Data-driven reliability assessment of dynamic structures based on power spectrum classification

Marco Behrendt[a,b], Masaru Kitahara[a], Takeshi Kitahara[c], Liam Comerford[b], Michael Beer[a,b,d]

[a] *Institute for Risk and Reliability, Leibniz Universität Hannover, Germany*
[b] *Institute for Risk and Uncertainty, University of Liverpool, United Kingdom*
[c] *Department of Civil Engineering, Kanto Gakuin University, Japan*
[d] *International Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University, Shanghai, China*

Abstract

The power spectral density function is a widely used tool to determine the frequency components and amplitudes of environmental processes, such as earthquakes or wind loads. It is an important technique especially in the engineering field of vibration analysis and in determining the response of structures. When using a large amount of data, a load model can be generated, which describes the characteristics of the underlying stochastic process. This load model enables artificially generated excitations to be generated within the framework of Monte Carlo simulations. If multiple data records are utilised, a problem that can occur is that the individual records have a high variance in the frequency domain and are therefore too dissimilar from each other, even though they appear to be similar in the time domain. A load model derived from this data does not represent the entire data set, because not the whole spectral range is covered. Therefore, every attempt must be made to group the data according to their characteristics and thus combine similar data to derive two or more load models accordingly. In this work, an approach is proposed to classify real earthquake ground motion records using the k-means algorithm based on similarities within the data ensemble as determined by the Bhattacharyya distance. The silhouette method enables the identification of the optimal number of groups for the classification. The classified data thus forms a subset of the entire data set from which load models can be generated and can be applied separately to the structure under investigation, leading to more accurate simulation results. The advantages of this classification approach are illustrated by means of an academic example and a seismic-isolated bridge pier model as a non-linear dynamic system.

Keywords: Power spectral density function, Stochastic processes, Stochastic dynamics, Reliability assessment, Uncertainty quantification, Earthquake engineering

1. Introduction

The simulation and subsequent reliability assessment of buildings and structures under specific loads has become increasingly important in engineering in the recent decades [1, 2, 3, 4]. In particular, structures that are subject to environmental processes such as wind and earthquake loads and thus exhibit dynamic system behaviour are of special interest [5, 6, 7]. A general understanding of the dynamic behaviour of structures, especially under earthquake loads, is given in [8]. To describe the environmental processes, which can be characterised as stochastic processes, in terms of their frequency components and governing frequencies, the Power Spectral Density (PSD) function can be utilised [9, 10]. The PSD function describes the stochastic process in the frequency domain and thus provides information about the frequencies, which are particularly important in structural dynamics. Through the PSD function, suitable stochastic processes can be generated in the time domain [11], which may be used for numerical simulations within the framework of extensive Monte Carlo (MC) simulations in order to obtain the response of the structure under investigation [12, 13, 14]. Modal analysis and frequency decomposition methods by singular value decomposition of the PSD function is an alternative approach to MC simulations for characterising system responses, see for example [15, 16, 17, 18, 19].

Especially in simulations involving dynamic system behaviour due to environmental processes, accurate simulation results are important to evaluate existing structures in terms of their resistance and durability or for the design of new buildings. An overview about risk assessment of earthquakes is provided in [20]. Simulations are necessary to provide an understanding of the real case and to obtain initial assessments of the response behaviour of a structure. A direct application of the safety specifications for structures at risk in civil engineering, such as defined in [21], is often not possible due to structural complexity or incomplete information about the system. Such a model can be investigated with regard to different excitations. The simulation and evaluation of the dynamic response of structures under specific loads, and in particular under seismic loads, has become increasingly important, with a particular emphasis on variability and uncertainties: variability of the model and variability of the input seismic motion. Accounting for uncertainties in both the structure and the input ground motions is important for a rigorous assessment of the seismic capacity of the structure. A suitable method for this purpose is Incremental Dynamic Analysis (IDA) [22, 23, 24], which applies earthquake loads with different scaled intensities to a structure. This yields functions that enables a comparison of different system responses to a range of intensity levels of the excitation. This method can be used to determine system responses for different potential earthquake excitations and to design the structure accordingly. Performance based engineering demand approaches, specifically fragility functions are utilised for defining the probability that a component exceeds a certain limit state depending on the excitation, e.g. the peak ground acceleration (PGA). An overview of different methods for determining fragility functions can be found in [25, 26, 27, 28, 29]. In [30] a computationally efficient method for analysing the seismic fragility of structures is proposed, while in [31] fragility analyses are linked to artificial neural networks. Seismic fragility analysis is combined with Bayesian linear regression demand models in [32], yielding more accurate results compared to traditional methods. Other works deal with fragility analysis for specific structures, such as highway bridges [33], concrete dams [34] or railway bridges [35]. Both, the fragility analysis and IDA are concerned with the selection of seismic ground motions and with the efficient and sufficient intensity measures [36, 37] of ground motions [38].

The definition of appropriate seismic intensities plays a key role in the earthquake engineering and engineering seismology to reduce the variability of the analysis results. The variability would strongly increase if the input ground motions have no similarity, which in general is always the case. Thus, it is valuable to classify the real earthquake records, which is the objective of this work, and define appropriate seismic loading models with smaller uncertainty to obtain more reliable results. To support this and to improve the simulation results, real data records can be used instead of artificially generated data. An overview of the data analysis of real data can be found, for instance, in [39, 40, 41]. Thanks to the ever increasing databases of environmental processes (e.g. [42, 43]), a large amount of data is available from which corresponding load models can be generated. Although a pre-selection of data can be made based on seismological criteria such as magnitude, epicentral distance, depth of the earthquake or site conditions, these data are never identical due to the nature of earthquakes. Furthermore, soil conditions the path and the source mechanisms, such as normal, inverse or strike-flip faults, influence the ground motions, see [44] for an overview. In all cases, even when using similar ground motion criteria and a similar building model, a large variability of the building response might be observed, which is difficult even for data of the same region [45]. In addition, uncertainties due to, for example, measurement errors, incorrect calibration or damage to the sensor or total failure can complicate the selection and subsequent analysis. Despite the fact that the data can be pre-selected according to the criteria mentioned, they may still be too different to obtain reliable results, i.e. with reduced variability. In such a case, a fatal assessment of the situation can emerge. For those problems, the temporal similarity can be defined considering time or frequency parameters [46]. In some cases, the data ensemble has a high spectral variance in the frequency domain, so that a single PSD function estimate is not sufficient to adequately represent the process statistics. It can reasonably be assumed that a more realistic representation of the spectral range of the process is captured by estimating two or more PSD functions to better represent the spectral range of the process. Therefore, it is necessary to define the *spectral similarity* that can be used to categorise individual data sets.

A number of different methods for classifying earthquake ground motions can be found in the literature. Many of these methods rely on heuristic methods such as the k-means algorithm [47] that can be used for fast local solutions [48]. For example, in [49] a method is presented that takes the spectral shape into account. In [50], different frequency content based parameters are used to classify the earthquakes using k-means and self-organizing maps (SOM). The moment magnitude and the Joyner-Boore distance [51] are used in [52, 53] to classify earthquake ground motions with the k-means algorithm as well, while in [54] and [55] fuzzy-based approaches are employed. All these approaches require different parameters from the time and/or frequency domain for the classification of earthquake ground motions. This presupposes a prior knowledge of the data used. In addition, the choice of parameters can lead to different results of the classifications, which in turn influence the simulation results. To simplify the classification and provide more robustness, this paper proposes a method where only the similarity in the frequency domain needs to be determined, and subsequently the earthquake ground motions can be grouped using the k-means algorithm. Furthermore, suitable load models can be generated from the classified PSD functions of the earthquakes.

The proposed approach is to first define the number of spectral groups and then optimise the arrangement of the data sets between the groups by minimising their respective spectral distance. The classification is carried out in the frequency domain only, as the frequency components of a time signal can thus be determined unambiguously. In addition, the signals in the time domain show hardly any or only small differences, whereas the transformation into the frequency domain often reveals larger differences. Furthermore, these signals are used in the field of stochastic processes and dynamic systems, which is why it is useful to classify ground motions based on their frequency characteristics. For the classification, the Bhattacharyya distance [56, 57, 58] is used to determine the similarity of the individual PSD functions. With the k-means algorithm, the PSD functions are classified into two or more groups. It is expected that in most cases, considering multiple spectral models will result in a more accurate overall response statistics than a single model. However, the latter requires that several different simulations of the structural response are carried out, which can be time consuming for large model analyses. Therefore, an approach based on the silhouette method is proposed to determine the optimal number for classification, which results in avoiding to perform structural response simulations more than necessary. This type of data processing leads to a more accurate analysis of structures and buildings, especially in the area of reliability analysis and assessment, and can reveal system failure that would not be detected when utilising a single PSD function estimate of the data set. The proposed method also enables to estimate the system response considering the probability of occurrence of each load model. Therefore, this method is useful in particular when utilising multiple real data records for the reliability assessment of real structures. The novelty is in the combination of the different basic tools and their further development and adjustment to solve the given classification problem. The proposed method improves the quality of the reliability assessment for large structures utilising site- and source-specific information. In addition, the classification does not require any parameters, except for a maximum number of groups to be determined, and is automatic, including the determination of the number of optimal groups, whereas other approaches require a set of parameters and prior knowledge. The classification approach presented in this work is valuable from an engineering view point, especially in probabilistic seismic engineering as it contributes significantly to the selection of appropriate real data for other commonly used methods in earthquake engineering. The selection of suitable real data is essential for reliable simulation results and especially for reducing the variability of the results. The classification approach can thus be transferred to other methods in probabilistic earthquake engineering.

In this work, real earthquake ground motion records are used, which are provided by the National Research Institute for Earth Science and Disaster Resilience in the K-NET and KiK-net databases [42]. This demonstrates that the proposed method is also feasible for practical application. The developed load models from the real data records are applied to a linear spring-mass-damper system with one degree of freedom and a seismic-isolated bridge pier model as an example of non-linear dynamic systems.

This work is organised as follows: Section 2 summarises briefly the theoretical background used in this work. In Section 3 the approach for classifying PSD functions is explained. Section 4 provides the classification approach for two examples of real earthquake ground motions. In Section 5 the classified ensembles are applied to two numerical examples to show the strength of the approach. The work concludes with Section 6.

2. Stochastic processes and power spectrum estimation

A stochastic (or random) process is influenced by random phenomena and fluctuations, so that it cannot be described completely deterministically. The value of the stochastic process at any point in time is determined by random variables [59]. The frequency composition of a zero-mean stationary stochastic process $X(t)$ can be derived via the Fourier transform of its autocorrelation function $R_X(\tau) = E[X(t)X(t+\tau)]$

$$S_X(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R_X(\tau)e^{-i\omega\tau}d\tau \tag{1}$$

and the inverse Fourier transform

$$R_X(\tau) = \int_{-\infty}^{\infty} S_X(\omega)e^{i\omega\tau}d\omega, \tag{2}$$

where $S_X(\omega)$ describes the PSD function. Eq. 1 and Eq. 2 are called Wiener-Khintchine theorem (e.g. [6, 59, 60]).

For generating simulated stochastic processes, the Spectral Representation Method (SRM) can be utilised [11]

$$X(t) = \sum_{n=0}^{N-1} \sqrt{4S_X(\omega_n)\,\Delta\omega}\cos(\omega_n t + \varphi_n), \tag{3}$$

where

$$\omega_n = n\Delta\omega, \quad n = 0, 1, 2, \ldots, N-1 \tag{4}$$

with $N \to \infty$ and φ_n as uniformly distributed random phase angles in the range $[0, 2\pi]$ and t as time vector. This provides a suitable method for generating compatible time signals derived from and carrying the characteristics of an underlying PSD function.

The estimation of the PSD function of a stationary stochastic process can be obtained by the periodogram [60, 6], which is formed by the squared absolute value of the discrete Fourier transform of the signal $X(t)$. The periodogram reads as follows

$$\hat{S}_X(\omega_k) = \lim_{T\to\infty} \frac{\Delta t^2}{T} \left| \sum_{t=0}^{T-1} X(t) e^{-\frac{i2\pi}{T}kt} \right|^2 , \tag{5}$$

where Δt is the time step size, T is the total length of the record, t describes the data point index in the record and k is the integer frequency for $\omega_k = \frac{2\pi k}{T}$. As the squared amplitude of the DFT is directly involved in equation 5, it is of particular interest in this work to propagate the uncertainty in the time signals.

3. Classification of spectral groups within ensembles

In this section, a brief overview of the problem is given and the method for classifying an ensemble of PSD functions is explained using an academic example. In addition, a method is presented which determines the optimal number of groups.

3.1. Problem statement

In most cases, no differences can be detected in the time domain, the time signals seem to be almost identical, see left side of Fig. 1. The time signals given here are derived from two different source PSD functions but the exact same random variables for generating the time signal (Eq. 3) are utilised. The PSD functions estimated from the time signals are given on the right side of Fig. 1. Despite the small differences in the time domain, significant differences are evident in the frequency domain, where it is clear that the PSD functions differ in spectral density and peak frequency. Although this is an academic example it illustrates that only infinitesimal differences in the time domain can cause significant differences in the frequency domain. Such a problem can occur when working with a large amount of data. Therefore, a thorough investigation of the data must be carried out. In a case like this, it may be useful to define two or more load models.

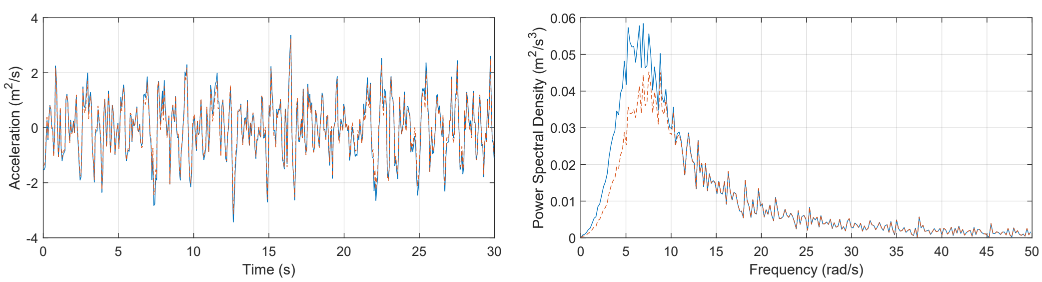

Figure 1: Example of two signals that show hardly any differences, but reveal the differences in the frequency domain.

3.2. Methodology

To identify different groups of PSD functions, the spectral similarity between two PSD functions P_1 and P_2 must be determined. In this work, it is proposed to use the Bhattacharyya distance

$$D_B\left(P_1, P_2\right) = -\log\left(\sum_{\omega\in\Omega}\sqrt{P_1\left(\omega\right)P_2\left(\omega\right)}\right). \tag{6}$$

Due to it's definition, the Bhattacharyya distance is a suitable distance measure for determining the similarity of the individual PSD functions within the ensemble. Indirectly it accounts, for instance, for the total power and shape (i.e. the distribution of frequency power) of the PSD functions. Therefore, two PSD functions with e.g., the same total power but different shape will have a larger distance than two PSD functions with the same total power and similar shape.

To determine the similarity of the PSD functions in the ensemble, the ensemble mean is first used as a reference spectrum. Using the Bhattacharyya distance Eq. (6), the similarity of each individual spectrum to the ensemble mean is determined. Therefore, Eq. (6) is evaluated for each individual spectrum and the ensemble mean. The resulting distance values are similar for PSD functions with similar shape and power. These distance values are used to determine similarity clusters using the k-means algorithm and to divide the entire data set into clusters, or so-called groups, i.e., similar distance values to the ensemble means result in the assignment to the same group. To perform this procedure, the number of desired groups must be defined beforehand. In general, a higher number yields in load models covering wider spectral ranges, while it requires larger computational burden. Therefore, it is important to determine the optimal number of groups. A method for this purpose is proposed in Section 3.3.

3.3. Optimal number of spectral groups

There are several methods available for determining the optimal number of groups k_{opt}, such as elbow method, for instance. However, these methods require often a visual assessment of the analyst to determine k_{opt} with respect to certain statistics. Another problem is that k_{opt} is subjective as it depends on the given data and the methods used to measure the distances. In this work, it is suggested to use the silhouette method [61, 62] for the determination of k_{opt}. Other approaches can be found in [63] and the references therein.

A maximum number of groups k_{max} has to be defined beforehand and the previously described procedure of classifying the PSD functions will be performed $k_{max} - 1$ times, i.e. for $2, 3, \ldots, k_{max}$ groups. The maximum number of groups k_{max} is case-dependent. For instance, an ensemble of power spectra with a high spectral variance might need more groups than with lower spectral variance. In any case, it is practical to choose a low number of k_{max}, for instance $5 \leq k_{max} \leq 10$. A very high k_{max} would not be reasonable because then it would also be possible to apply all given data sets individually to the structural model. This would no longer correspond to the intended classification. In order to obtain the most accurate classification possible, as many groups as necessary should be obtained, but as few as possible.

The silhouette coefficient provides a measure of the quality of a clustering that is independent of the number of clusters. The silhouette coefficient is defined as the arithmetic mean of all silhouette values $s(i)$

$$s_C = \frac{1}{n_i} \sum_{i=1}^{n_i} s(i) \tag{7}$$

where n_i describes the total number of data points and the silhouette values $s(i)$ are defined as

$$s(i) = \frac{b(i) - a(i)}{\max\left(a(i), b(i)\right)}. \tag{8}$$

According to [61], $a(i)$ is the average distance of the sample i to the other samples within the same cluster A and $b(i)$ is the average distance of sample i to the other samples in another cluster C which is closest to cluster A. The silhouette value can range from $-1 \leq s(i) \leq 1$, while a high silhouette value implies a high similarity to sample i's cluster. In the proposed method, the number of groups corresponding to the highest silhouette coefficients is chosen as the optimal number of groups k_{opt}.

In order to illustrate the silhouette method, a short academic example is given in Fig. 2. In this example 3 different underlying analytical expressions of power spectra are utilised to generate 10 PSD functions each, with different values, to simulate a certain randomness. From the given example it can be clearly seen that the optimal number of groups is $k_{opt} = 3$. The example therefore only aims to illustrate the proposed method. In this case, it can be seen that the mean value of the entire ensemble (dashed line) is unsuitable for deriving a load model for the ensemble. Especially at frequencies around 2.5 rad/s, the three classified groups are completely disjointed, which clearly shows that a classification is useful. This can be confirmed by determining the optimal number of groups k_{opt}. The silhouette values for this example are calculated by Eq. (8) and are depicted in Fig. 3 for the classification into 2, 3, 4 and 5 groups. It can be appreciated that, especially for the classification in group 3, very high individual silhouette values are achieved as all of them are close to 1. When classified in 4 or 5 groups, on the other hand, silhouette values with lower quality are more frequent, showing that the classification is not well-suited for certain PSD functions. To determine the optimal number of groups k_{opt}, Eq. (7) is used to compute the silhouette coefficient for each individual classification, which is the mean value of all silhouette values for the according classification. This results in the silhouette coefficients shown in Fig. 4. The maximum of all silhouette coefficients reveals the optimal number of groups, accordingly $k_{opt} = 3$ in this example.

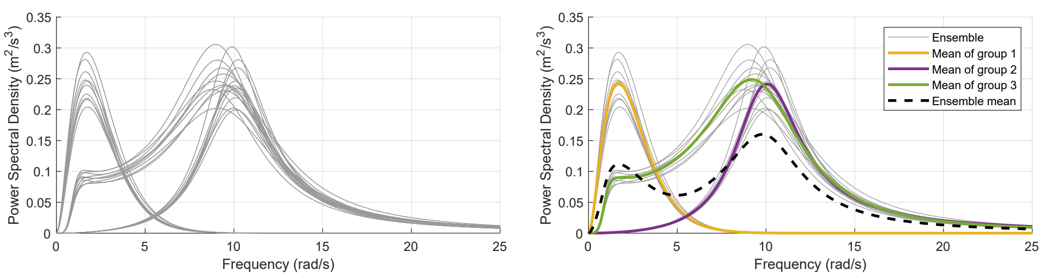

Figure 2: Unclassified ensemble (left) and the corresponding mean values of the classified groups and the entire ensemble (right).

3.4. Usage of the method

When using real data, usually given in the time domain, there are two possibilities, for both of which it can be argued why they are useful. After transforming the data from time domain to frequency domain and carrying out the classification, the options are:

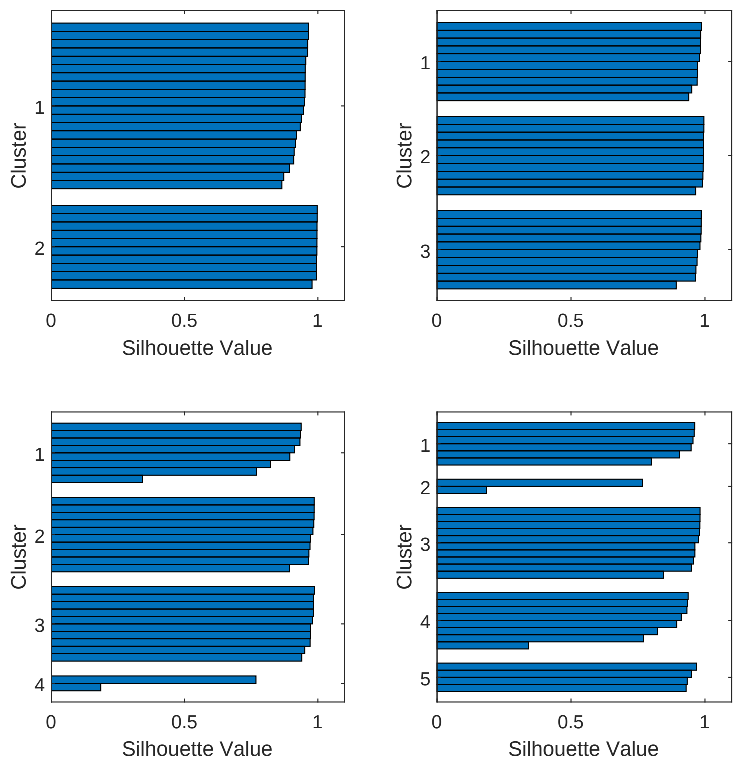

Figure 3: Silhouette values for the classification into 2, 3, 4 and 5 groups.

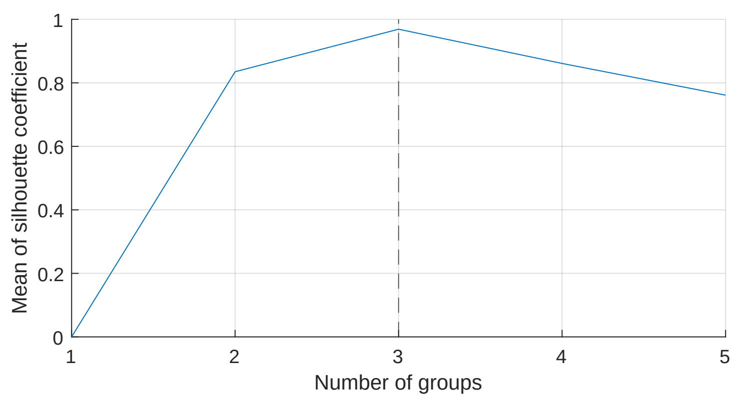

Figure 4: Silhouette coefficients for different numbers of groups. The optimal number of groups is $k_{opt} = 3$ as the maximum value is obtained there.

(i) Set up a load model based on the mean spectra of the classified groups. Then utilise Eq. 3, for instance, to generate time signals, which can be applied via MC simulation to structures.

(ii) The data in time domain can be applied directly for the respective groups in order to carry out a reliability analysis.

Whether to use option (i) or (ii) is dependent on the amount of given data. If the size is small, option (i) seems to be appropriate in order to set up a load model from which data with similar characteristics within the classified groups can be generated. If real data is available in a large amount option (ii) might be the better choice as it can be applied directly to the system. However, in the numerical examples in this work in the following sections, the focus is on case (i).

4. Classification of real data records

The real data records utilised in this work are provided by the K-NET and KiK-net database [42] and were chosen and downloaded by the authors. Thus, no pre-existing data selected by other authors were used. In general, there are mainly two ways to characterise ground motions, namely source-specific and site-specific characterisation. For source-specific characterisation only records of the same earthquake event but from different monitoring sites are utilised. Site-specific characterisation means that records of the same monitoring site but from different earthquake events are used. In the following, both ways of characterising ground motion are illustrated.

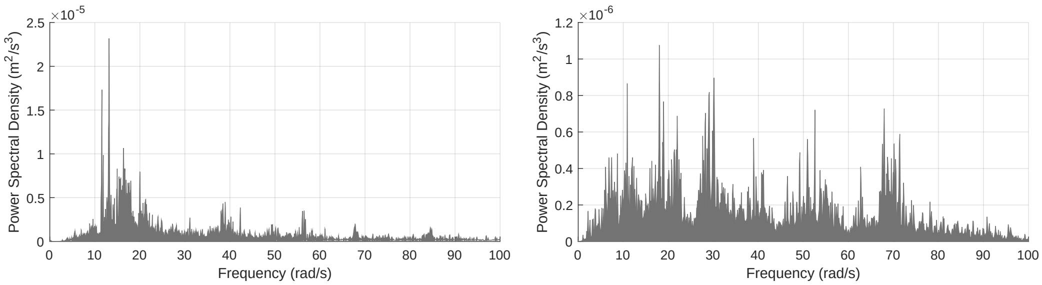

Figure 5: Source-specific (left) and site-specific ensemble (right).

4.1. Source-specific data ensemble

For the classification of data and application to structures, real data of a specific earthquake event was utilised, see Fig. 5 (left). The earthquake occurred at 20:50 on 17/07/2021 at a depth of 80 km at latitude 33.6N and longitude 131.9E with a magnitude of 5.1. Data was collected from 313 monitoring sites. All given ground motions have a total length $T = 120$ s and time step size $\Delta t = 0.01$ s. For a reliable classification, however, the data was preselected according to their PGA, as it is meaningless to compare and classify data with completely different amplitudes. Therefore, only ground motions in the range $0.02\ \mathrm{m/s^2} \leq PGA \leq 0.06\ \mathrm{m/s^2}$ were utilised in the following. The resulting data ensemble consists of 168 earthquake ground motions. The data were transformed into the frequency domain according to Eq. (5) and then classified using the Bhattacharyya distance (Eq. 6) and k-means algorithm. This was done for $k = 2$ and $k = 3$ groups. Fig. 6 shows the mean PSD functions of the resulting groups. For the classification into $k = 2$ groups, group 1 and group 2 both consists of 84 PSD functions. The classification in $k = 3$ groups yields 58 PSD functions in group 1, 63 in group

2 and 47 in group 3.

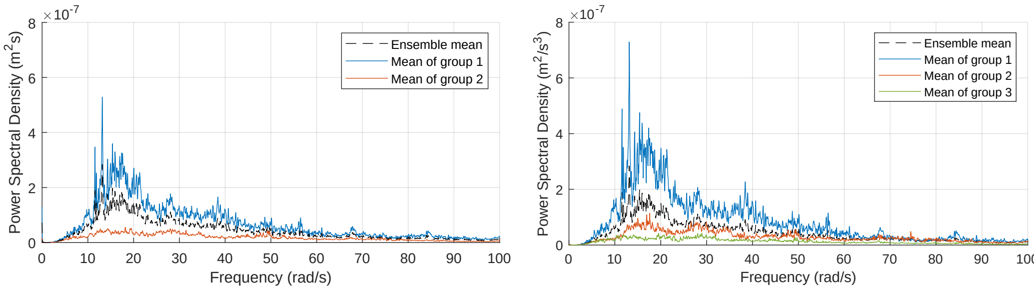

Figure 6: Classified mean PSD functions estimated from source-specific seismic ground motions with total length $T = 120$ s.

According to the silhouette method described in Section 3.3, the optimal number of groups is $k_{opt} = 2$, as shown in Fig. 7 and 8. For illustration purposes, however, the classification and simulation is carried out for both $k = 2$ and for $k = 3$ groups.

In order to verify the classification not only by sight, which can solely be an indicator, this is also substantiated by the total power and the peak frequency values of the classified group in frequency domain. For each group, the respective maximum and minimum are determined and given in Table 1. In the time domain, minimum, maximum and mean value of the PGA of the classified groups were determined and are shown in Table 2. Overlapping intervals of the minimum and maximum values with regard to the different groups are permissible here, since a combination of these factors influence the classification. However, a clear trend in the values can be recognised.

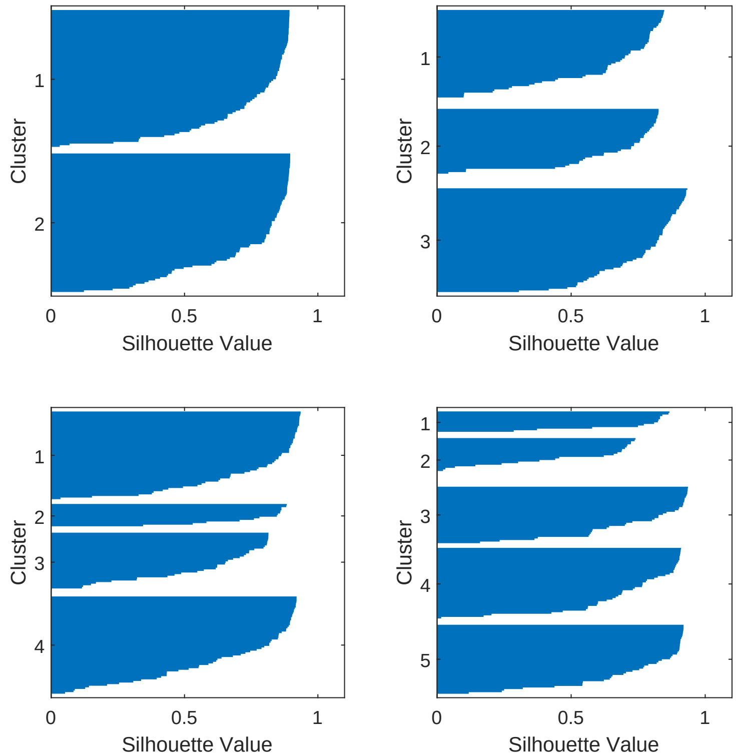

Figure 7: Silhouette values for the ensemble of source-specific data for different numbers of groups.

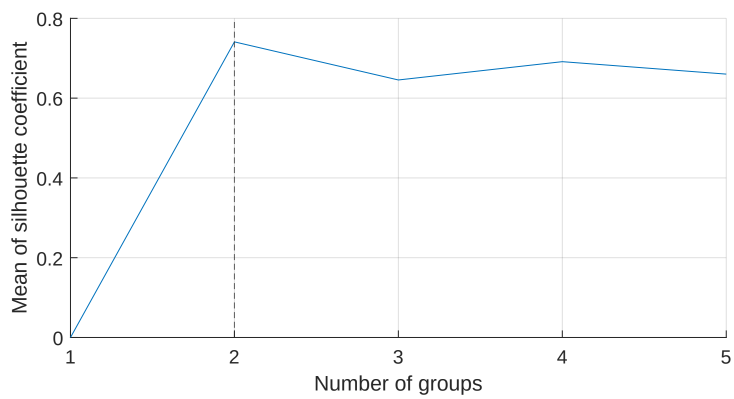

Figure 8: Silhouette coefficients for the source-specific ensemble for different numbers of groups. The optimal number of groups is $k_{opt} = 2$.

Table 1: Classification values for source-specific data in frequency domain

Classification	Group	Total power	Peak frequency value
$k = 2$	1	$[5.40 \cdot 10^{-5}, 5.29 \cdot 10^{-4}]$	$[1.54 \cdot 10^{-7}, 2.32 \cdot 10^{-5}]$
	2	$[7.9 \cdot 10^{-6}, 7.39 \cdot 10^{-5}]$	$[3.39 \cdot 10^{-8}, 1.94 \cdot 10^{-6}]$
$k = 3$	1	$[9.00 \cdot 10^{-5}, 5.29 \cdot 10^{-4}]$	$[4.27 \cdot 10^{-7}, 2.32 \cdot 10^{-5}]$
	2	$[3.83 \cdot 10^{-5}, 1.24 \cdot 10^{-4}]$	$[1.54 \cdot 10^{-7}, 5.05 \cdot 10^{-6}]$
	3	$[7.9 \cdot 10^{-6}, 5.04 \cdot 10^{-5}]$	$[3.39 \cdot 10^{-8}, 8.78 \cdot 10^{-7}]$

Table 2: Classification values for source-specific data in time domain

Classification	Group	PGA	mean(PGA)
$k = 2$	1	$[0.0218, 0.0595]$	0.0432
	2	$[0.0202, 0.0488]$	0.0282
$k = 3$	1	$[0.0271, 0.0595]$	0.0446
	2	$[0.0218, 0.0595]$	0.0347
	3	$[0.0202, 0.0475]$	0.0262

4.2. Site-specific data ensemble

For the site-specific classification of earthquake ground motions, data from the K-NET monitoring station in Tokyo, Japan (site code TKY007, site name Shinjuku) at latitude 35.7107N, longitude 139.6859E, and elevation 34 m were used, see Fig. 5 (right). Data from July 2010 to July 2021 were used for the classification. The utilised earthquake ground motions have a total length of $T = 60$ s and a time step size $\Delta t = 0.01$ s. As for the source-specific data before, the data was preselected according to the PGA. In this example, ground motions in the range $0.005\ \text{m/s}^2 \leq PGA \leq 0.015\ \text{m/s}^2$ are utilised. The resulting data ensemble consists of a total of 64 individual records. After transforming the data into the frequency domain according to Eq. (5) and classification using the Bhattacharyya distance (Eq. 6) and the k-means algorithm for $k = 2$ and $k = 3$ groups, the corresponding mean

PSD functions are obtained in Fig. 9. For the classification into $k = 2$ groups, group 1 consists of 35 PSD functions and group 2 consists of 29 PSD functions. The classification in $k = 3$ groups yields 25 PSD functions in group 1, 22 in group 2 and 17 in group 3.

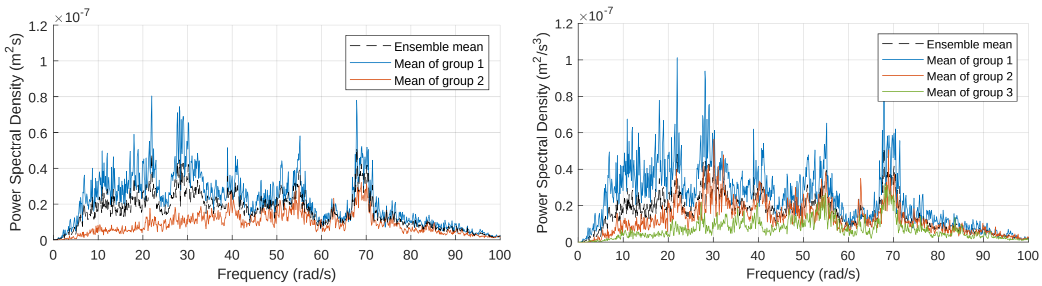

Figure 9: Classified mean PSD functions estimated from site-specific seismic ground motions with total length $T = 60$ s.

Determining the optimal number of groups using the silhouette methods yields $k_{opt} = 3$, as shown in Fig. 10 and 11.

The classification is verified by the total power and the peak frequency values. For each group, the respective maximum and minimum are determined and given in Table 3. The minimum, maximum and mean values of the PGA in the time domain of the groups were also determined and are shown in Table 4. As with the source-specific classification, a clear trend can also be seen in these values.

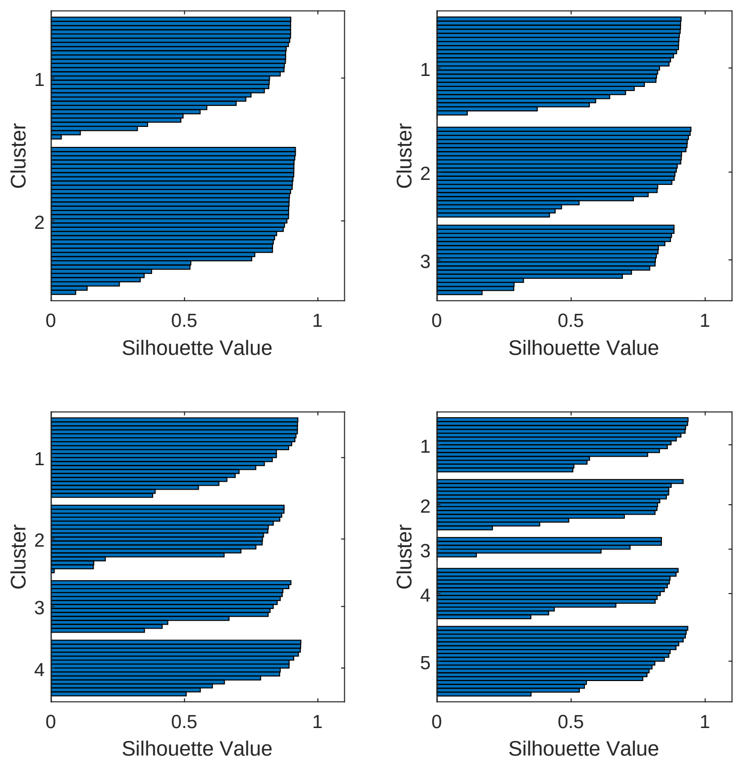

Figure 10: Silhouette coefficients for the ensemble of site-specific data for different numbers of groups.

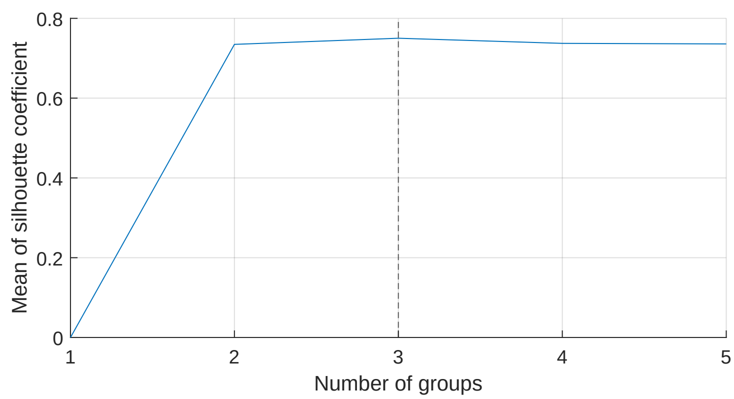

Figure 11: Silhouette coefficients for the site-specific ensemble for different numbers of groups. The optimal number of groups is $k_{opt} = 2$.

Table 3: Classification values for site-specific data in frequency domain

Classification	Group	Total power	Peak frequency value
$k=2$	1	$[1.24\cdot10^{-5}, 3.56\cdot10^{-5}]$	$[1.34\cdot10^{-7}, 1.08\cdot10^{-6}]$
	2	$[4.35\cdot10^{-6}, 1.39\cdot10^{-5}]$	$[3.77\cdot10^{-8}, 4.08\cdot10^{-7}]$
$k=3$	1	$[1.73\cdot10^{-5}, 3.56\cdot10^{-5}]$	$[2.57\cdot10^{-7}, 1.08\cdot10^{-6}]$
	2	$[8.85\cdot10^{-6}, 1.73\cdot10^{-5}]$	$[1.06\cdot10^{-7}, 8.96\cdot10^{-7}]$
	3	$[4.35\cdot10^{-6}, 9.33\cdot10^{-6}]$	$[3.77\cdot10^{-8}, 2.46\cdot10^{-7}]$

Table 4: Classification values for site-specific data in time domain

Classification	Group	PGA	mean(PGA)
$k=2$	1	$[0.0088, 0.0148]$	0.0124
	2	$[0.0051, 0.0142]$	0.0095
$k=3$	1	$[0.0092, 0.0148]$	0.0129
	2	$[0.0061, 0.0142]$	0.0107
	3	$[0.0051, 0.0127]$	0.0089

5. Numerical examples

In this section, generated load models from the classified ensembles of real earthquake ground motions are applied to two numerical examples in order to show the strength of the novel approach. The first example aims to demonstrate the effectiveness of the proposed classification approach and verify the identified optimal classification number, using a linear mass-spring-damper system considering different scenarios in the relationship between the system's natural frequency and dominant frequencies of input ground motions. The second example, on the other hand, aims to show the feasibility of the proposed method for reliability assessment of non-linear dynamic systems using a seismic-isolated bridge pier model.

The ensembles classified into 2 and 3 groups (Fig. 6 and Fig. 9) are used in the following for the numerical examples. For these, SRM (Eq. 3) is utilised to generate adequate time signals as the system's excitation. The derived mean PSD functions of the individual groups are used as the underlying PSD function required for SRM. For each classified group 10,000 MC samples were generated and applied to the structure. For each sample, the maximum displacement of the system in the time domain is determined, from which a cumulative distribution function (CDF) is calculated that can be used to estimate the probability of failure for specific displacements.

For the mass-spring-damper system discussed in Section 5.1, stationary stochastic processes are generated, whereas for the bridge pier model in 5.2, non-stationary stochastic processes emulated by an envelope function are generated, since the response property of non-linear dynamic systems are strongly affected by the non-stationarity of input ground motions. The envelope function is given by

$$g(t) = k\left(e^{-at} - e^{-bt}\right), \tag{9}$$

with $k = 500$, $a = 0.05$ and $b = 0.8$. This is to emulate a strong earthquake ground motion. Two examples of a generated stationary and a non-stationary process are given in Fig. 12.

5.1. Linear mass-spring-damper system

The first numerical example is performed using a Single-Degree-of-Freedom (SDOF) mass-spring-damper system. The system can be described by the

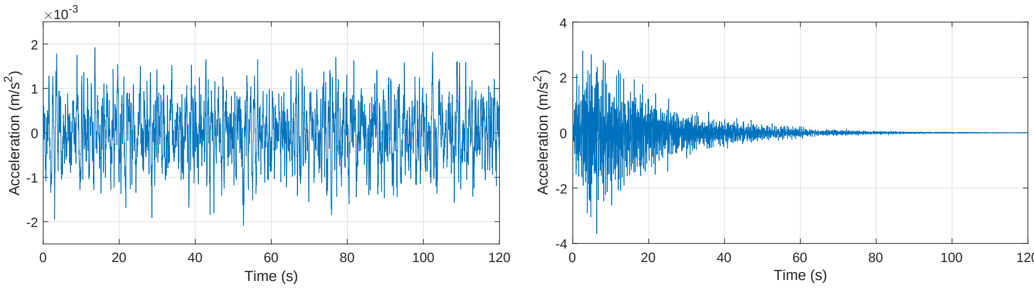

Figure 12: Stationary (left) and non-stationary ground motion acceleration (right), generated by SRM (Eq. 3) and the envelope function (Eq. 9).

following equation of motion

$$m\ddot{x} + c\dot{x} + kx = F(t), \tag{10}$$

with m as mass, k as stiffness and c as damping coefficient. The natural frequency is $\omega_0 = \sqrt{k/m}$ and the damping ratio is $\xi = c/(2\omega_0 m)$. x, $\dot{x}$ and $\ddot{x}$ denote displacement, velocity and acceleration of the system, respectively. The excitation $F(t)$ on the right-hand side is modelled by a stochastic processes based on the classified PSD functions derived in Section 4. An explicit Runge-Kutta scheme [64] is used to solve Eq. (10).

To show not only the influence of the input ensemble, but also of the system and its parameters, 2 different scenarios are calculated for each input ensemble, which will be called A and B for the source-specific ensemble and C and D for the site-specific ensemble in the following. The scenarios A and C represent the cases where the natural frequencies of the system and the dominant frequencies of the input ground motions differ, while scenarios B and D represent the cases where they are close to each other. The respective system parameters are given in Table 5.

Table 5: Parameters of the SDOF system for different scenarios.

Data set	Scenario	m (kg)	k (N/m)	c (Ns/m)	ω_0 (rad/s)	ξ (–)
source-specific	A	50	1922	15	6.2	0.024
	B	10	2800	15	16.733	0.045
site-specific	C	19	1922	15	10.058	0.039
	D	10	4835	15	21.989	0.034

5.1.1. Results of source-specific data

The resulting CDFs of the maximum system displacements for the classified source-specific data for scenario A are shown in Fig. 13. The CDFs for the classification into 2 groups are given on the left and for the classification into 3 groups on the right. For a better comparability, the results are shown for the mean value of the entire ensemble as well as for the mean values of the individual groups. In addition, a weighted mean CDF is given, taking into account the probability of occurrence of each load model (i.e., the ratio between the number of real ground motion records assigned to each group).

It can be seen, that the simulation results are more accurately for the classified load models compared to the load model of the entire ensemble. The distribution of the maximum system displacements varies considerably depending on the used load models defined by the groups. The results clearly show that a significantly higher range is covered by defining different load models. The individual load models themselves only cover a smaller range, but the load models considered as a whole reach a larger range. This shows that the definition of a single load model is not sufficient to cover all possible ranges of maximum displacement. Such a load model can lead to large system displacements not being identified in the simulation and a possible system failure remaining undetected. This is particularly evident when comparing the CDF of group 1 and group 2 for the classification into 2 groups. Where the CDF of group 2 reaches its maximum, the CDF of group 1 is almost identical to 0, which confirms that based on the two distinct simulations completely different values for the maximum displacements can be obtained. A similar result can be seen for the classification into 3 groups.

Furthermore, it can be easily recognised that in the example with the classification into 3 groups, group 2 and 3 hardly differ from each other. This is because the PSD functions of the group 2 and 3 are close to each other at the system's natural frequency. The weighted mean CDF, calculated from the individual CDFs of the groups taking into account their weights, reveals a slight shift compared to the CDF of the entire ensemble. This indicates that a more accurate system response was calculated by considering the weights of the individual groups because the weighted mean CDF can consider the probability of occurrence of each load model which will also affect the determination of the system reliability or decisions for planning buildings and structures in the future. The weighted mean CDFs are in a similar range to the CDF of the entire ensemble regardless of the number of classification, which supports that $k_{opt} = 2$ is reasonable and correct.

In Fig. 14 the results for the SDOF system for scenario B are depicted. Since the natural frequency of the system has changed due to the use of other

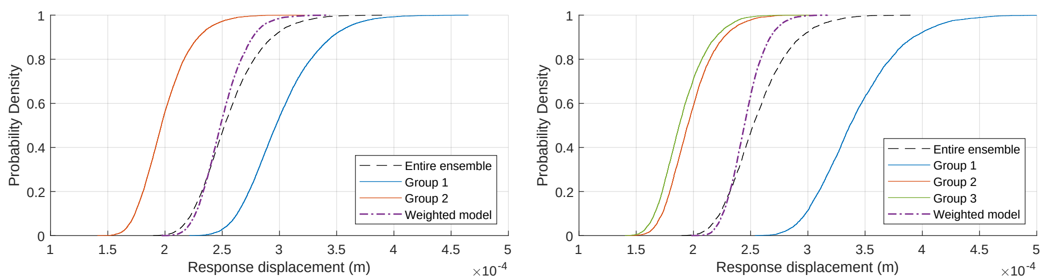

Figure 13: CDFs of the maximum response displacement of the linear mass-spring-damper system for scenario A for the classification into 2 groups (left) and into 3 groups (right) of the source-specific data ensemble.

system parameters, different simulation results arise. Compared to scenario A, where the spectral densities of the different groups were very close to each other at the natural frequency, now the natural frequency is around the area of the largest differences in the spectral densities of the ensemble. This can be seen in particular on the right-hand side of Fig. 14, as the CDFs of group 2 and 3 are significantly further apart than they were in scenario A. This also causes the weighted mean CDFs in both cases to shift to the left into the range of smaller system displacements. Accordingly, with the system parameters of scenario B, there is no overlap of the weighted mean CDF and the CDF of the entire ensemble. Nevertheless, the weighted mean CDFs are still in a similar range, which supports that the classification into 2 groups is sufficient.

Although the silhouette coefficients are often very close, see for example Fig. 8, the application of the classified models, however, shows that it is indeed effective. This is particularly evident in the results in Fig. 13. The classification results in the optimal number of groups $k_{opt} = 2$, which yields

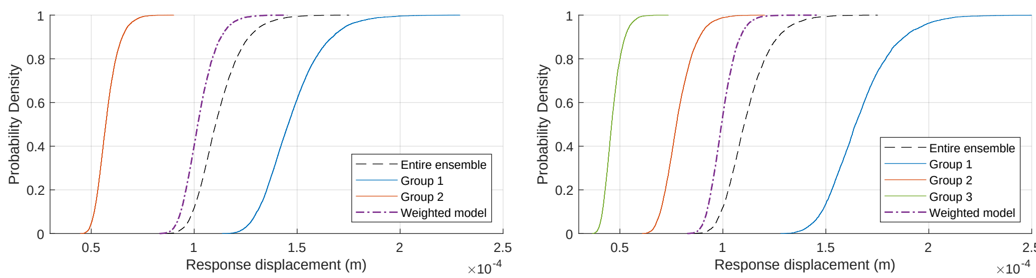

Figure 14: CDFs of the maximum response displacement of the linear mass-spring-damper system for scenario B for the classification into 2 groups (left) and into 3 groups (right) of the source-specific data ensemble.

reasonable results. When classified into 3 groups, the results of group 1 and group 2 are fairly close, indicating that they can form one group. This supports the argument that the classification into 2 groups is optimal.

5.1.2. Results of site-specific data

The results of the site-specific data and their classification show a similar behaviour as the results of the source-specific classification. In Fig. 15 the CDFs for the classification into 2 groups (left) and into 3 groups (right) for scenario C are shown. Again, the individual groups show a more accurate distribution of the maximum system displacements. Without a prior classification into groups, smaller and larger system displacements can hardly be recognised; this is only made possible by the classification. The overall model, which takes into account the weighted individual groups, also shows a more accurate representation of the maximum system displacements. Compared to the source-specific data, the optimal number of groups has been determined to be $k_{opt} = 3$. The weighted mean CDF for the classification into 3 groups is further shifted to the left side from the ensemble mean CDF compared to that for the classification into 2 groups. It supports that the classification into 2 groups is not enough for accurate estimation of system responses, and thus the classification into 3 groups is reasonable.

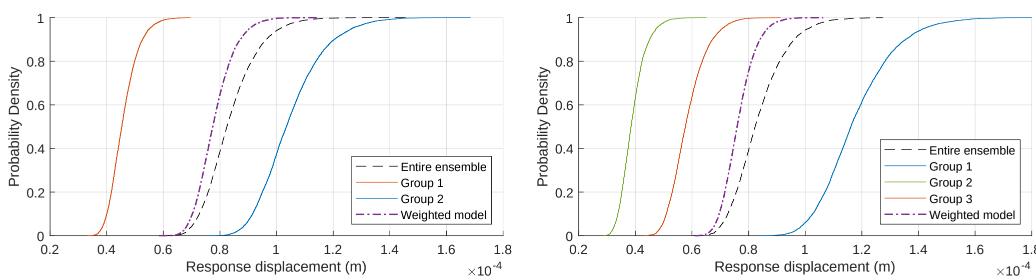

Figure 15: CDFs of the maximum response displacement of the linear mass-spring-damper system for scenario C for the classification into 2 groups (left) and into 3 groups (right) of the site-specific ensemble.

The results of scenario D, using the site-specific data and its classification, are shown in Fig. 16. As the natural frequency has changed due to the use of other system parameters, correspondingly different simulation results can be obtained. Since the spectral densities are now somewhat higher compared to scenario C, the system displacements are also in part significantly higher. On the left side of Fig. 16 the results for the classification into 2 groups are shown, while on the right side the results for the classification into 3 groups are given. In particular, the classification into 3 groups reveals high distances between the CDFs of the individual groups and also the weighted mean CDF is slightly further shifted to left side from the ensemble mean CDF for the case classified into 3 groups than the case classified into 2 groups, which indicates that a classification into 3 groups is optimal. In both cases, reasonable weighted mean CDFs are calculated based on a refined subdivision of the ensemble.

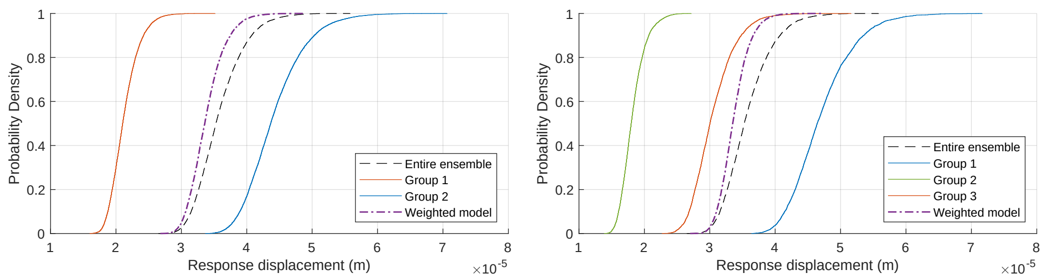

Figure 16: CDFs of the maximum response displacement of the linear mass-spring-damper system for scenario D for the classification into 2 groups (left) and into 3 groups (right) of the site-specific ensemble.

5.2. Non-linear bridge pier model

For the numerical investigation of a non-linear system a seismic-isolated bridge pier model with rubber bearings is utilised. The model is based on the design specifications for highway bridges of the Japan Road Association [65] and the manual on bearings for highway bridges [66]. The bridge pier is modelled as a 2-DOF lumped mass system and consists of a superstructure and the RC pier, which is modelled as a non-linear horizontal spring, see Fig. 17. The rubber bearings are idealised as a bilinear model, while for the RC pier a bilinear model with elastoplastic characteristics and stiffness degradation model is used, the so-called Takeda model [67]. A fixed boundary condition is assumed for the connection to the surface. Rayleigh damping is adopted, with the damping ratios of 0% for the bearing and 2 % for the pier, respectively. For the numerical solution, a dynamic response analysis is performed using the Newmark-beta method with $\gamma = 1/2$, $\beta = 1/4$ and the time step size $\Delta t = 0.01$ s. The utilised structural parameters are given in Table 6.

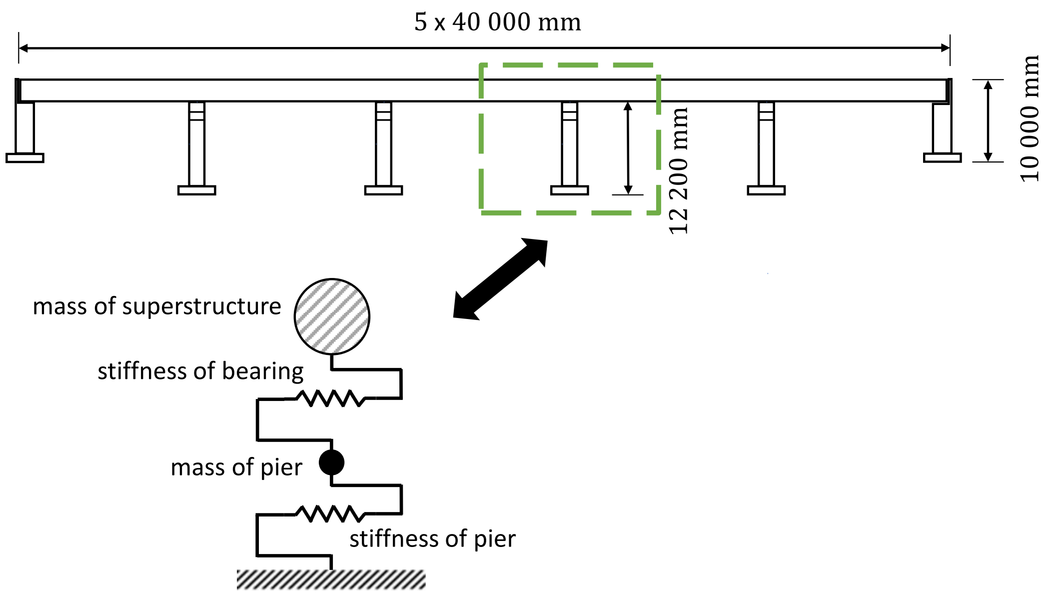

Figure 17: 2-DOF lumped mass model for the target bridge pier.

Table 6: Model parameters of the bridge pier.

Model parameter		Nominal value
Superstructure	Mass M_S (ton)	604
Rubber bearing	Yield strength (kN)	1118
	Yield stiffness K_{B1}	40,000
	Post-yield stiffness K_{B2}	6000
RC pier	Mass M_p (ton)	346.2
	Yield strength (kN)	3374
	Yield displacement (m)	0.0306
	Ultimate displacement (m)	0.251
	Yield stiffness K_p (kN/m)	110100

An example of the non-linear force-displacement behaviour of the rubber bearing of the bridge pier model is depicted in Fig. 18.

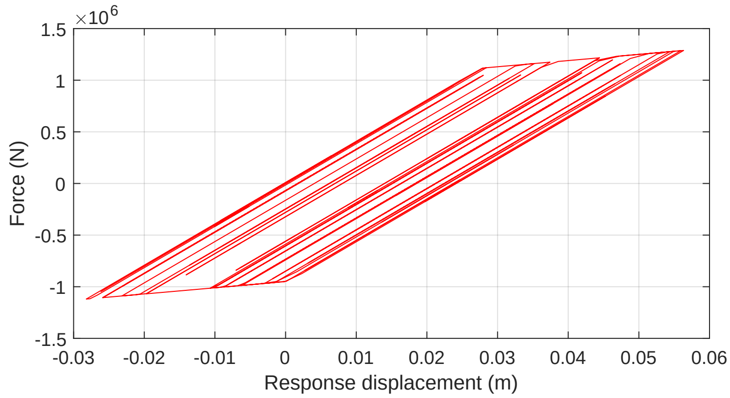

Figure 18: Force-displacement behaviour of the rubber bearings.

As in the example of the linear mass-spring-damper model in the preceding section, a total of 10,000 MC samples were generated and applied to the bridge pier model. For each sample, the maximum displacement of the system, i.e. the maximum displacements at the RC pier and at the rubber bearings, is determined in the time domain, from which the CDF is calculated again. In this case, only non-stationary earthquake ground motions were utilised as samples to provide a more realistic example. It is important to note that unlike the previous case of the linear system, it is difficult to discuss about the validity of the identified optimal number for classification, since response properties of the non-linear dynamic systems are significantly affected by the structural non-linearity and non-stationarity of the input ground motions. This example rather aims to demonstrate the feasibility of the proposed classification approach in reliability assessment of non-linear dynamic systems and thus, for the sake of brevity, only the results with the optimal number of classifications are presented.

5.2.1. *Results of the non-linear bridge pier model*

The results of the source-specific data are given in Fig. 19. The CDFs of the maximum displacements at the RC pier of the bridge for the optimal number of groups $k_{opt} = 2$ and of those at the rubber bearings are shown. The results demonstrate for the non-linear model that the classification of the ensemble yields more accurate results. With the classification into 2 groups, it can be seen that higher overall system displacements can be calculated with the load model generated from group 1 than with the load model of the entire ensemble. In this example it is again confirmed that the classification of an ensemble leads to more accurate results. It can also be seen that the weighted mean CDF deviates slightly from the CDF of the entire ensemble for the RC pier case, while they overlap each other for the rubber bearings.

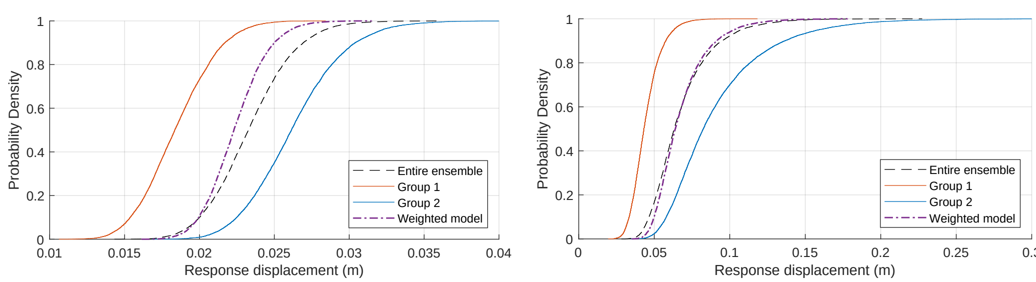

Figure 19: CDFs of the maximum response displacement of the seismic-isolated bridge pier model for the source-specific ensemble for the classification into $k_{opt} = 2$ groups. Results for the RC pier are shown on the left, results for the rubber bearings are shown on the right.

In Fig. 20 the results of the site-specific data are shown. The CDFs for the optimal number of $k_{opt} = 3$ groups are shown for the RC pier (left) and for the rubber bearings (right). It can be seen that the classification leads to more accurate results instead of considering the entire ensemble and determine a load model from it. The classification into 3 groups shows a high diversity of the CDFs, which is as a consequence of the optimal number of groups having been determined to be $k_{opt} = 3$. This also leads to the fact that the weighted mean CDF here partly deviates strongly from that of the entire ensemble for each of the cases. Overlaps can only be seen in the range of small system displacements.

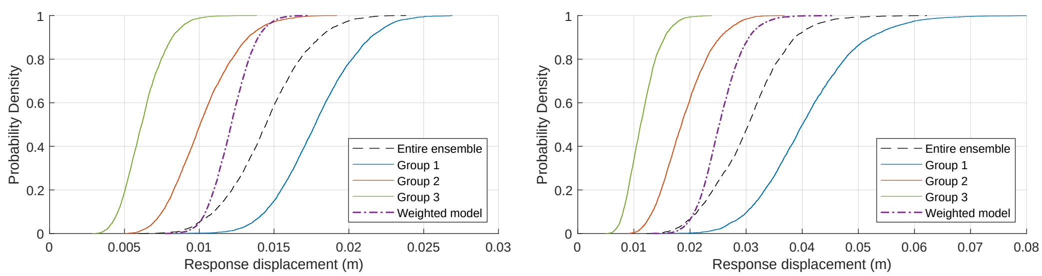

Figure 20: CDFs of the maximum response displacement of the seismic-isolated bridge pier model for the site-specific ensemble for the classification into $k_{opt} = 3$ groups. Results for the RC pier are shown on the left, results for the rubber bearings are shown on the right.

The results demonstrate that the classification of the ensemble can cover wider ranges of the system responses. Moreover, except for the results at the rubber bearings for the source-specific case, the weighted mean CDFs provide more accurate results than the ensemble CDF. These results thus demonstrate the feasibility of the proposed method for reliability assessment of non-linear dynamic systems.

6. Conclusion

A new technique has been proposed for developing load models from ensembles that exhibit high spectral variance. Using the Bhattacharyya distance, groups of similar PSD functions in the frequency domain can be determined applying the k-means algorithm. Classification in the frequency domain is necessary because differences in the time domain often cannot be detected; similar signals in the time domain can lead to highly differing PSD functions in the frequency domain. The dissimilarities can often only be revealed there. The classification of the ensemble leads to more accurate simulation results, which can be important especially for the reliability assessment of the structure under investigation. In many cases, the higher number of load models yields in higher system displacements that would otherwise remain undetected and a possible system failure would thus not be detected. However, it requires that multiple, distinct simulations of the structural behaviour must be carried out, which could equate to a significant time investment for large model analysis. Therefore, a method for identifying the optimal number for classification based on the silhouette method was also proposed to avoid performing more simulations than necessary. The results of the individual groups can be weighted considering the probability of occurrence of each load model to obtain a more accurate overall system response, which can then be evaluated for design purposes. This may allow the use of modified system parameters in the design of the structure or lead to cost savings in the computations of the simulations. The validity of the identified optimal number for classification and the strength of the proposed method were first investigated using a linear mass-spring-damper system, and then the proposed method was applied to a seismic-isolated bridge pier model to demonstrate its feasibility in reliability assessment of non-linear dynamic systems. While the application in this work is based on seismic ground motions, the developed approach is also suitable for other stochastic processes, such as wind or wave loads subject to structures. The prerequisite for the application of this approach to other stochastic processes is that they exhibit similar characteristics among themselves after the transformation into the frequency domain, otherwise a classification would not be useful as it would be obvious that the data are dissimilar. If they show similar characteristics but high variance, classification is indispensable.

Acknowledgement

This research is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) with grants BE 2570/4–1 and CO 1849/1–1.

References

[1] A. Powell, S. Crandall, Random Vibration, The Technology Press of the Massachusetts Institute of Technology, Cambridge, 1958.

[2] T. Soong, M. Grigoriu, Random Vibration of Mechanical and Structural Systems, PTR Prentice Hall, 1993.
URL https://books.google.de/books?id=6JVRAAAAMAAJ

[3] J. B. Roberts, P. D. Spanos, Random vibration and statistical linearization, Courier Corporation, 2003.

[4] L. D. Lutes, S. Sarkani, Random vibrations: analysis of structural and mechanical systems, Butterworth-Heinemann, 2004.

[5] Y.-K. Lin, G.-Q. Cai, Probabilistic structural dynamics: advanced theory and applications, McGraw-Hill New York, 1995.

[6] J. Li, J. Chen, Stochastic Dynamics of Structures, John Wiley & Sons, 2009.

[7] G. I. Schuëller, M. Shinozuka, Stochastic Methods in Structural Dynamics, Vol. 10 of Mechanics: Dynamical Systems, Springer Science & Business Media, 2012.
URL `https://books.google.de/books?id=BaXzCAAAQBAJ`

[8] A. K. Chopra, Dynamics of structures: Theory and applications to earthquake engineering, Englewood Cliffs, N.J: Prentice Hall, 1995.

[9] R. A. Muller, G. J. MacDonald, Ice ages and astronomical causes: data, spectral analysis and mechanisms, Springer Science & Business Media, 2002.

[10] G. W. Housner, Characteristics of strong-motion earthquakes, Bulletin of the Seismological Society of America 37 (1) (1947) 19–31.

[11] M. Shinozuka, G. Deodatis, Simulation of stochastic processes by spectral representation, Applied Mechanics Reviews 44 (4) (1991) 191–204. doi:10.1115/1.3119501.

[12] G. I. Schuëller, Efficient monte carlo simulation procedures in structural uncertainty and reliability analysis - recent advances, Structural Engineering and Mechanics 32 (1) (2009) 1–20. doi:https://doi.org/10.12989/SEM.2009.32.1.001.

[13] E. Zio, The Monte Carlo Simulation Method for System Reliability and Risk Analysis, Springer London, 2013. doi:https://doi.org/10.1007/978-1-4471-4588-2.

[14] M. Grigoriu, Stochastic Calculus: Applications in Science and Engineering, Springer Science & Business Media, 2013.

[15] Rune Brincker and Lingmi Zhang and Palle Andersen, Modal identification of output-only systems using frequency domain decomposition, Smart Materials and Structures 10 (3) (2001) 441–445. doi:10.1088/0964-1726/10/3/303.
URL https://doi.org/10.1088/0964-1726/10/3/303

[16] B. Peeters, G. De Roeck, Stochastic System Identification for Operational Modal Analysis: A Review, Journal of Dynamic Systems, Measurement, and Control 123 (4) (2001) 659–667. arXiv:https://asmedigitalcollection.asme.org/

dynamicsystems/article-pdf/123/4/659/5514360/659_1.pdf, doi:10.1115/1.1410370.
URL https://doi.org/10.1115/1.1410370

[17] S. Gade, Frequency domain techniques for operational modal analysis, The Shock and Vibration Digest 38 (6) (2006) 537–538.
URL https://link.gale.com/apps/doc/A152997313/AONE?u=anon~9b6dd377&sid=googleScholar&xid=2a59f38d

[18] T. Wang, O. Celik, F. Catbas, L. Zhang, A frequency and spatial domain decomposition method for operational strain modal analysis and its application, Engineering Structures 114 (2016) 104–112. doi:https://doi.org/10.1016/j.engstruct.2016.02.011.
URL https://www.sciencedirect.com/science/article/pii/S0141029616000997

[19] Ç. Hızal, E. Aktaş, Probabilistic investigation of error propagation in frequency domain decomposition-based operational modal analysis, Structural Control and Health Monitoring 28 (8) (2021) e2759. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/stc.2759, doi:https://doi.org/10.1002/stc.2759.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/stc.2759

[20] B. R. Ellingwood, Earthquake risk assessment of building structures, Reliability Engineering & System Safety 74 (3) (2001) 251–262.

doi:https://doi.org/10.1016/S0951-8320(01)00105-3.
URL https://www.sciencedirect.com/science/article/pii/S0951832001001053

[21] P. Code, Eurocode 8: Design of structures for earthquake resistance-part 1: general rules, seismic actions and rules for buildings, Brussels: European Committee for Standardization (2005).

[22] D. Vamvatsikos, C. A. Cornell, Incremental dynamic analysis, Earthquake Engineering & Structural Dynamics 31 (3) (2002) 491–514. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/eqe.141, doi:https://doi.org/10.1002/eqe.141.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/eqe.141

[23] D. Vamvatsikos, C. A. Cornell, Applied Incremental Dynamic Analysis, Earthquake Spectra 20 (2) (2004) 523–553. arXiv:https://doi.org/10.1193/1.1737737, doi:10.1193/1.1737737.
URL https://doi.org/10.1193/1.1737737

[24] D. Vamvatsikos, M. Fragiadakis, Incremental dynamic analysis for estimating seismic performance sensitivity and uncertainty, Earthquake Engineering & Structural Dynamics 39 (2) (2010) 141–163. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/eqe.935, doi:https://doi.org/10.1002/eqe.935.

URL https://onlinelibrary.wiley.com/doi/abs/10.1002/eqe.935

[25] Keith Porter and Robert Kennedy and Robert Bachman, Creating Fragility Functions for Performance-Based Earthquake Engineering, Earthquake Spectra 23 (2) (2007) 471–489. arXiv:https://doi.org/10.1193/1.2720892, doi:10.1193/1.2720892.
URL https://doi.org/10.1193/1.2720892

[26] M. G. Sfahani, H. Guan, Y.-C. Loo, Seismic Reliability and Risk Assessment of Structures Based on Fragility Analysis - A Review, Advances in Structural Engineering 18 (10) (2015) 1653–1669. arXiv:https://doi.org/10.1260/1369-4332.18.10.1653, doi:10.1260/1369-4332.18.10.1653.
URL https://doi.org/10.1260/1369-4332.18.10.1653

[27] I. Zentner, M. Gündel, N. Bonfils, Fragility analysis methods: Review of existing approaches and application, Nuclear Engineering and Design 323 (2017) 245–258. doi:https://doi.org/10.1016/j.nucengdes.2016.12.021.
URL https://www.sciencedirect.com/science/article/pii/S0029549316305209

[28] Ghosh, Swarup and Ghosh, Shyamal and Chakraborty, Subrata, Seismic fragility analysis in the probabilistic performance-based earthquake engineering framework: an overview, International Journal of Advances in

Engineering Sciences and Applied Mathematics 13 (1) (2021) 122–135. doi:10.1007/s12572-017-0200-y.
URL https://doi.org/10.1007/s12572-017-0200-y

[29] S. Rajkumari, K. Thakkar, H. Goyal, Fragility analysis of structures subjected to seismic excitation: A state-of-the-art review, Structures 40 (2022) 303–316. doi:https://doi.org/10.1016/j.istruc.2022.04.023.
URL https://www.sciencedirect.com/science/article/pii/S2352012422002843

[30] G. Lupoi, P. Franchin, A. Lupoi, P. E. Pinto, Seismic Fragility Analysis of Structural Systems, Journal of Engineering Mechanics 132 (4) (2006) 385–395. doi:10.1061/(ASCE)0733-9399(2006)132:4(385).

[31] N. D. Lagaros, Y. Tsompanakis, P. N. Psarropoulos, E. C. Georgopoulos, Computationally efficient seismic fragility analysis of geostructures, Computers & Structures 87 (19) (2009) 1195–1203, civil-Comp Special Issue. doi:https://doi.org/10.1016/j.compstruc.2008.12.001.
URL https://www.sciencedirect.com/science/article/pii/S0045794908002721

[32] S. Ghosh, S. Chakraborty, Seismic fragility analysis of structures based on Bayesian linear regression demand models, Probabilistic Engineering Mechanics 61 (2020) 103081. doi:https:

//doi.org/10.1016/j.probengmech.2020.103081.
URL https://www.sciencedirect.com/science/article/pii/S0266892020300680

[33] A. M. Billah, M. S. Alam, Seismic fragility assessment of highway bridges: a state-of-the-art review, Structure and Infrastructure Engineering 11 (6) (2015) 804–832. arXiv:https://doi.org/10.1080/15732479.2014.912243, doi:10.1080/15732479.2014.912243.
URL https://doi.org/10.1080/15732479.2014.912243

[34] M. A. Hariri-Ardebili, V. E. Saouma, Seismic fragility analysis of concrete dams: A state-of-the-art review, Engineering Structures 128 (2016) 374–399. doi:https://doi.org/10.1016/j.engstruct.2016.09.034.
URL https://www.sciencedirect.com/science/article/pii/S014102961630654X

[35] S. Misra, J. E. Padgett, Seismic Fragility of Railway Bridge Classes: Methods, Models, and Comparison with the State of the Art, Journal of Bridge Engineering 24 (12) (2019) 04019116. doi:10.1061/(ASCE)BE.1943-5592.0001485.

[36] X. Chen, N. Xiang, Z. Guan, J. Li, Seismic vulnerability assessment of tall pier bridges under mainshock-aftershock-like earthquake sequences using vector-valued intensity measure, Engineering Structures 253 (2022) 113732. doi:https:

//doi.org/10.1016/j.engstruct.2021.113732.
URL https://www.sciencedirect.com/science/article/pii/S0141029621018186

[37] A. R. Ghotbi, E. Taciroglu, Ground motion selection based on a multi-intensity-measure conditioning approach with emphasis on diverse earthquake contents, Earthquake Engineering & Structural Dynamics 50 (5) (2021) 1378–1394. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/eqe.3383, doi:https://doi.org/10.1002/eqe.3383.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/eqe.3383

[38] Matthieu Perrault and Philippe Guéguen, Correlation between Ground Motion and Building Response using California Earthquake Records, Earthquake Spectra 31 (4) (2015) 2027–2046. arXiv:https://doi.org/10.1193/062413EQS168M, doi:10.1193/062413EQS168M.
URL https://doi.org/10.1193/062413EQS168M

[39] Christine A Goulet and Tadahiro Kishida and Timothy D Ancheta and Chris H Cramer and Robert B Darragh and Walter J Silva and Youssef MA Hashash and Joseph Harmon and Grace A Parker and Jonathan P Stewart and Robert R Youngs, PEER NGA-East database, Earthquake Spectra 37 (1_suppl) (2021) 1331–1353. doi:10.1177/

87552930211015695.
URL https://doi.org/10.1177/87552930211015695

[40] Berry, Michael and Parrish, Myles and Eberhard, Marc, PEER structural performance database user's manual (version 1.0), University of California, Berkeley (2004).

[41] K. A. Porter, An overview of PEER's performance-based earthquake engineering methodology, in: Proceedings of ninth international conference on applications of statistics and probability in civil engineering, 2003, pp. 1–8.

[42] National Research Institute for Earth Science and Disaster Resilience, NIED K-NET, KiK-net (2019). doi:10.17598/NIED.0004.

[43] T. Kishida, V. Contreras, Y. Bozorgnia, N. A. Abrahamson, Ahdi, T. S.K., Ancheta, D. Boore, K. Campbell, B. Chiou, R. Darragh, N. Gregor, N. Kuehn, D. Kwak, A. Kwok, P. Lin, H. Magistrale, S. Maazoni, S. Muin, S. Midorikawa, H. Si, W. Silva, J. Stewart, K. Wooddell, R. R. Youngs, NGA-Sub Ground Motion Database, Proceedings of the Eleventh U.S. National Conference on Earthquake Engineering (2018).

[44] E. I. Katsanos, A. G. Sextos, G. D. Manolis, Selection of earthquake ground motion records: A state-of-the-art review from a structural engineering perspective, Soil Dynamics and Earthquake Engineering 30 (4) (2010) 157–169. doi:https:

//doi.org/10.1016/j.soildyn.2009.10.005.
URL https://www.sciencedirect.com/science/article/pii/S0267726109001602

[45] Subash Ghimire and Philippe Guéguen and Ariana Astorga, Analysis of the efficiency of intensity measures from real earthquake data recorded in buildings, Soil Dynamics and Earthquake Engineering 147 (2021) 106751. doi:https://doi.org/10.1016/j.soildyn.2021.106751.
URL https://www.sciencedirect.com/science/article/pii/S0267726121001731

[46] Burks, Lynne S. and Baker, Jack W., Validation of Ground-Motion Simulations through Simple Proxies for the Response of Engineered Systems, Bulletin of the Seismological Society of America 104 (4) (2014) 1930–1946. arXiv:https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/104/4/1930/2676473/1930.pdf, doi:10.1785/0120130276.
URL https://doi.org/10.1785/0120130276

[47] A. K. Jain, Data clustering: 50 years beyond K-means, Pattern Recognition Letters 31 (8) (2010) 651–666, award winning papers from the 19th International Conference on Pattern Recognition (ICPR). doi:https://doi.org/10.1016/j.patrec.2009.09.011.
URL https://www.sciencedirect.com/science/article/pii/S0167865509002323

[48] M. Honarkhah, J. Caers, Stochastic simulation of patterns using distance-based pattern modeling, Mathematical Geosciences 42 (5) (2010) 487–517. doi:10.1007/s11004-010-9276-7.

[49] H. B. Seed, C. Ugas, J. Lysmer, Site-dependent spectra for earthquake-resistant design, Bulletin of the Seismological Society of America 66 (1) (1976) 221–243. arXiv:https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/66/1/221/5320247/bssa0660010221.pdf, doi:10.1785/BSSA0660010221.
URL https://doi.org/10.1785/BSSA0660010221

[50] S. Yaghmaei-Sabegh, A novel approach for classification of earthquake ground-motion records, Journal of Seismology 21 (4) (2017) 885–907. doi:10.1007/s10950-017-9642-8.

[51] W. B. Joyner, D. M. Boore, Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 imperial valley, California, earthquake, Bulletin of the Seismological Society of America 71 (6) (1981) 2011–2038. arXiv:https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/71/6/2011/5330048/bssa0710062011.pdf, doi:10.1785/BSSA0710062011.
URL https://doi.org/10.1785/BSSA0710062011

[52] Y. Ding, Y. Peng, J. Li, A stochastic semi-physical model of seismic ground motions in time domain, Journal of Earthquake and

Tsunami 12 (03) (2018) 1850006. arXiv:https://doi.org/10.1142/S1793431118500069, doi:10.1142/S1793431118500069.
URL https://doi.org/10.1142/S1793431118500069

[53] Y. Ding, Y. Peng, J. Li, Cluster analysis of earthquake ground-motion records and characteristic period of seismic response spectrum, Journal of Earthquake Engineering 24 (6) (2020) 1012–1033. arXiv:https://doi.org/10.1080/13632469.2018.1453420, doi:10.1080/13632469.2018.1453420.
URL https://doi.org/10.1080/13632469.2018.1453420

[54] A. Alimoradi, S. Pezeshk, F. Naeim, H. Frigui, Fuzzy patterm classification of strong ground motion records, Journal of Earthquake Engineering 9 (3) (2005) 307–332. arXiv:https://doi.org/10.1080/13632460509350544, doi:10.1080/13632460509350544.
URL https://doi.org/10.1080/13632460509350544

[55] A. Azarbakht, Z. Minaei, Adaptive fuzzy c-mean clustering of ground motion prediction equations, Journal of Seismology and Earthquake Engineering 19 (1) (2017) 25–35.

[56] A. Bhattacharyya, On a measure of divergence between two statistical populations defined by their probability distributions, Bull. Calcutta Math. Soc. 35 (1943) 99–109.

[57] A. Bhattacharyya, On a Measure of Divergence between Two Multino-

mial Populations, Sankhyā: The Indian Journal of Statistics (1933-1960) 7 (4) (1946) 401–406.
URL http://www.jstor.org/stable/25047882

[58] S. Bi, S. Prabhu, S. Cogan, S. Atamturktur, Uncertainty quantification metrics with varying statistical information in model calibration and validation, AIAA Journal 55 (10) (2017) 3570–3583. arXiv:https://doi.org/10.2514/1.J055733, doi:10.2514/1.J055733.
URL https://doi.org/10.2514/1.J055733

[59] B. Priestley, Spectral Analysis and Time Series, Academic Press, 1982.

[60] D. Newland, An Introduction to Random Vibrations, Spectral & Wavelet Analysis: Third Edition, Dover Civil and Mechanical Engineering, Dover Publications, 2012.
URL https://books.google.de/books?id=mE9owVg1U1IC

[61] P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, Journal of Computational and Applied Mathematics 20 (1987) 53–65. doi:https://doi.org/10.1016/0377-0427(87)90125-7.
URL https://www.sciencedirect.com/science/article/pii/0377042787901257

[62] L. Kaufman, P. J. Rousseeuw, Finding groups in data: an introduction

to cluster analysis, Vol. 344, John Wiley & Sons, 2009. doi:10.1002/9780470316801.

[63] C. Patil, I. Baidari, Estimating the optimal number of clusters k in a dataset using data depth, Data Science and Engineering 4 (2) (2019) 132–140. doi:10.1007/s41019-019-0091-y.

[64] L. F. Shampine, M. W. Reichelt, The MATLAB ODE Suite, SIAM Journal on Scientific Computing 18 (1) (1997) 1–22. arXiv:https://doi.org/10.1137/S1064827594276424, doi:10.1137/S1064827594276424.
URL https://doi.org/10.1137/S1064827594276424

[65] Japan Road Association, Specifications for highway bridges, Part V: Seismic Design, Maruzen, Tokyo, Japan (2016).

[66] Japan Road Association, Manual on bearings for highway bridges (in Japanese), Maruzen, Tokyo, Japan (2004).

[67] T. Takeda, M. A. Sozen, N. N. Nielsen, Reinforced concrete response to simulated earthquakes, Journal of the Structural Division 96 (12) (1970) 2557–2573. arXiv:https://ascelibrary.org/doi/pdf/10.1061/JSDEAG.0002765, doi:10.1061/JSDEAG.0002765.
URL https://ascelibrary.org/doi/abs/10.1061/JSDEAG.0002765