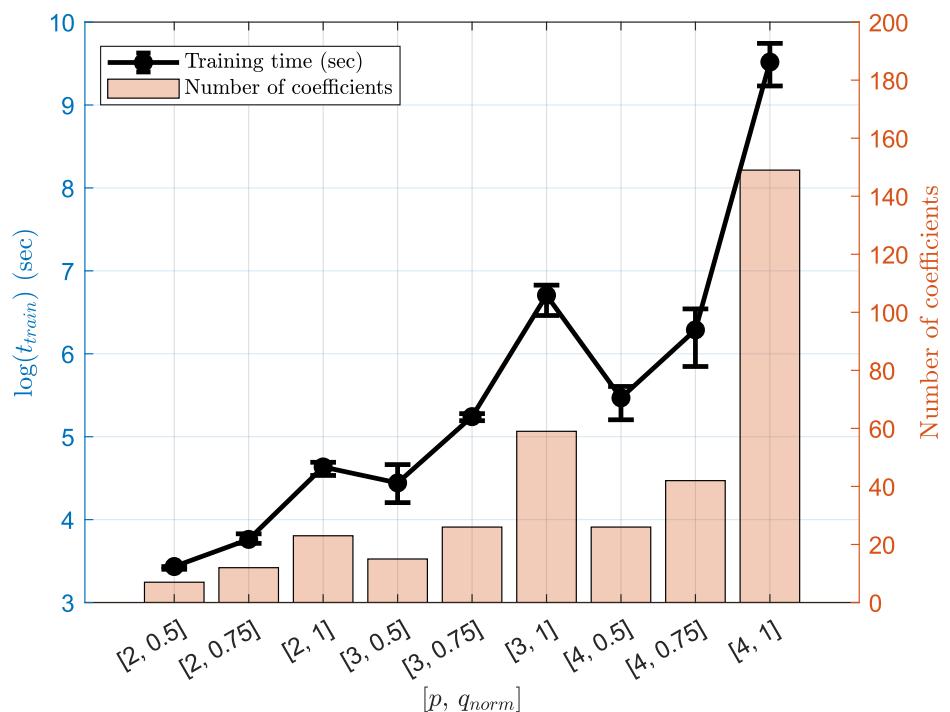
1   # Initialized the list for coeffs and optimizer
2   coeffs_SPCE = []
3   optimizer = Optimizer(
4       # Define searching bounds
5       dimensions=[lower_bound, upper_bound, name='Sigma')],
6       # Using Gaussian Process estimator
7       base_estimator='gp',
8       # Using the Latin hypercube sampling technique for DOE
9       initial_point_generator = 'lhs'
10      )
11
12  # Start optimization process, run for 10 iteration
13  for i in range(10):
14      # Evaluate 5 sample points in each iteration
15      x = optimizer.ask(n_points=5)
16      # Allocate 5 computations into 5 cores
17      # GP_Objfun is the objective function
18      y_Parallel = Parallel(n_jobs=5)(delayed(GP_Objfun)(v) for v in x)
19      # Collect results in each core parallelly
20      y_gp_obj, coeffs = zip(*y_Parallel)
21      coeffs_SPCE.extend(coeffs)
22      # Return the best result and coeffs
23      optimizer.tell(x, y_gp_obj)

```

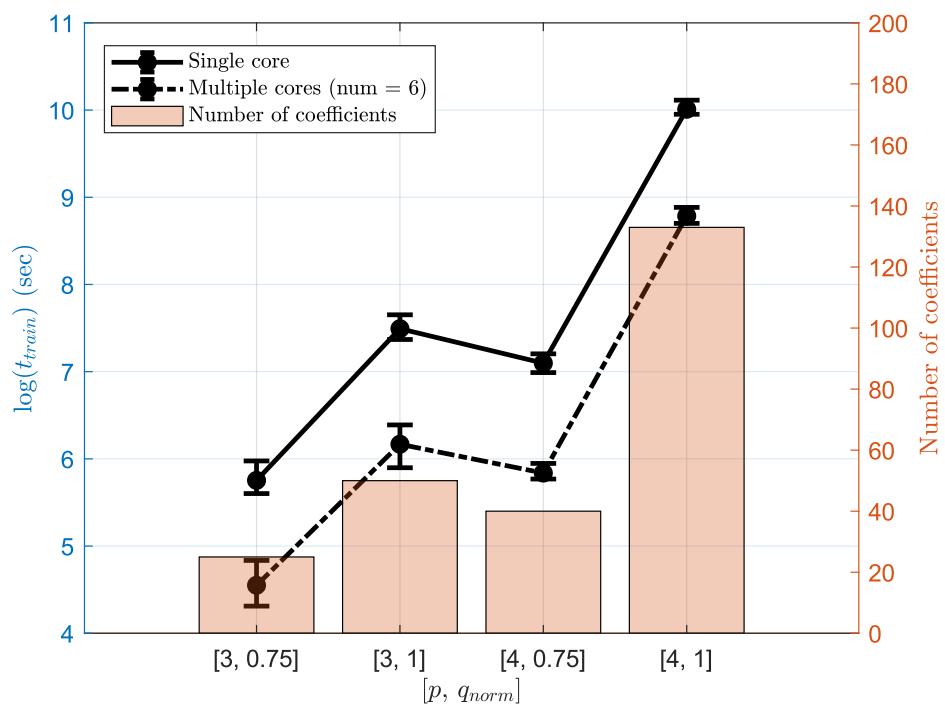
**Listing 6.2:** Pseudocode of parallel computing implementation



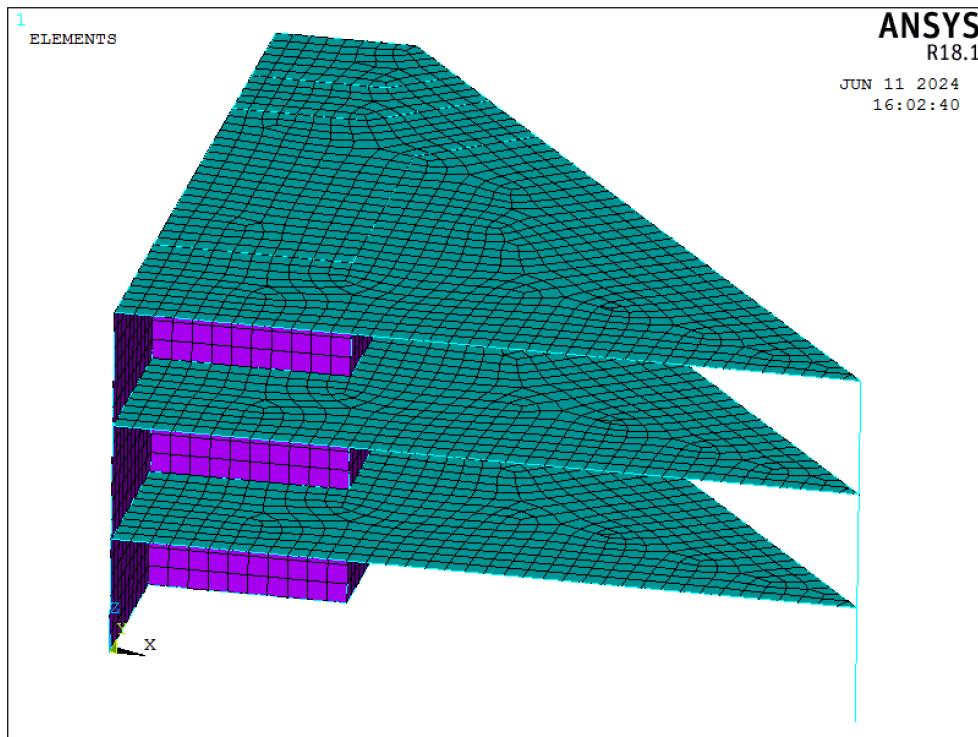
**Figure 6.19:** Time comparison of different sizes of training data  $N$ ,  $p = 4$ ,  $q_{norm} = 1$



**Figure 6.20:** Time comparison of different combinations of  $p$  and  $q_{norm}$



**Figure 6.21:** Time comparison of running in single/multiple cores.



**Figure 7.1:** Building Geometry at geothermal energy site in Taufkirchen. 1<sup>st</sup> Perspective.

**Table 7.1:** Material properties of building

**Building material properties**

Density $\rho_b$	2500	$kg/m^3$
Poisson's ratio $\nu_b$	0.27	-
Young's modulus $E_b$	$\sim \mathcal{N}(3 \cdot 10^{10}, 1.5 \cdot 10^9)$	$N/m^2$
Modal damping ratio $\zeta_b$	$\sim \mathcal{LN}(-3.2, 0.1)$	-

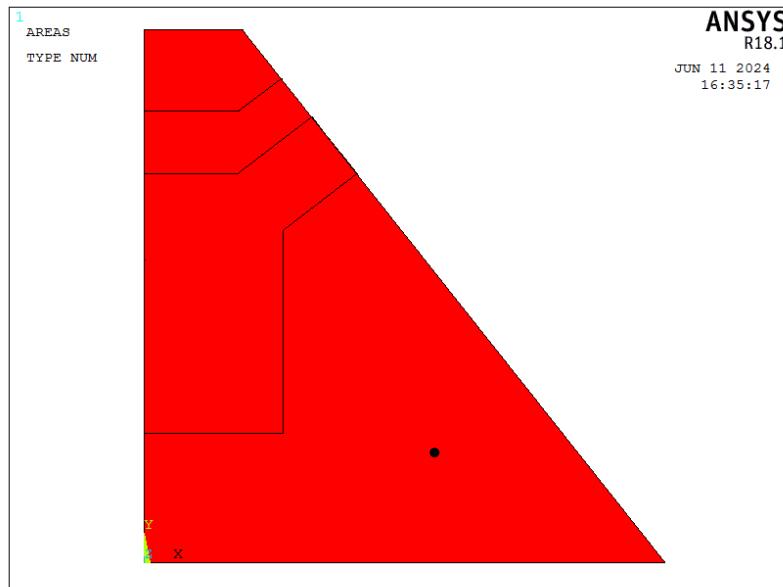
**Table 7.2:** Soil properties

**Soil material properties**

Density of soil $\rho_s$	2300	$kg/m^3$
Poisson's ratio of soil $\nu_s$	0.33	-
Shear wave velocity $v_s$	450	$m/s$

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.

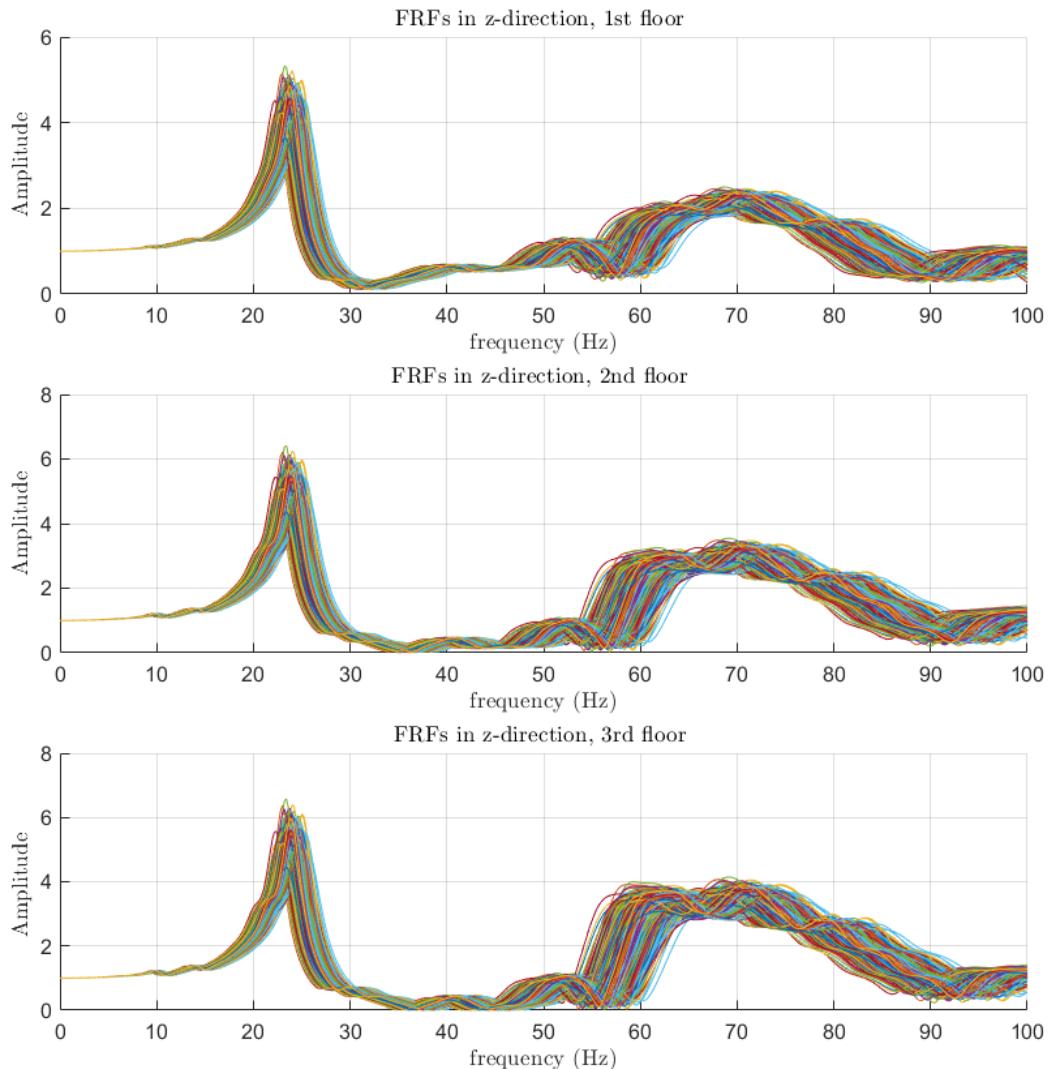


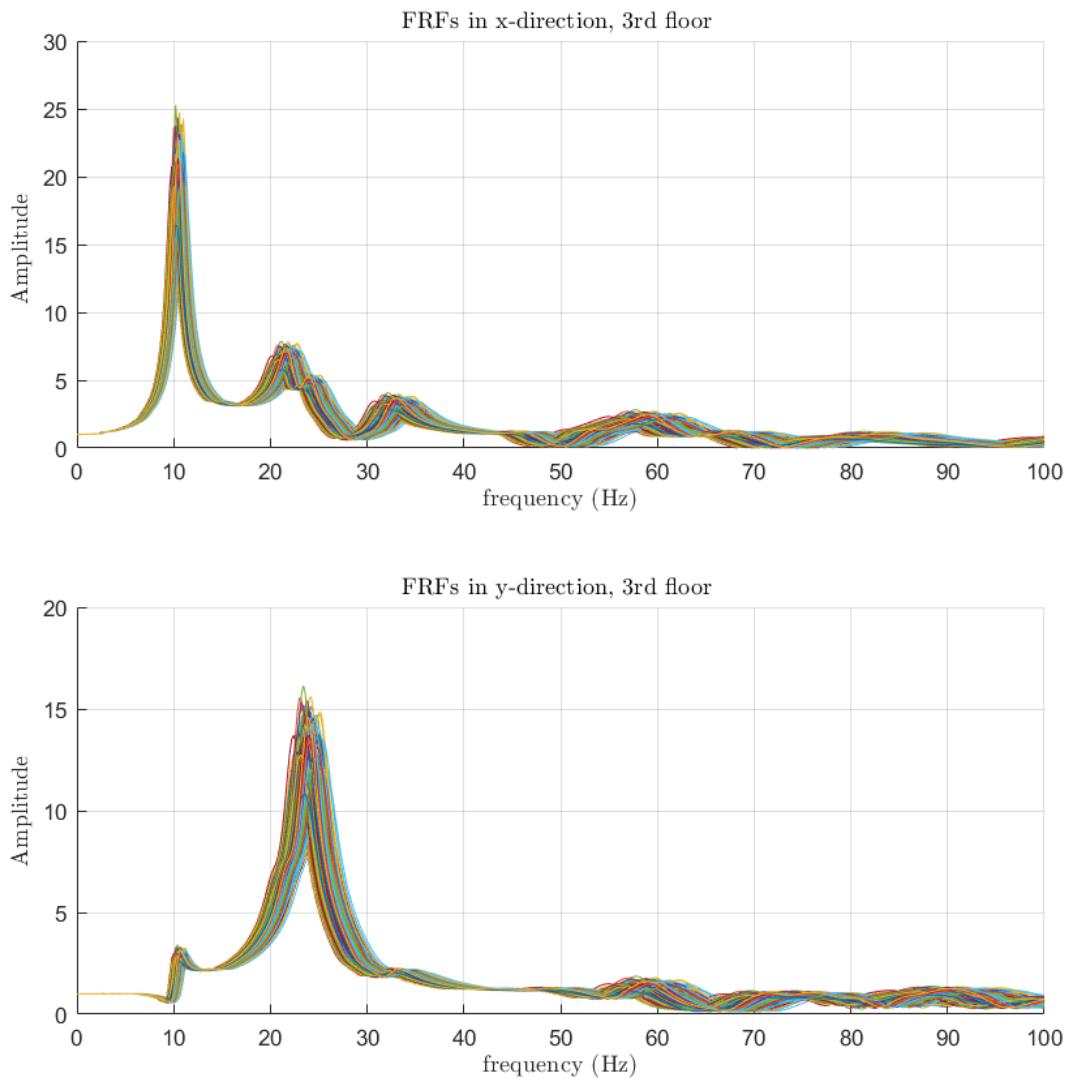
**Figure 7.2:** Top view of the building, the black lines inside the red area are the placement of walls. The black dot is the target point recording the FRFs, whose coordinate is [8.75, 4.8] (m) for x and y directions; the z coordinates correspond to the height of each floor.

**Table 7.4:** Computer specification for simulation

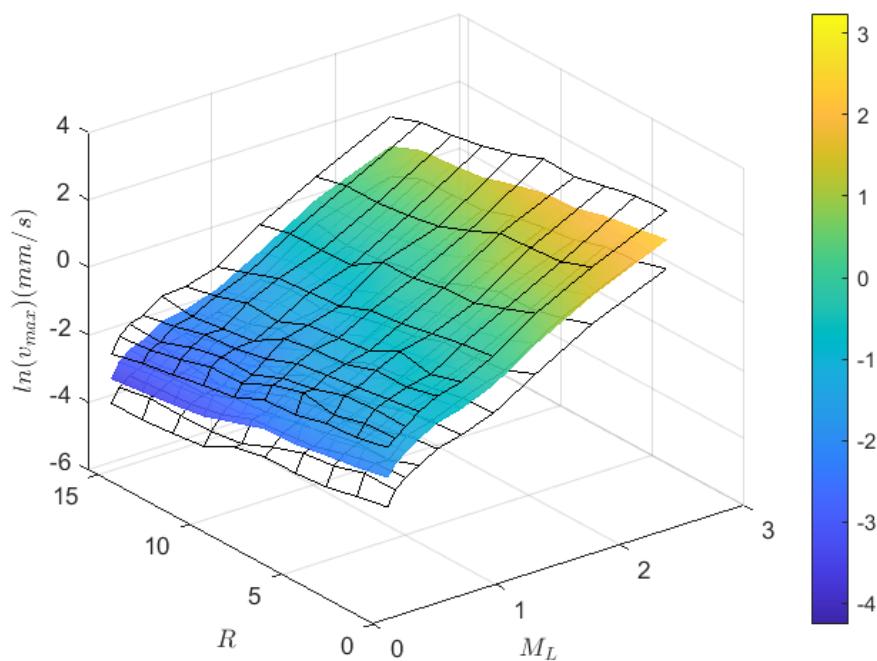
Computer specification for simulation

CPU	AMD FX-8150 Eight Core 8x
CPU speed	3.60 GHz
RAM	32.0 GB
OS version	Windows 11 64-bit

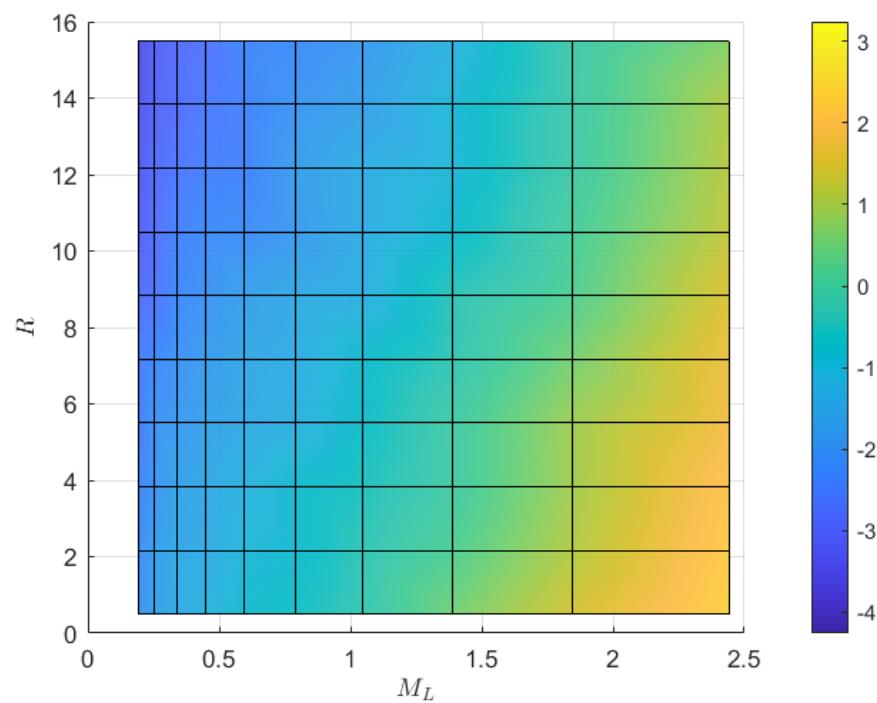
**Figure 7.3:** 500 frequency response function of example buildings in the vertical direction at each floor, where Young's modulus and damping ratio are uncertain, following the probability distribution shown in table 7.1. The parameter for SSI is included in table 7.2.



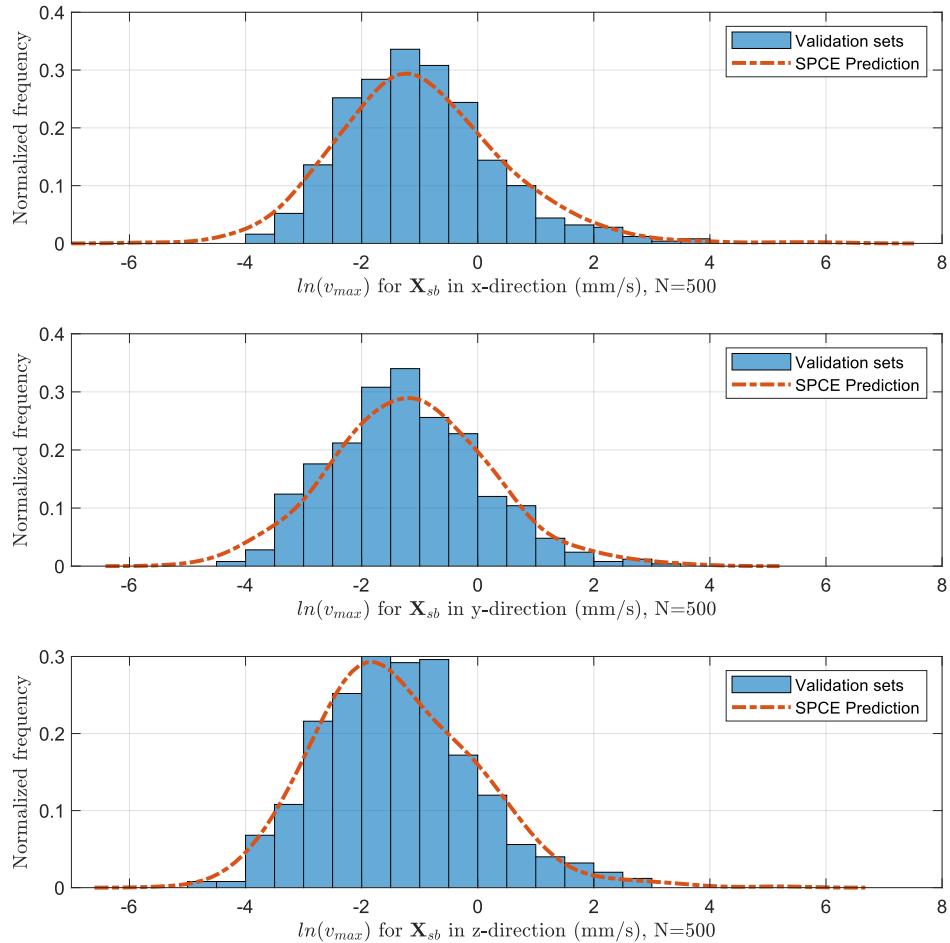
**Figure 7.4:** 500 frequency response function of example buildings in both horizontal directions at the top-most floor, where Young's modulus and damping ratio are uncertain, following the probability distribution shown in table 7.1. The parameter for SSI is included in table 7.2.



**Figure 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



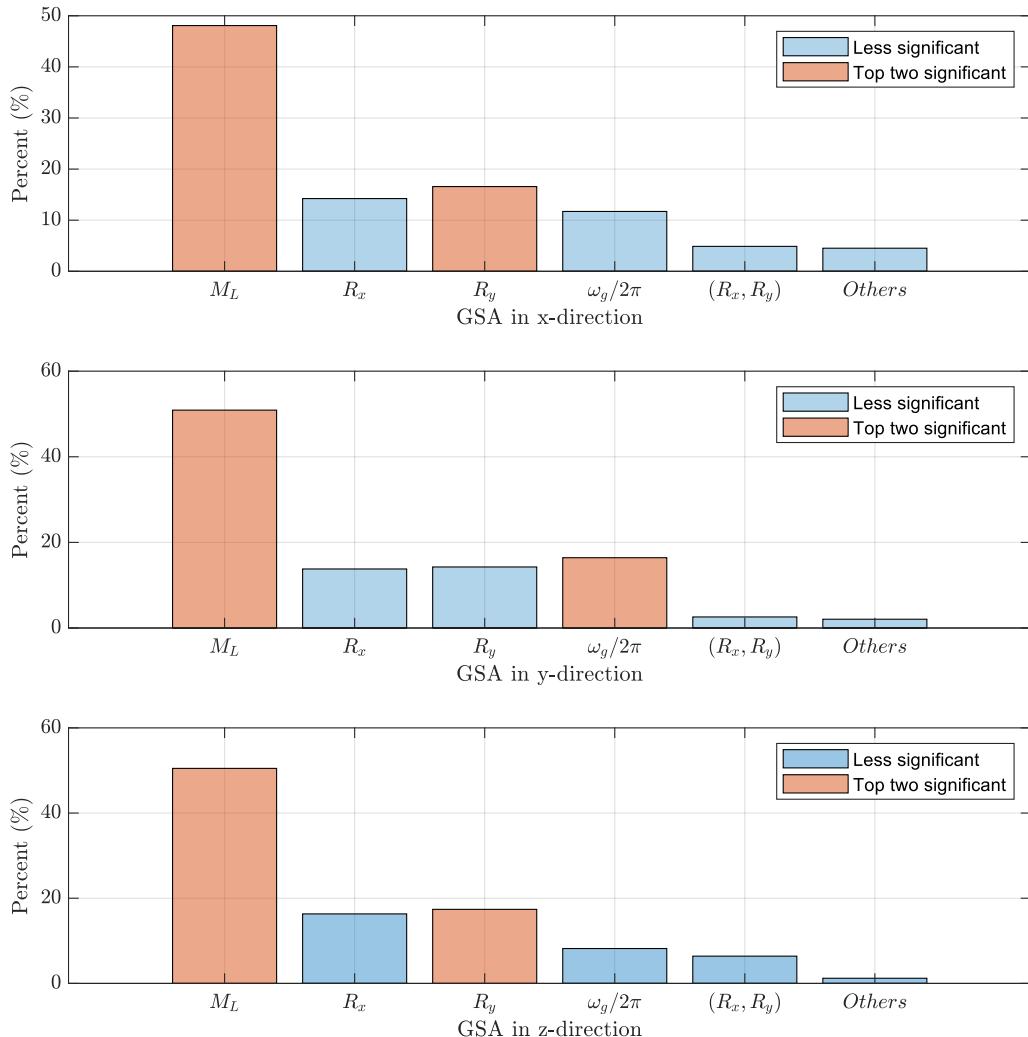
**Figure 7.6:** 2D Response function built by SPCE against  $M_L$  and  $R$  in z-direction



**Figure 7.7:** Comparison of validation data and SPCE prediction of maximum velocities in 3 directions

**Table 7.5:** Global  $\varepsilon_{val}$  in three directions, 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

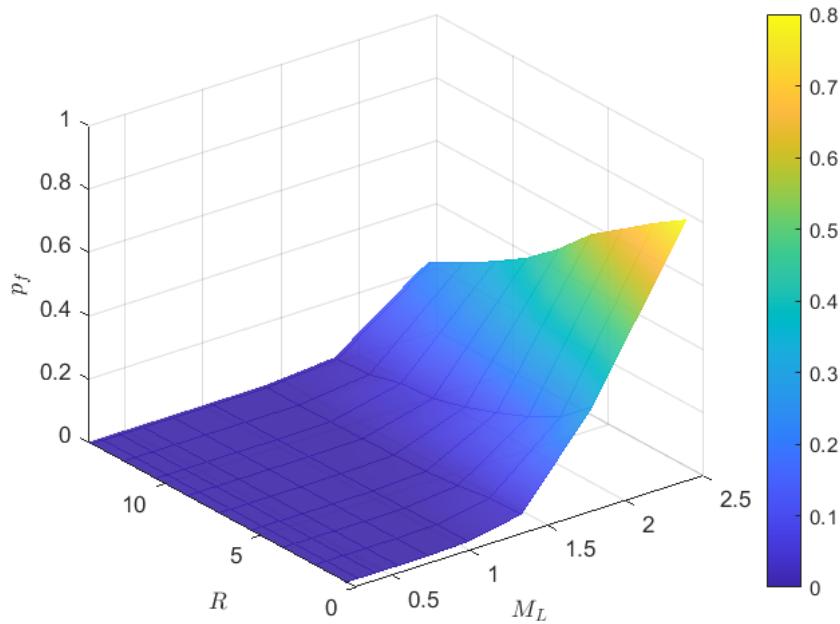
	$\mathbb{E}[\varepsilon_{val}]$	$\sigma$
$\varepsilon_{val}$ in x-direction	0.04266	0.02775
$\varepsilon_{val}$ in y-direction	0.02287	0.00941
$\varepsilon_{val}$ in z-direction	0.02186	0.00873



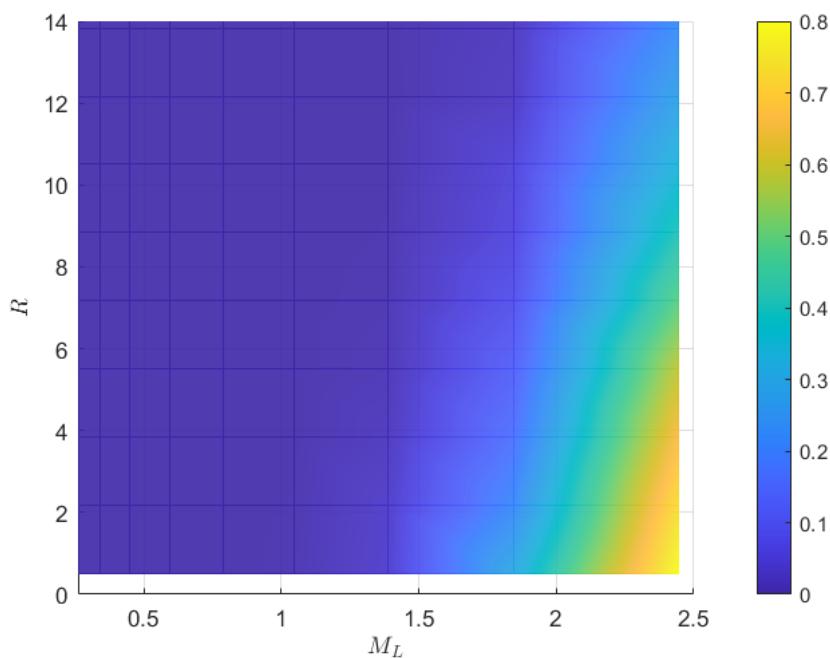
**Figure 7.8:** QoI-based Sobol's indices in three directions. The orange bars indicate the top two significant parameters where the large SPCE coefficients are set. The blue bars are the comparably less significant parameters affecting the system. Here, the label " $(R_x, R_y)$ " represents the higher-order index, and the label "Others" is the remaining indices that are insignificant.

## 7.4 Serviceability and comfortability assessment

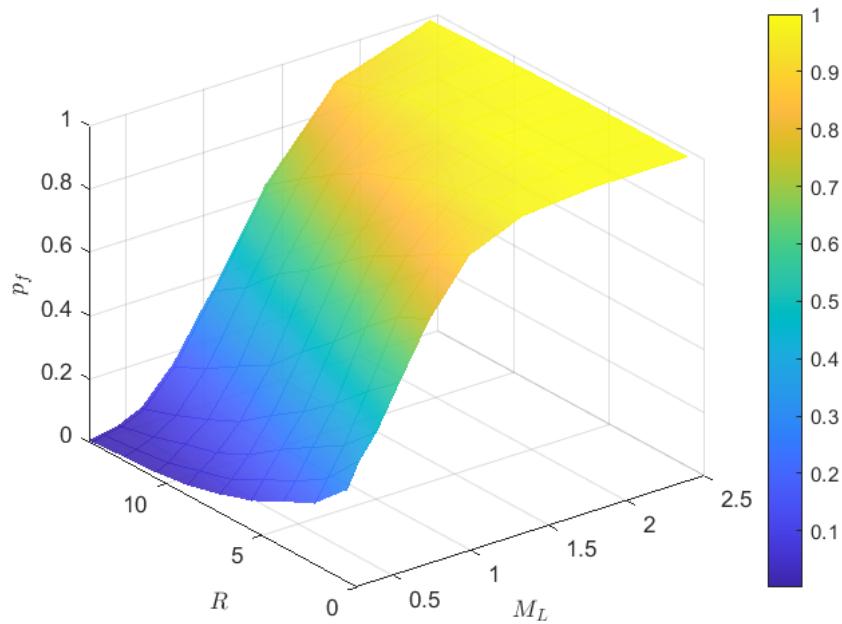
Serviceability and comfortability assessments are performed in this section. The criteria are chosen by assuming the building is in a purely residential area. For DIN-4150-2,  $A_0 = 3$  and  $A_0 = 0.2$  are chosen as criteria for events in day and night, where the empirical constant  $c_F$  is set as 0.8. Furthermore,  $v_{max} = 15$  (mm/s) and  $v_{max} = 20$  (mm/s) are selected in



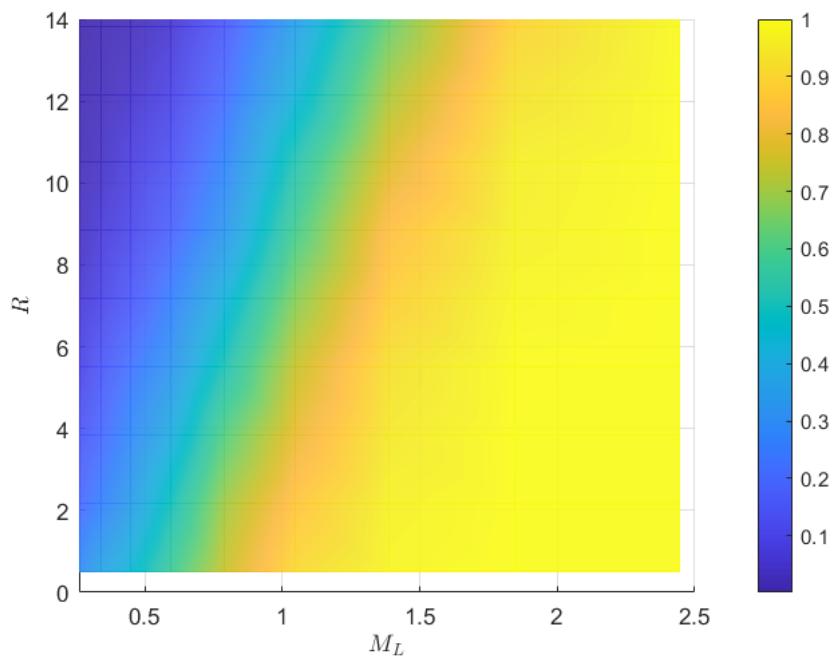
**Figure 7.9:** 3D Fragility curve for Comfortability assessment, following the standard of DIN-4150-2 for daytime, where  $A_0 = 3$  and  $c_F = 0.8$ . The z-axis describes the probability  $p_f$  of whether the response will exceed the limit state or not.



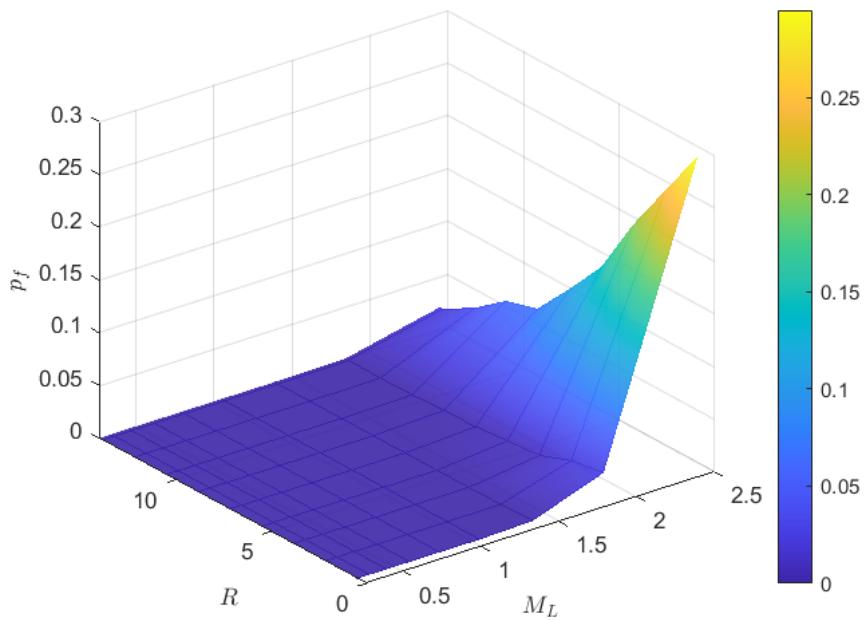
**Figure 7.10:** 2D Fragility curve for Comfortability assessment, following the standard of DIN-4150-2 for daytime, where  $A_0 = 3$  and  $c_F = 0.8$ . The z-axis describes the probability  $p_f$  of whether the response will exceed the limit state or not.



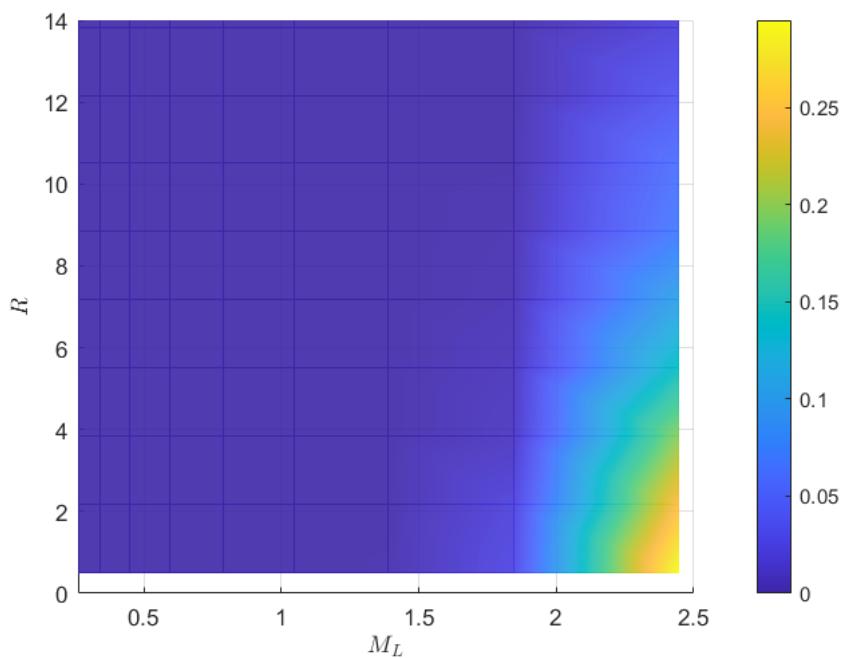
**Figure 7.11:** 3D Fragility curve for Comfortability assessment, following the standard of DIN-4150-2 for night, where  $A_0 = 0.2$  and  $c_F = 0.8$ . The z-axis describes the probability  $p_f$  of whether the response will exceed the limit state or not.



**Figure 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.



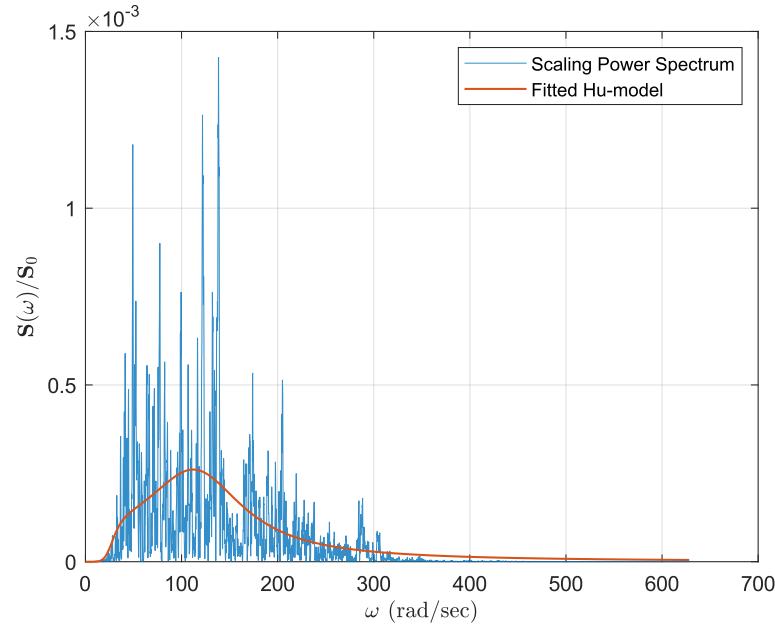
**Figure 7.13:** 3D Fragility curve for Serviceability assessment, following the standard of DIN-4150-3, where  $v_{max} = 15$  (mm/s). The z-axis describes the probability  $p_f$  of whether the response will exceed the limit state or not.



**Figure 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.

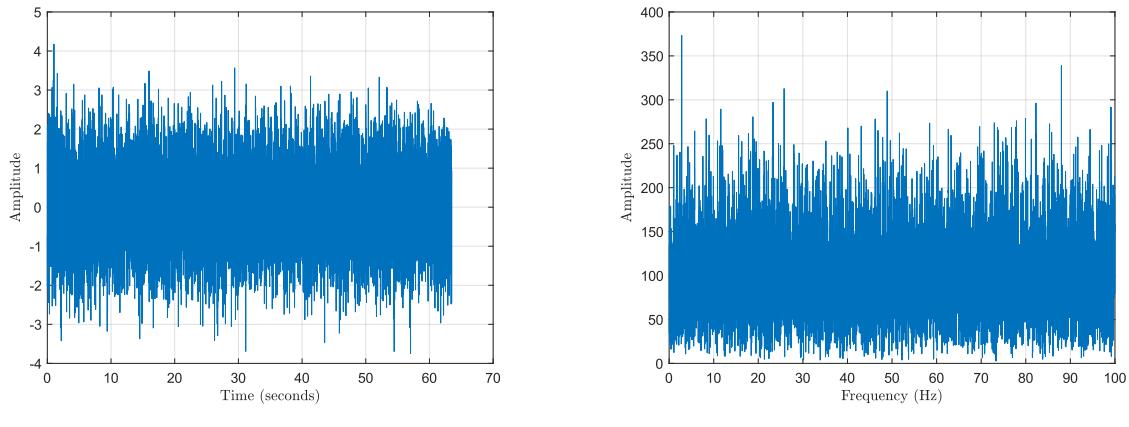
## A.2 Example flowchart of artificial ground motion generator with fitting procedure

1. Fit the Hu model with a recorded ground motion by the nonlinear least square method. Return the optimal model parameters  $\mathbf{X}_{a,opt}$  and its corresponding FRF.



**Figure A.1:** Step 1: Fitting the FRF of Hu model with recorded ground motion

2. Generate the GWN and transfer it from the time domain to the frequency domain by FFT. This GWN is artificial rock acceleration assumed by the Hu model.

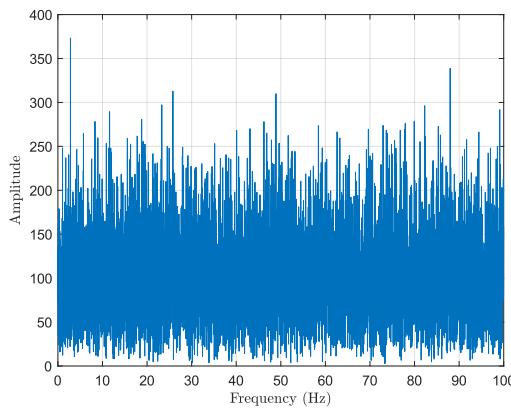


**(a)** The GWN in the time domain

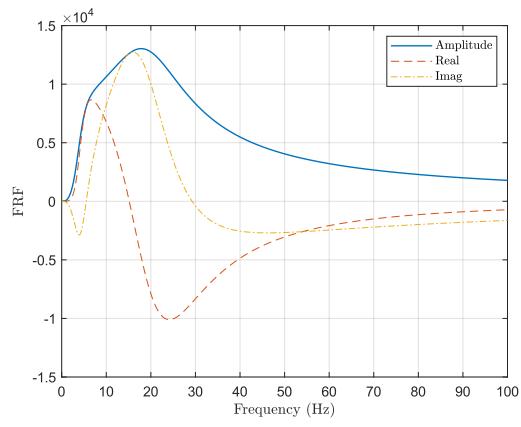
**(b)** The GWN in the frequency domain

**Figure A.2:** Step 2: Transfer the GWN from the time domain to the frequency domain.

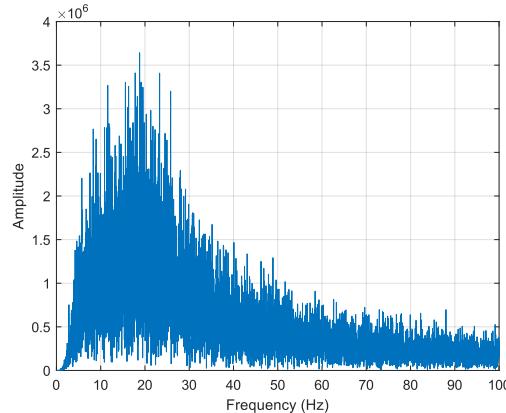
3. Dot product the GWN in the second step with FRF generated in the first step. This process can also be done using the convolution technique in the time domain. Here, the artificial ground motion of the Hu model in the frequency domain is formed. Then, IFFT transfers this artificial ground motion back to the time domain. It is now spectral and temporal stationary.



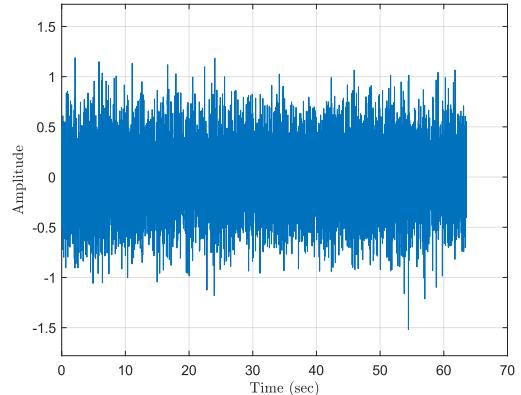
**(a)** The GWN in the frequency domain



**(b)** FRF of Hu model formed by  $\mathbf{X}_{a,opt}$



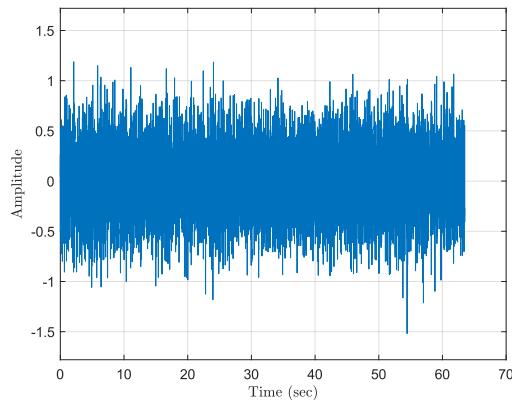
**(c)** Dot product of the GWN and the FRF



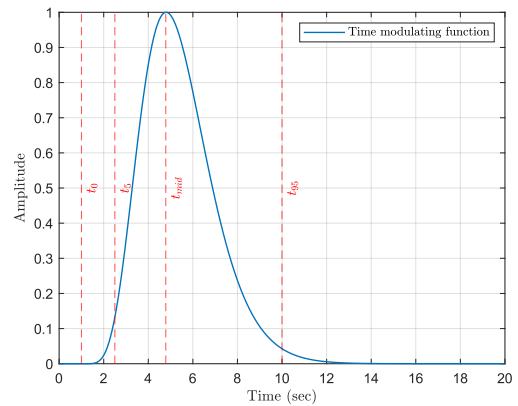
**(d)** Transfer the result back to the time domain

**Figure A.3:** Step 3: Generate the ground motion by the Hu model.

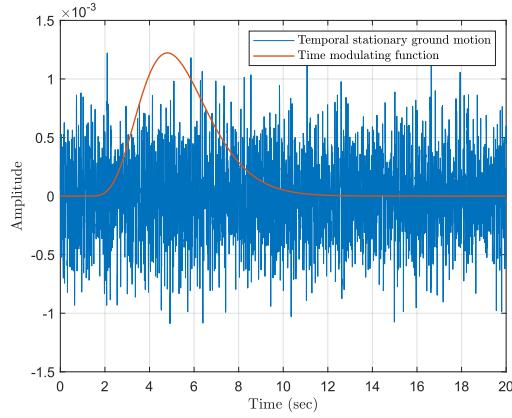
4. Dot product again this spectral and temporal stationary ground motion with time modulating function to make it temporal non-stationary.



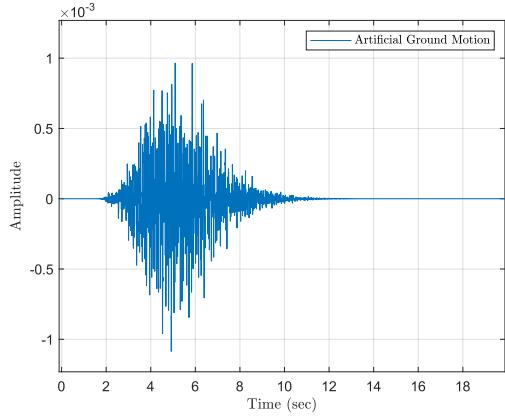
**(a)** Ground motion from the Hu model in the time domain



**(b)** Time modulating function formed by time information



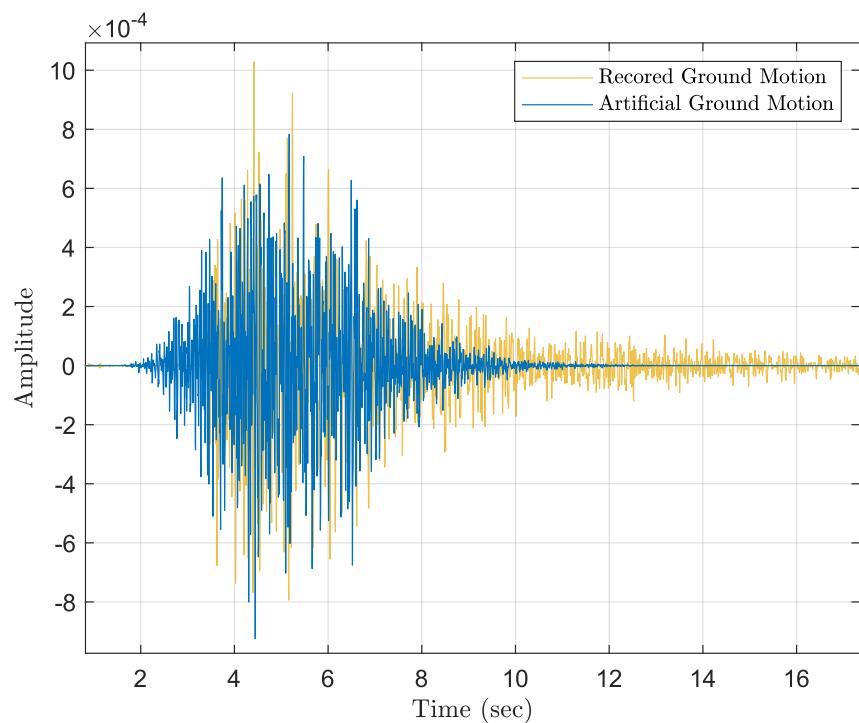
**(c)** Dot product of ground motion and time modulating function



**(d)** Generate the artificial ground motion with temporal non-stationary

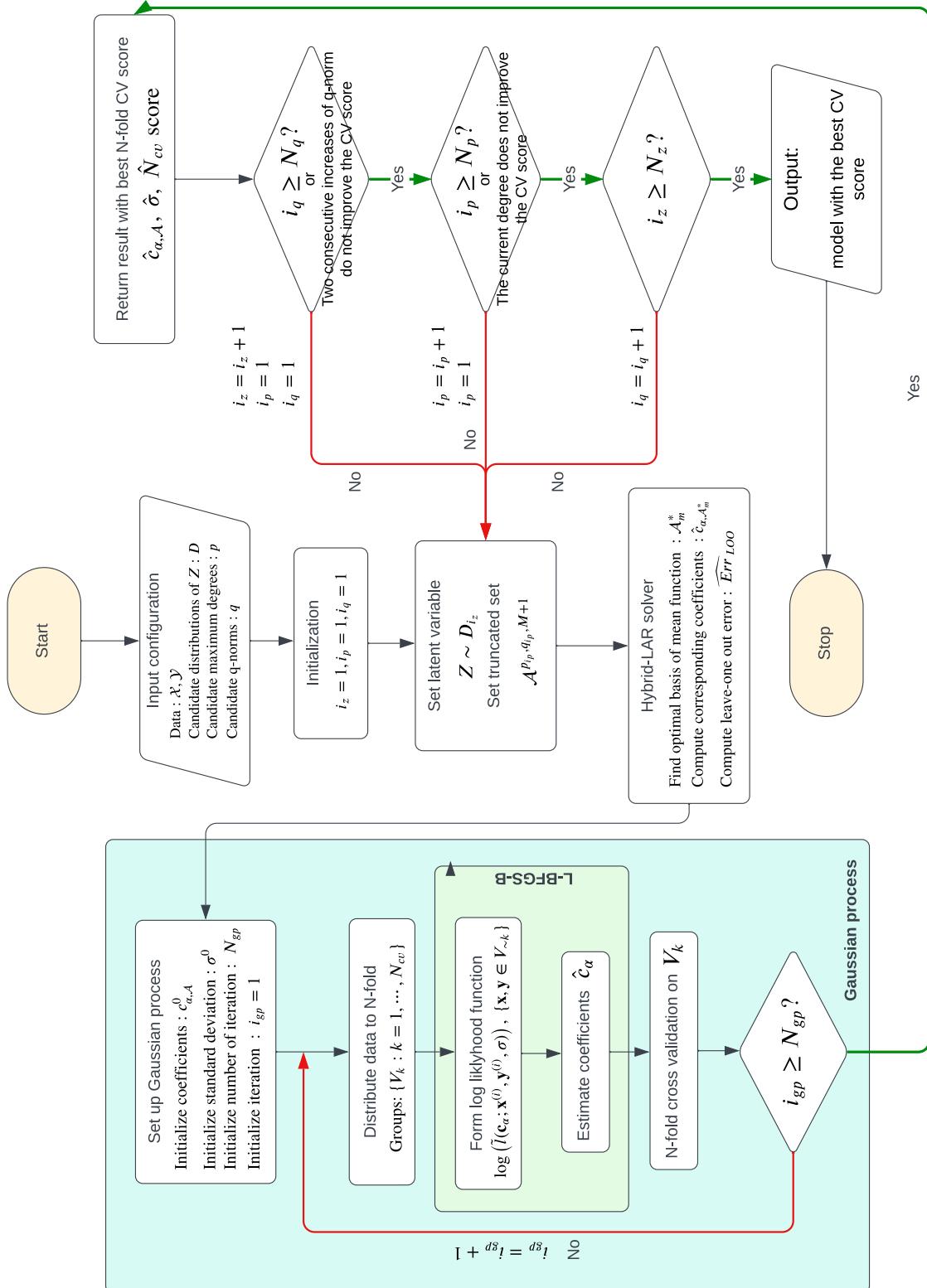
**Figure A.4:** Step 4: Generate the artificial ground motion based on Hu model and time modulating function

5. Compare it with recorded ground motion.



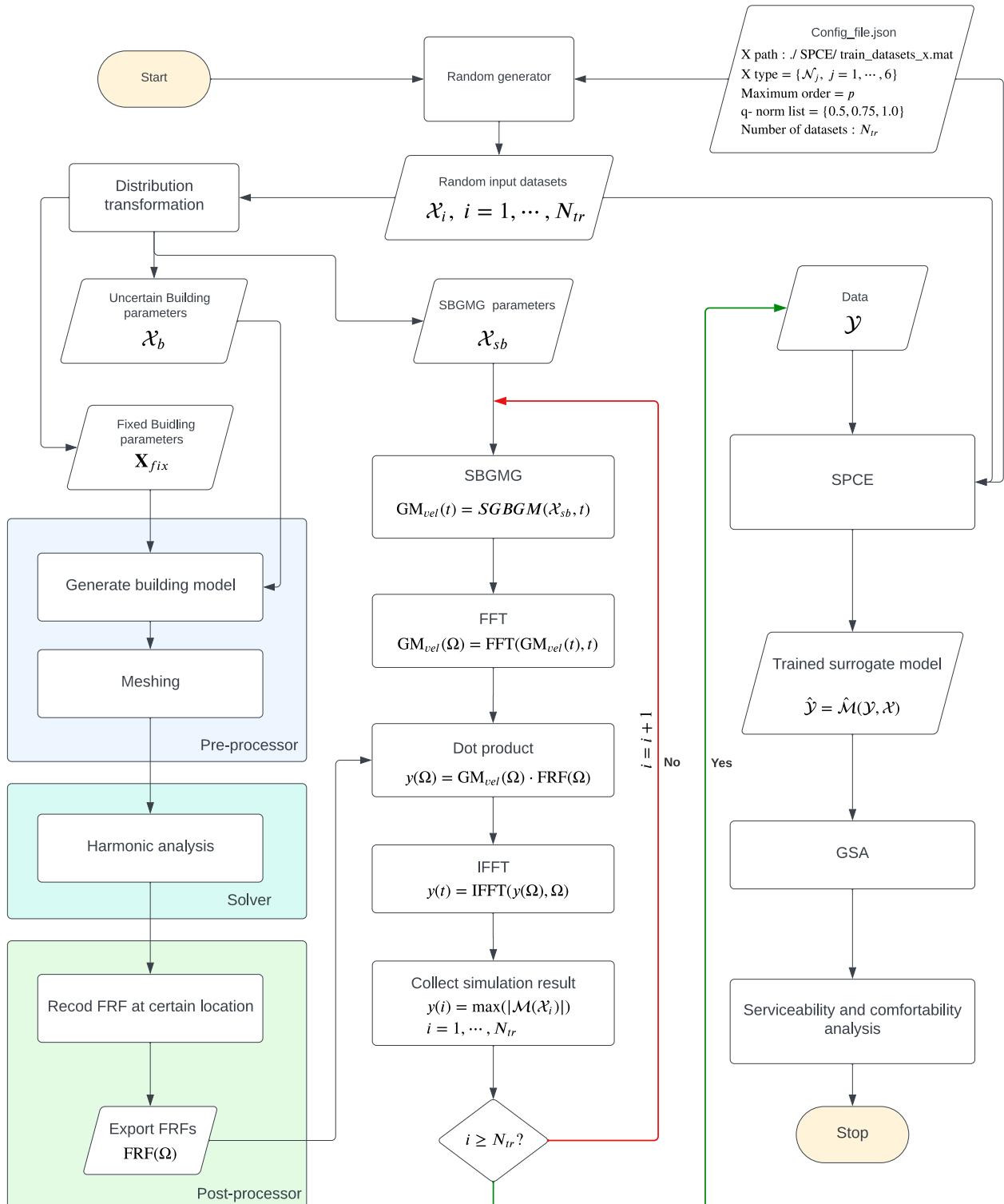
**Figure A.5:** Step 5: The comparison of artificial ground motion and recorded one.

### A.3 Stochastic polynomial chaos expansion framework



**Figure 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].

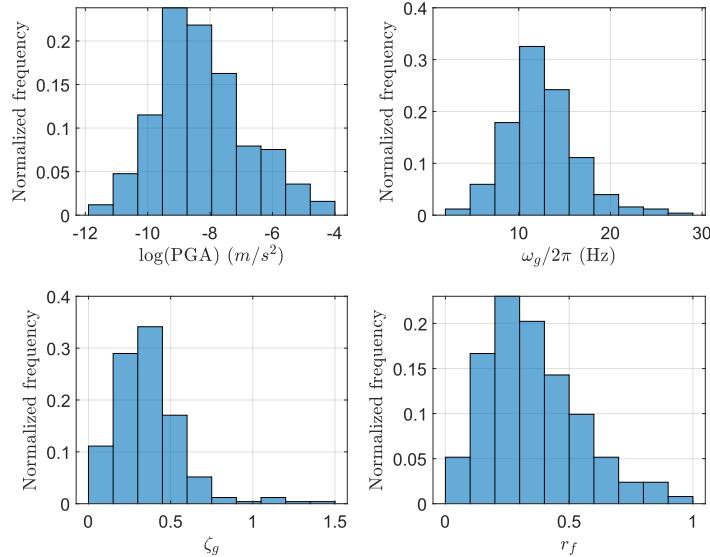
## A.4 Overall flowchart of simulation and training



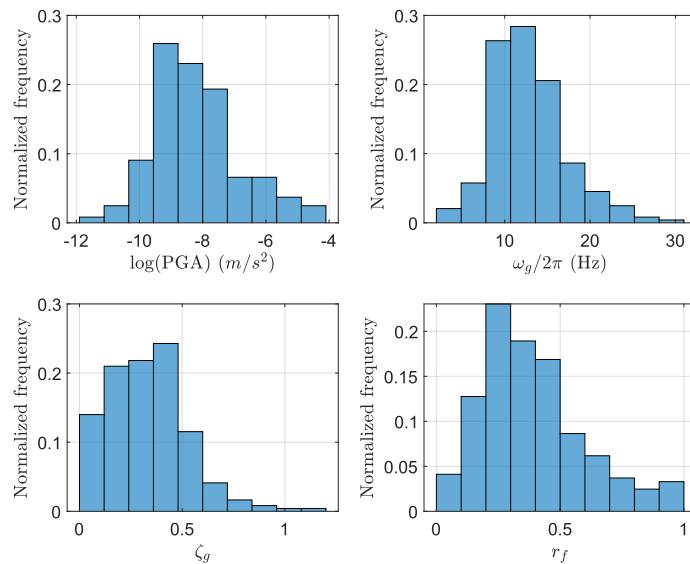
**Figure A.7:** Overall flowchart of the simulation and training. The simulation has three parts: the pre-processor, solver, and post-processor. The modelings and simulations for building are run in ANSYS ®.

## A.5 Additional plots of numerical example

### A.5.1 Histogram of model parameters

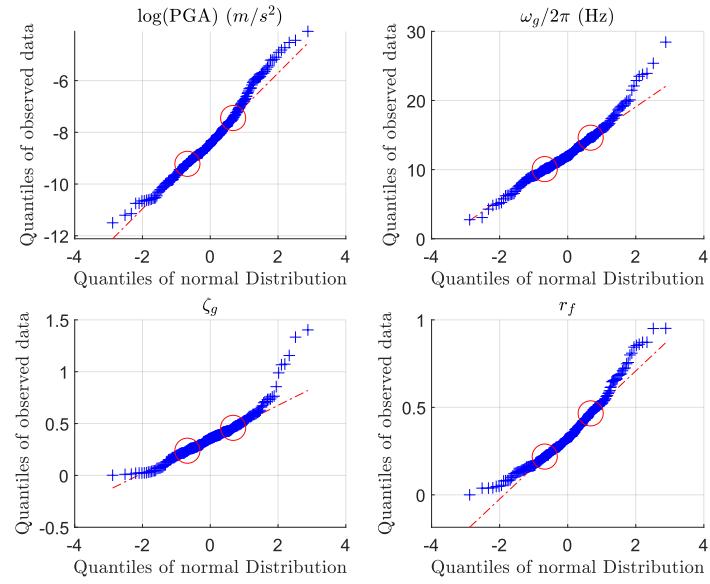


**Figure A.8:** Histograms of model parameters  $\omega_g$ ,  $\zeta_g$ ,  $r_h$ ,  $\text{PGA}$  in x-direction. The frequency in the histogram is normalized by the relative probability, where the  $\text{PGA}$  is presented on the logarithm scale.

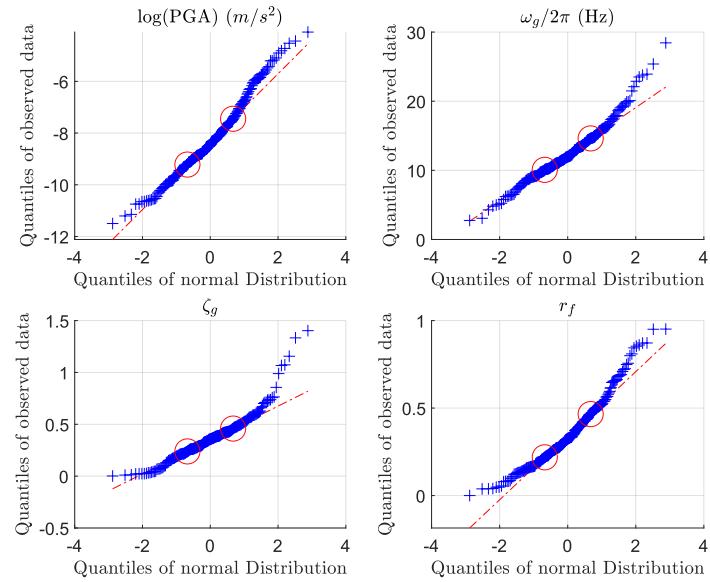


**Figure A.9:** Histograms of model parameters  $\omega_g$ ,  $\zeta_g$ ,  $r_h$ ,  $\text{PGA}$  in y-direction. The frequency in the histogram is normalized by the relative probability, where the  $\text{PGA}$  is presented on the logarithm scale.

### A.5.2 Quantile plots of model parameters



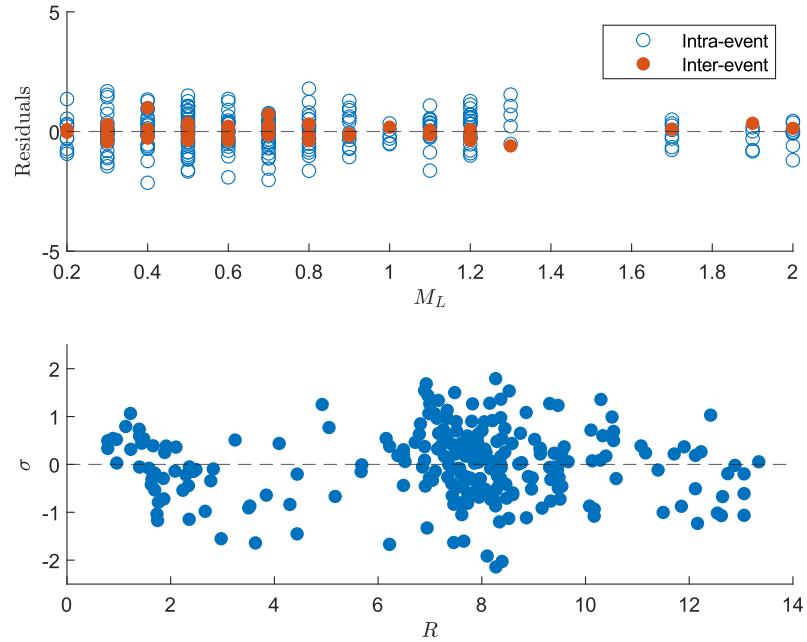
**Figure A.10:** Quantile plots of the model parameter in x-direction. Hollow circles indicate the first and the third quartiles.



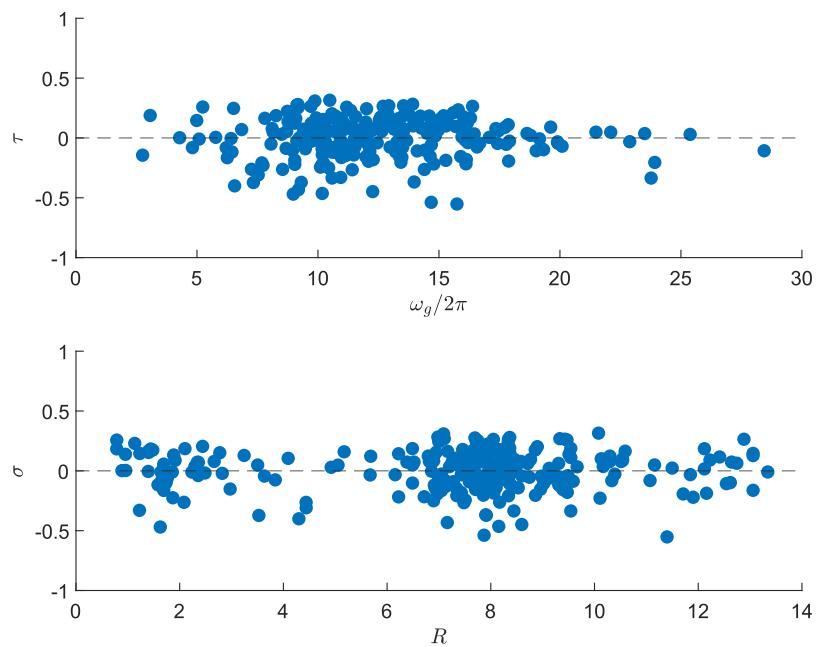
**Figure A.11:** Quantile plots of the model parameter in y-direction. Hollow circles indicate the first and the third quartiles.

### A.5.3 Fitting result of Linear mixed effect model

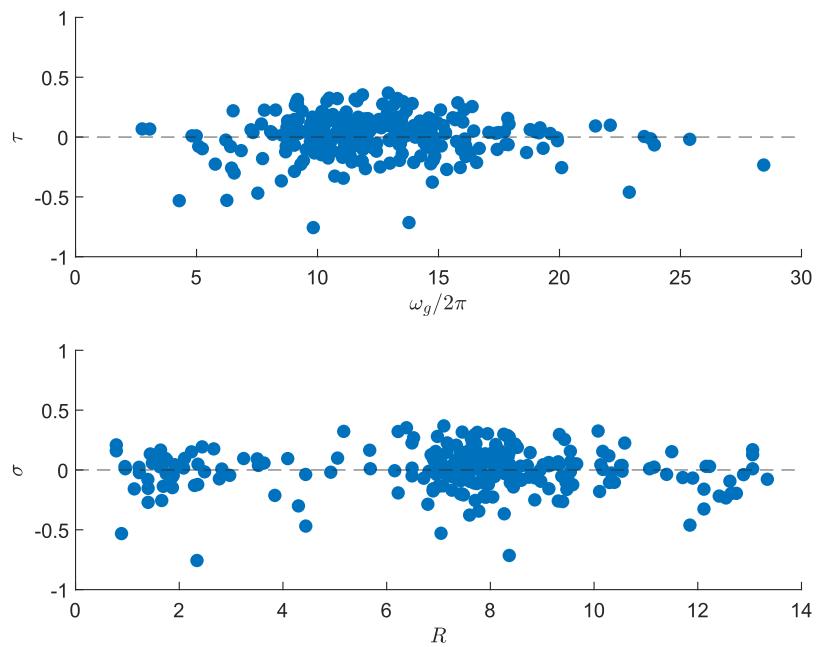
#### A.5.4 Scatter plots of residuals of fitting linear mixed effect model



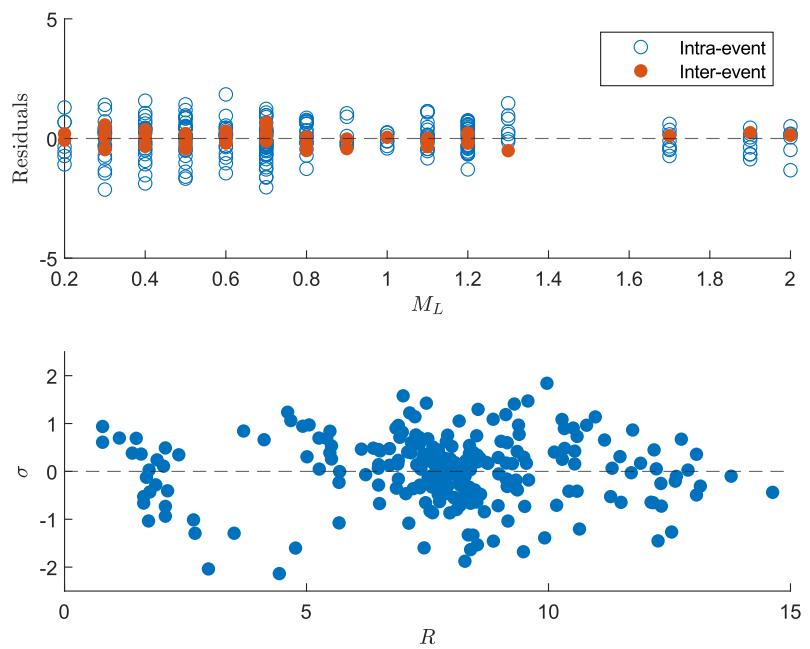
**Figure A.12:** Scatter plots of residuals of  $\ln(PGA)$  against  $M_L$  and  $R$  in x-direction



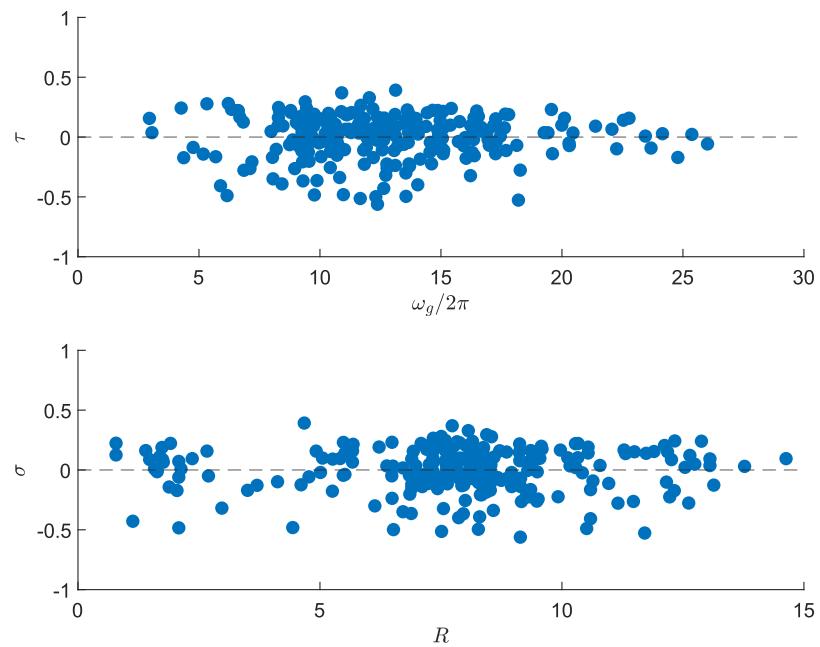
**Figure A.13:** Scatter plots of residuals of  $\omega_f$  against  $\omega_g$  and  $R$  in x-direction



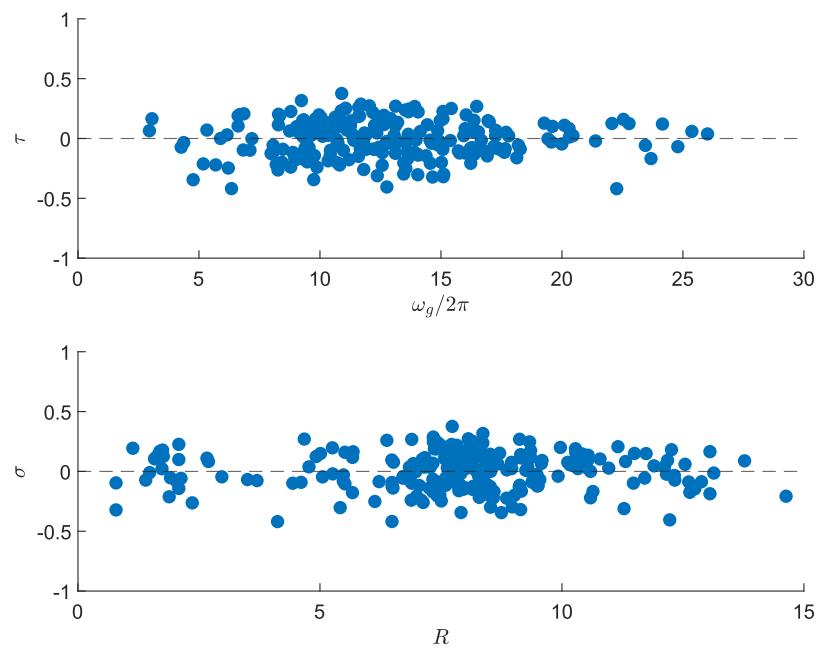
**Figure A.14:** Scatter plots of residuals of  $\zeta_g$  against  $\omega_g$  and  $R$  in x-direction



**Figure A.15:** Scatter plots of residuals of  $\ln(\text{PGA})$  against  $M_L$  and  $R$  in y-direction

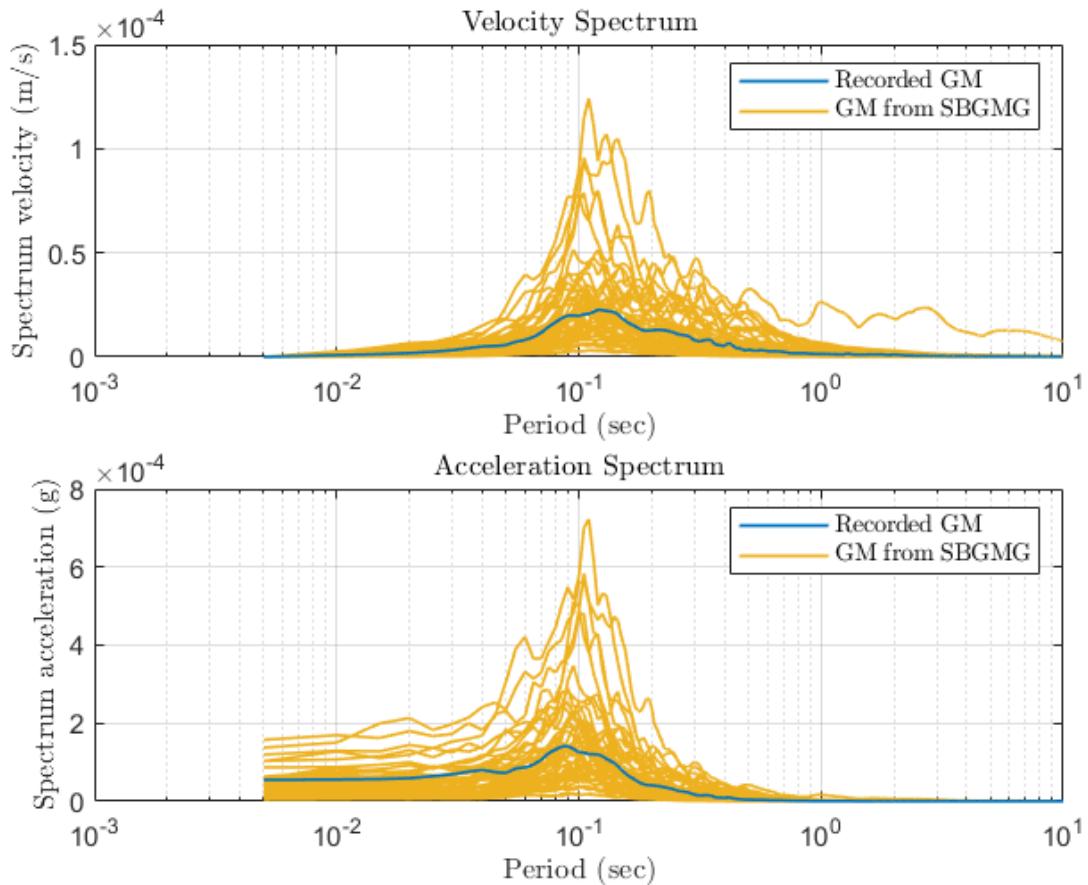


**Figure A.16:** Scatter plots of residuals of  $\omega_f$  against  $\omega_g$  and  $R$  in y-direction

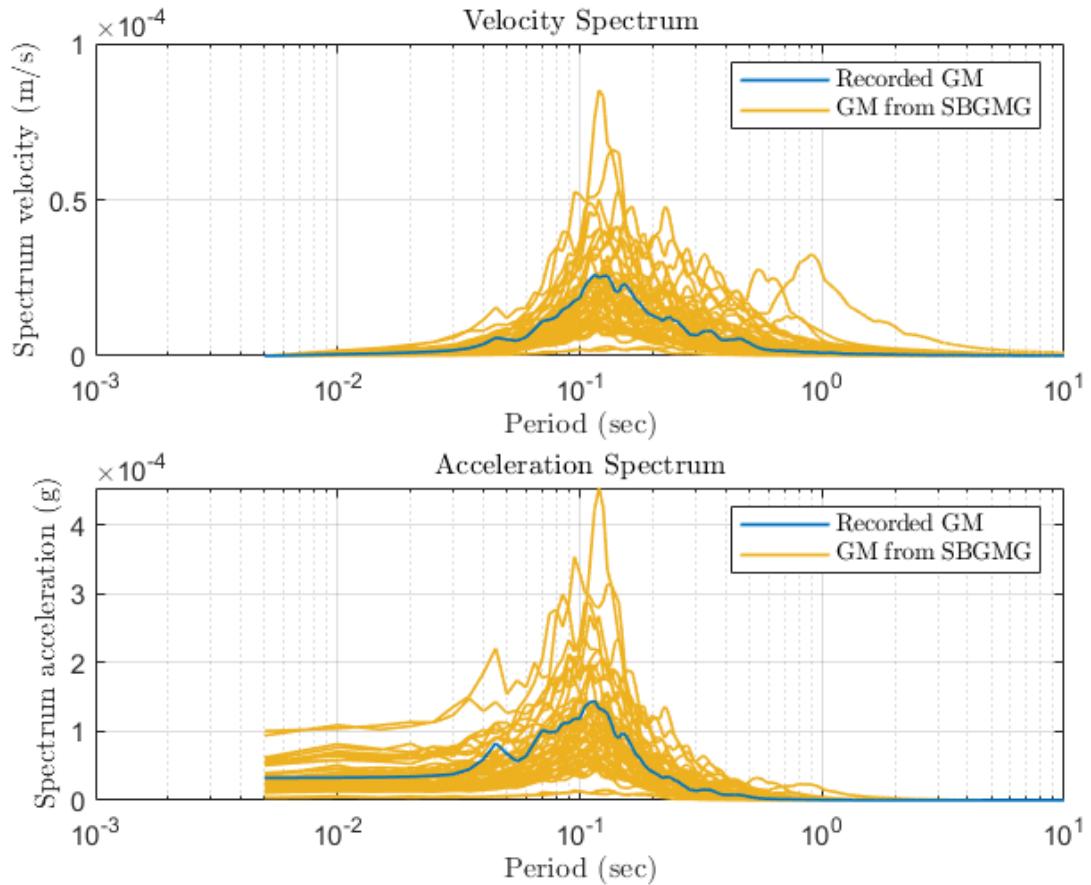


**Figure A.17:** Scatter plots of residuals of  $\zeta_g$  against  $\omega_g$  and  $R$  in y-direction

### A.5.5 Two further examples of response spectral analysis



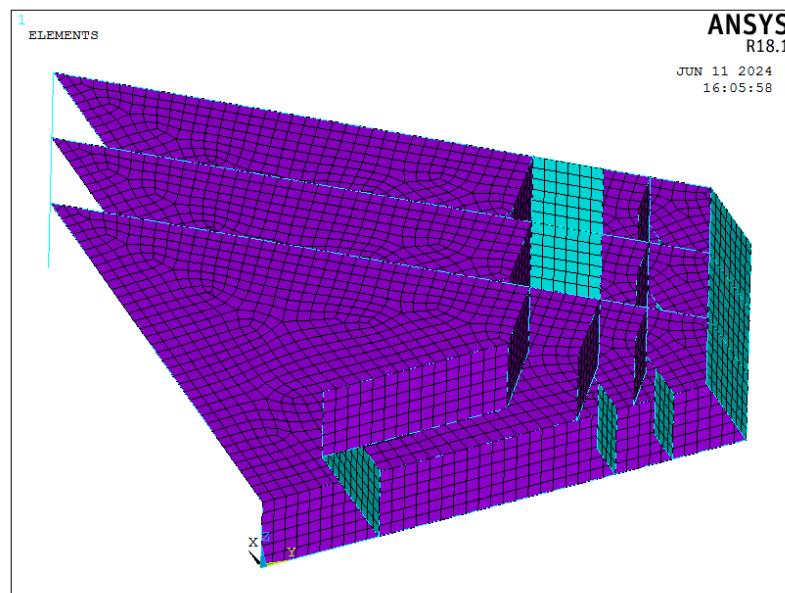
**Figure A.18:** Response spectral analysis. The top figure shows the velocity response, and the bottom shows the acceleration response. The blue line is the spectral response of the ground motion, recorded at 14:18:43 on March 27<sup>th</sup>, 2021, from the seismic station INS3, Insheim, Germany. The 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.



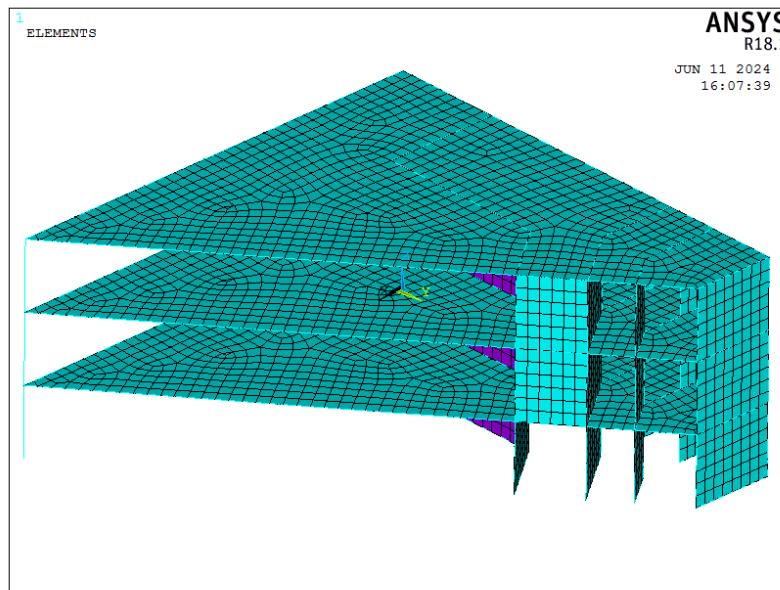
**Figure A.19:** Response spectral analysis. The top figure shows the velocity response, and the bottom shows the acceleration response. The blue line is the spectral response of the ground motion, recorded at 16:28:45 on February 26<sup>th</sup>, 2021, from the seismic station TMO20, Insheim, Germany. 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.

## A.6 Additional plots of case study

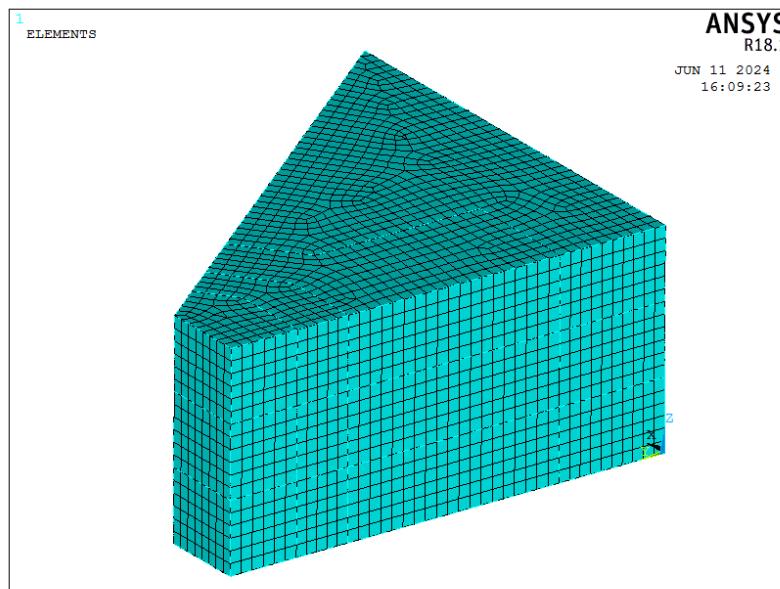
### A.6.1 Different perspectives of Building model



**Figure A.20:** Building Geometry at geothermal energy site in Taufkirchen. 2<sup>nd</sup> Perspective.

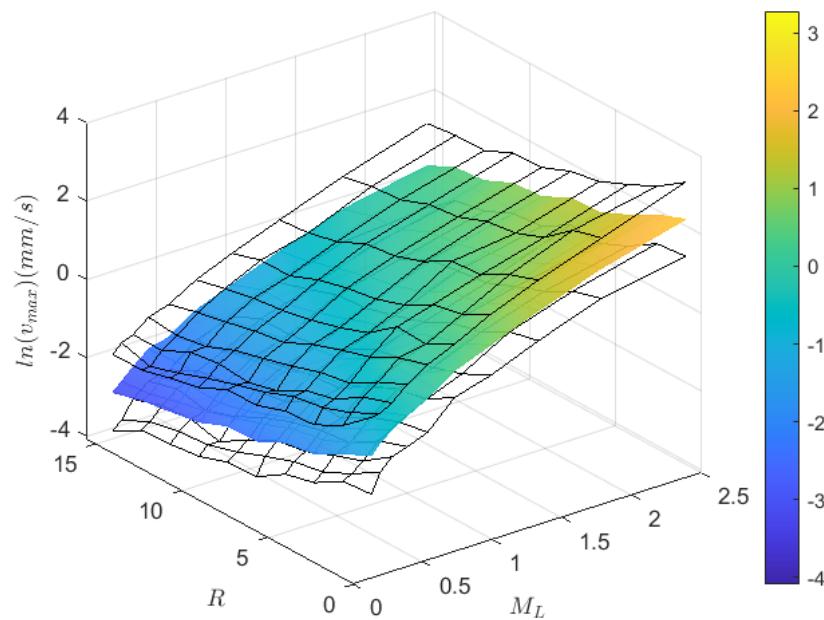


**Figure A.21:** Building Geometry at geothermal energy site in Taufkirchen. 3<sup>rd</sup> Perspective.

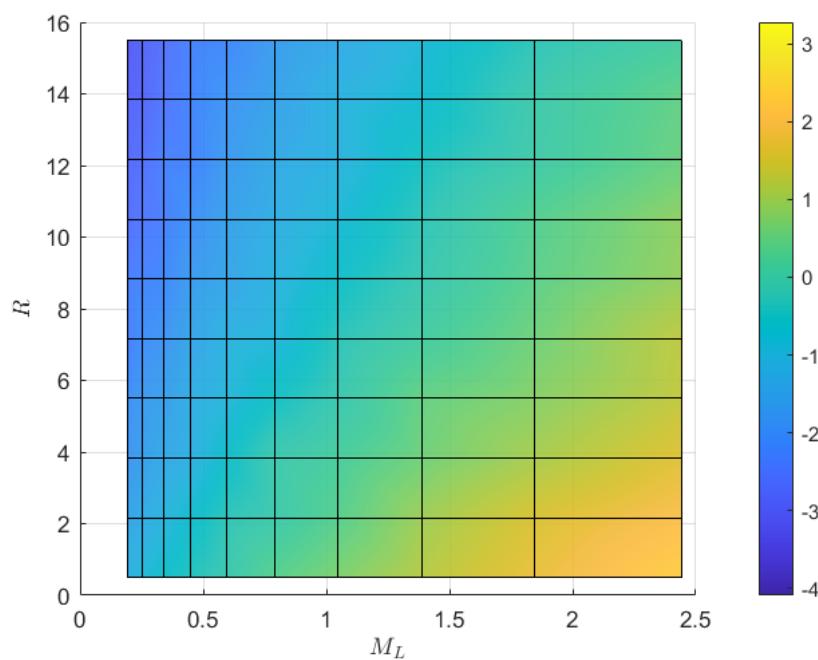


**Figure A.22:** Building Geometry at geothermal energy site in Taufkirchen. 4<sup>th</sup> Perspective.

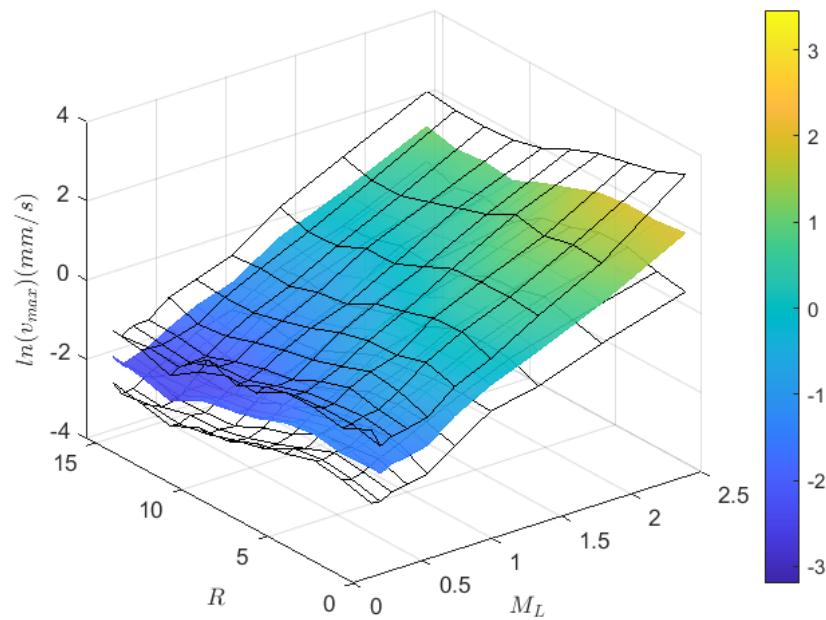
### A.6.2 Response surfaces of the trained SPCE model



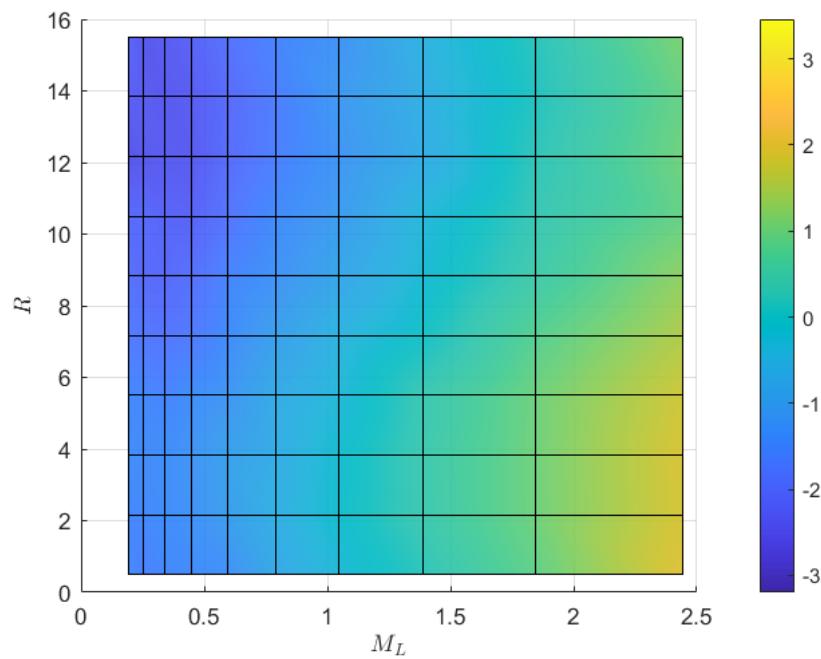
**Figure A.23:** 3D Response function built by SPCE against  $M_L$  and  $R$  in x-direction. The black-mesh surfaces are the standard deviation computed from 100 replications



**Figure A.24:** 2D Response function built by SPCE against  $M_L$  and  $R$  in x-direction



**Figure 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



**Figure A.26:** 2D Response function built by SPCE against  $M_L$  and  $R$  in y-direction