

Uncertainty modelling in power spectrum estimation of environmental processes

Von der Fakultät für Bauingenieurwesen und Geodäsie
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des Grades

Doktor-Ingenieur
Dr.-Ing.

genehmigte Dissertation
von

Marco Behrendt, M. Sc.

2022

Referent: Prof. Dr.-Ing. Michael Beer
Korreferenten: Prof. Dr. Jianbing Chen
Prof. Dr.-Ing. Ingo Neumann
Prof. Dr. sc. nat. ETH Insa Neuweiler

Tag der Promotion: 23. November 2022

Eidesstattliche Versicherung

Hiermit versichere ich, die Regeln der geltenden Promotionsordnung zu kennen und eingehalten zu haben und mit einer Prüfung nach den Bestimmungen der Promotionsordnung einverstanden zu sein, die Dissertation selbst verfasst zu haben, keine Textabschnitte von Dritten oder eigener Prüfungsarbeiten ohne Kennzeichnung übernommen und alle von ihr oder ihm benutzten Hilfsmittel und Quellen in seiner Arbeit angegeben zu haben, Dritten weder unmittelbar noch mittelbar geldwerte Leistungen für Vermittlungstätigkeiten oder für die inhaltliche Ausarbeitung der Dissertation erbracht zu haben, die Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung oder die gleiche oder eine in wesentlichen Teilen ähnliche Arbeit bei einer anderen Hochschule als Dissertation eingereicht zu haben.

Hannover, 12. September 2022

.....
(Unterschrift)

Zusammenfassung

Für eine effiziente Zuverlässigkeitsanalyse von Gebäuden und Strukturen sind robuste Belastungsmodelle in der stochastischen Dynamik erforderlich, die insbesondere aus Umweltprozessen wie Erdbeben oder Windlasten geschätzt werden können. Um das Antwortverhalten eines dynamischen Systems unter solchen Belastungen zu bestimmen, ist die Funktion der Leistungsspektraldichte (PSD) ein weit verbreitetes Werkzeug zur Identifizierung der Frequenzkomponenten und der entsprechenden Amplituden von Umweltprozessen. Da die zu diesem Zweck benötigten realen Datensätze häufig mit aleatorischen und epistemischen Unsicherheiten behaftet sind und der PSD-Schätzprozess selbst weitere Unsicherheiten induzieren kann, ist eine strenge Quantifizierung dieser Unsicherheiten unerlässlich, da andernfalls ein sehr ungenaues Belastungsmodell erzeugt werden könnte, das zu fehlerhaften Simulationsergebnissen führen kann. Ein eigentlich katastrophales Systemverhalten kann so in einen akzeptablen Bereich verschoben werden, so dass das System als sicher eingestuft wird, obwohl es einem hohen Risiko der Beschädigung oder des Zusammenbruchs ausgesetzt ist. Um diese Probleme anzugehen, werden alternative Belastungsmodelle vorgeschlagen, die probabilistische und nicht-deterministische Modelle verwenden, welche in der Lage sind, diese Unsicherheiten effizient zu berücksichtigen und die Belastungen entsprechend zu modellieren. Bei der Erstellung dieser Lastmodelle werden verschiedene Methoden verwendet, die insbesondere nach dem Charakter der Daten und der Anzahl der verfügbaren Datensätze ausgewählt werden.

Wenn mehrere Datensätze verfügbar sind, können zuverlässige statistische Informationen aus einer Reihe ähnlicher PSD-Funktionen extrahiert werden, die sich z.B. nur geringfügig in Form und Spitzenfrequenz unterscheiden. Auf der Grundlage dieser Statistiken wird ein Modell der PSD-Funktion abgeleitet, das subjektive Wahrscheinlichkeiten verwendet, um die epistemischen Unsicherheiten zu erfassen und diese Informationen effektiv darzustellen. Die spektralen Leistungsdichten werden als Zufallsvariablen charakterisiert, anstatt diskrete Werte zu verwenden, somit stellt die PSD-Funktion selbst einen nicht-stationären Zufallsprozess dar, der einen Bereich möglicher gültiger PSD-Funktionen für einen gegebenen Datensatz umfasst.

Wenn nur eine begrenzte Anzahl von Datensätzen zur Verfügung steht, ist es nicht möglich, solche zuverlässigen statistischen Informationen abzuleiten. Daher wird ein intervallbasierter Ansatz vorgeschlagen, der nur eine obere und untere Grenze bestimmt und sich nicht auf eine Verteilung innerhalb dieser Grenzen stützt. Ein Satz von diskret wertigen PSD-Funktionen wird in eine intervallwertige PSD-Funktion umgewandelt, indem die Gewichte von vorab abgeleiteten Basisfunktionen aus einem Radialbasisfunktionsnetz so optimiert werden, dass sie eine obere und

untere Grenze bilden, die den Datensatz umfassen. Damit wird ein Bereich möglicher Werte und Systemreaktionen anstelle diskreter Werte ermittelt, welche in der Lage sind, epistemische Unsicherheiten zu erfassen.

Bei der Erstellung eines solchen Lastmodells aus realen Datensätzen kann das Problem auftreten, dass die einzelnen Datensätze eine hohe spektrale Varianz im Frequenzbereich aufweisen und sich daher zu stark voneinander unterscheiden, obwohl sie im Zeitbereich ähnlich erscheinen. Ein aus diesen Daten abgeleitetes Lastmodell deckt möglicherweise nicht den gesamten Spektralbereich ab und ist daher nicht repräsentativ. Die Daten werden daher mit Hilfe der Bhattacharyya-Distanz und des k -means-Algorithmus nach ihrer Ähnlichkeit gruppiert, wodurch zwei oder mehr Belastungsmodelle aus dem gesamten Datensatz erzeugt werden können. Diese können separat auf die zu untersuchende Struktur angewandt werden, was zu genaueren Simulationsergebnissen führt. Dieser Ansatz kann auch zur Schätzung der spektralen Ähnlichkeit einzelner Datensätze im Frequenzbereich verwendet werden, was für die oben genannten Lastmodelle besonders relevant ist.

Wenn die Unsicherheiten direkt im Zeitsignal modelliert werden, kann es eine schwierige Aufgabe sein, sie effizient in den Frequenzbereich zu transformieren. Ein solches Signal kann möglicherweise nur aus zuverlässigen Grenzen bestehen, in denen das tatsächliche Signal liegt. Es wird eine Methode vorgestellt, mit der diese Intervallunsicherheit automatisch durch die diskrete Fourier Transformation propagiert werden kann, um die exakten Grenzen der Fourier-Amplitude und der Schätzung der PSD-Funktion zu erhalten. Die Methode ermöglicht es, ein solches Intervallsignal zu propagieren, ohne Annahmen über die Abhängigkeit und Verteilung des Fehlers über die Zeitschritte zu treffen.

Diese neuartigen Darstellungen von Lastmodellen sind in der Lage, epistemische Unsicherheiten zu quantifizieren, die in realen Datensätzen enthalten sind und durch den PSD-Schätzprozess induziert werden. Die Stärken und Vorteile dieser Ansätze in der Praxis werden anhand mehrerer numerischer Beispiele aus dem Bereich der stochastischen Dynamik demonstriert.

Schlüsselwörter: Stochastische Dynamik, Schätzung der Leistungsspektraldichte, Stochastische Prozesse, Zufällige Schwingungen, Quantifizierung von Unsicherheiten, Ungenaue Wahrscheinlichkeiten.

Abstract

For efficient reliability analysis of buildings and structures, robust load models are required in stochastic dynamics, which can be estimated in particular from environmental processes, such as earthquakes or wind loads. To determine the response behaviour of a dynamic system under such loads, the power spectral density (PSD) function is a widely used tool for identifying the frequency components and corresponding amplitudes of environmental processes. Since the real data records required for this purpose are often subject to aleatory and epistemic uncertainties, and the PSD estimation process itself can induce further uncertainties, a rigorous quantification of these is essential, as otherwise a highly inaccurate load model could be generated which may yield in misleading simulation results. A system behaviour that is actually catastrophic can thus be shifted into an acceptable range, classifying the system as safe even though it is exposed to a high risk of damage or collapse. To address these issues, alternative loading models are proposed using probabilistic and non-deterministic models, that are able to efficiently account for these uncertainties and to model the loadings accordingly. Various methods are used in the generation of these load models, which are selected in particular according to the characteristic of the data and the number of available records.

In case multiple data records are available, reliable statistical information can be extracted from a set of similar PSD functions that differ, for instance, only slightly in shape and peak frequency. Based on these statistics, a PSD function model is derived utilising subjective probabilities to capture the epistemic uncertainties and represent this information effectively. The spectral densities are characterised as random variables instead of employing discrete values, and thus the PSD function itself represents a non-stationary random process comprising a range of possible valid PSD functions for a given data set.

If only a limited amount of data records is available, it is not possible to derive such reliable statistical information. Therefore, an interval-based approach is proposed that determines only an upper and lower bound and does not rely on any distribution within these bounds. A set of discrete-valued PSD functions is transformed into an interval-valued PSD function by optimising the weights of pre-derived basis functions from a Radial Basis Function Network such that they compose an upper and lower bound that encompasses the data set. Therefore, a range of possible values and system responses are identified rather than discrete values, which are able to quantify the epistemic uncertainties.

When generating such a load model using real data records, the problem can arise that the individual records exhibit a high spectral variance in the frequency domain and therefore differ

too much from each other, although they appear to be similar in the time domain. A load model derived from these data may not cover the entire spectral range and is therefore not representative. The data are therefore grouped according to their similarity using the Bhattacharyya distance and k -means algorithm, which may generate two or more load models from the entire data set. These can be applied separately to the structure under investigation, leading to more accurate simulation results. This approach can also be used to estimate the spectral similarity of individual data sets in the frequency domain, which is particularly relevant for the load models mentioned above.

If the uncertainties are modelled directly in the time signal, it can be a challenging task to transform them efficiently into the frequency domain. Such a signal may consist only of reliable bounds in which the actual signal lies. A method is presented that can automatically propagate this interval uncertainty through the discrete Fourier transform, obtaining the exact bounds on the Fourier amplitude and an estimate of the PSD function. The method allows such an interval signal to be propagated without making assumptions about the dependence and distribution of the error over the time steps.

These novel representations of load models are able to quantify epistemic uncertainties inherent in real data records and induced due to the PSD estimation process. The strengths and advantages of these approaches in practice are demonstrated by means of several numerical examples concentrated in the field of stochastic dynamics.

Keywords: Stochastic dynamics, Power spectral density estimation, Stochastic processes, Random vibrations, Uncertainty quantification, Imprecise probabilities.

Contents

1	Introduction	1
1.1	Research background	2
1.1.1	Uncertainties	2
1.1.2	Stochastic processes and power spectral density estimation	12
1.2	Aims and objectives	27
1.3	Original contributions	28
1.4	Structure of the thesis	31
2	Relaxed power spectrum estimation from multiple data records utilising subjective probabilities	33
2.1	Introduction	34
2.2	Stochastic process representation and power spectrum estimation	37
2.3	Relaxed power spectrum and relaxed spectral representation method	37
2.3.1	Convergence of the relaxed PSD function	41
2.3.2	Influence on Generated Time Signals	42
2.4	Numerical examples	44
2.4.1	Linear SDOF oscillator	44
2.4.2	Non-linear Bouc-Wen-Baber-Noori model	46
2.5	Conclusion	50
3	Stochastic process generation from relaxed power spectra utilising stochastic harmonic functions	53
3.1	Introduction	54
3.2	Spectral representation and PSD function estimation	56
3.3	Stochastic harmonic function representation	57
3.4	Relaxed PSD function	58
3.4.1	Relaxed PSD function utilising a truncated normal distribution	58
3.4.2	Relaxed spectral representation method	59
3.4.3	Relaxed stochastic harmonic function representation	59
3.5	Benchmark setup and results	60
3.6	Conclusion	64
4	Estimation of an imprecise power spectral density function with optimised bounds from scarce data for epistemic uncertainty quantification	66
4.1	Introduction	67
4.2	Preliminaries	69
4.2.1	PSD estimation and stochastic processes	69
4.2.2	Radial basis function networks	71
4.3	Method development	72
4.3.1	Basis power spectrum	72
4.3.2	Fitting an RBF network	73
4.3.3	Obtaining optimised bounds	74
4.3.4	Optimisation of the hyperparameter	75

4.4	Academic examples	75
4.4.1	Estimation of an imprecise PSD function	76
4.4.2	Influence of the hyperparameter on the optimised bounds	78
4.5	Optimising the bounds of real data	79
4.6	Conclusions	82
5	Development of a relaxed stationary power spectrum using imprecise probabilities with application to high-rise buildings	84
5.1	Introduction	85
5.2	Stochastic Process Representation and Power Spectrum Estimation	87
5.3	Estimation of a Relaxed Power Spectrum	88
5.4	Numerical Example	89
5.4.1	Single-Degree-of-Freedom System	91
5.4.2	Multiple-Degree-of-Freedom System	92
5.5	Conclusion	96
6	Data-driven reliability assessment of dynamic structures based on power spectrum classification	98
6.1	Introduction	99
6.2	Stochastic processes and power spectrum estimation	103
6.3	Classification of spectral groups within ensembles	104
6.3.1	Problem statement	104
6.3.2	Methodology	104
6.3.3	Optimal number of spectral groups	105
6.3.4	Example	106
6.3.5	Usage of the method	107
6.4	Classification of real data records	108
6.4.1	Source-specific data ensemble	108
6.4.2	Site-specific data ensemble	109
6.5	Numerical examples	112
6.5.1	Linear mass-spring-damper system	114
6.5.2	Non-linear bridge pier model	117
6.6	Conclusions	120
7	Projecting interval uncertainty through the discrete Fourier transform: An application to time signals with poor precision	122
7.1	Introduction	123
7.2	Preliminaries	125
7.2.1	Interval arithmetic	125
7.2.2	Signal with interval uncertainty	126
7.2.3	Interval extensions	127
7.2.4	Repeated variables problem	128
7.2.5	PSD function estimation	128
7.3	The interval discrete Fourier transform algorithm	129
7.3.1	Obtaining the boundary of the united set	129

7.3.2	Determining the two anchoring points	132
7.3.3	Obtaining the interval amplitude spectrum	134
7.3.4	Computational cost	135
7.4	Technical application	137
7.4.1	Modelling of the excitation by sea waves	138
7.4.2	Computing the interval PSD function and the interval system response	139
7.4.3	Comparison with Monte Carlo	141
7.4.4	Sensitivity analysis	143
7.5	Conclusions	144
7.6	Replicability	145
8	Assessing the severity of missing data problems with the interval discrete Fourier transform algorithm	150
8.1	Introduction	151
8.2	Preliminaries	152
8.2.1	Power spectrum estimation	152
8.2.2	Generation of artificial time signals	152
8.3	The interval DFT algorithm	153
8.4	Missing data	154
8.5	Case studies	155
8.5.1	Sensitivity to interval uncertainty	156
8.5.2	Number of missing data	156
8.5.3	Gap size of missing data	158
8.5.4	Distribution of missing data	159
8.6	Conclusions	161
9	Conclusions and Outlook	163
9.1	Conclusions	163
9.2	Outlook	165
10	List of Publications	168

1 | Introduction

In structural engineering it is nowadays of enormous importance to determine the reliability of existing buildings and structures, or those designed for the future, and to confirm their stability. In particular, structures that are exposed to environmental processes are at risk of failure. For example, in almost all regions of the world, wind loads, especially on high-rise buildings, can cause slow damage due to dynamic system behaviour. Heavy winds and storms can cause immediate and severe damage or even cause the collapse of a building. Offshore structures located in the ocean can be subject to a combination of wind and waves. In some other regions of the world, such as China, Japan or Indonesia, just to name a few, buildings can be exposed to short but extreme events such as earthquakes, which can cause serious damage. Many recent events have shown that such environmental processes can have severe consequences.

Although a variety of safety criteria exist for the construction of new structures to counteract such events, it is often not possible or practical to apply them directly due to the complexity of the structures. An important tool in this case are computer simulations, such as finite element model simulations. Due to the ever-increasing computational power, there are more and more accurate models and approaches available to determine the reliability of structures, which is an tremendous advantage in risk assessment. With the help of simulation models, real buildings and structures can be abstracted and tested under certain loads with regard to given safety criteria. Often, simplifications and assumptions of the real case are necessary in a simulation in order not to make the model unnecessarily complex, as very accurate models have a negative impact on the simulation time. Nevertheless, meaningful statements can be made even with a simplified building model. A simulation therefore always represents a compromise between the accuracy of the results and simulation time. However, a significant advantage of simulations is that different scenarios, such as different extreme loads with different intensities, can be analysed, which could possibly occur in reality. Thus, a building can be planned for the worst possible case. Also, computer simulations are indispensable for assessing the reliability of existing structures and for designing new structures in the future.

Since the above-mentioned environmental processes cause a dynamic system behaviour in particular, a characterisation of the dynamic components is necessary. If the peak frequencies of the loads are in the range of the natural frequencies of the structure, this can lead to resonance and cause severe system behaviour. A spectral/frequency analysis of both the system and the load must be performed. An important tool for this purpose is the analysis of specific power spectral density functions, which transforms a signal so that the determining frequencies and their energy distribution are decomposed. Using this method, a system can be assessed with regard to the risk posed by a particular excitation in frequency domain.

Dealing with uncertainties, however, are a major problem in simulations and in predicting the system behaviour. These uncertainties can arise, for instance, from measurement errors, inaccurately calibrated sensors, or a limited amount of samples of the respective load. This can result

in a negative impact on the simulation results. In the worst case, a simulation model influenced by uncertainties can be assessed as safe, although the actual system behaviour would be catastrophic. A reliable prediction regarding critical system behaviour and/or system failure is thus made enormously difficult. An appropriate quantification of these uncertainties is therefore of utmost importance. The challenge and main objective of this work is to quantify and reduce the uncertainties such that reasonable and meaningful simulation results can be obtained. This will also yield in a reduced variability of the simulation results, ensuring that they are more reliable. Due to the problems mentioned above, uncertainties in the estimation of the power spectral density are investigated and quantified in this work.

1.1 Research background

This section outlines the essential subjects of interest in the context of this work, namely uncertainties and power spectral density estimation. A fundamental overview for both is given in this section.

1.1.1 Uncertainties

This section addresses uncertainties in engineering, in particular the sources, their characterisation and categorisation and the propagation of uncertainties in reliability analyses. There is broad agreement in the scientific world that uncertainties are characterised by *aleatory* and *epistemic* uncertainties [1]. Aleatory uncertainties occur due to intrinsic randomness and are irreducible, while epistemic uncertainties occur due to lack of knowledge and are therefore reducible, for example by collecting more data. Thus, a consolidated judgement can be determined, for example, about the distribution of a random variable. If such additional information is available and can be collected/measured, the uncertainties can be characterised as epistemic. If no further information is available such that the uncertainty can be reduced, those are classified as aleatory uncertainties. Often, however, a clear classification is not possible and it is up to the analyst to decide how to treat a particular uncertainty. Such a decision should always be combined with the underlying information and input data. In some cases, both types of uncertainties may exist in the given problem but cannot be strictly separated from each other. This type is called *hybrid* uncertainty, see for instance [2, 3].

Fig. 1.1 provides a first impression of the classification of uncertainties [4]. The upper part of the figure shows the entire spectrum between complete ignorance, when nothing is known about the model or the parameters, to certainty, when the entire model, its parameters and all other necessary information are completely known and the simulation results are absolutely certain. In between, however, is the present state of knowledge, which indicates that all information is only partly certain and partly uncertain. The uncertain part is described in more detail in the lower part of the figure. The present state of knowledge can be improved to a certain extent, namely by reducing the uncertainties, for example by acquiring more information. Once this

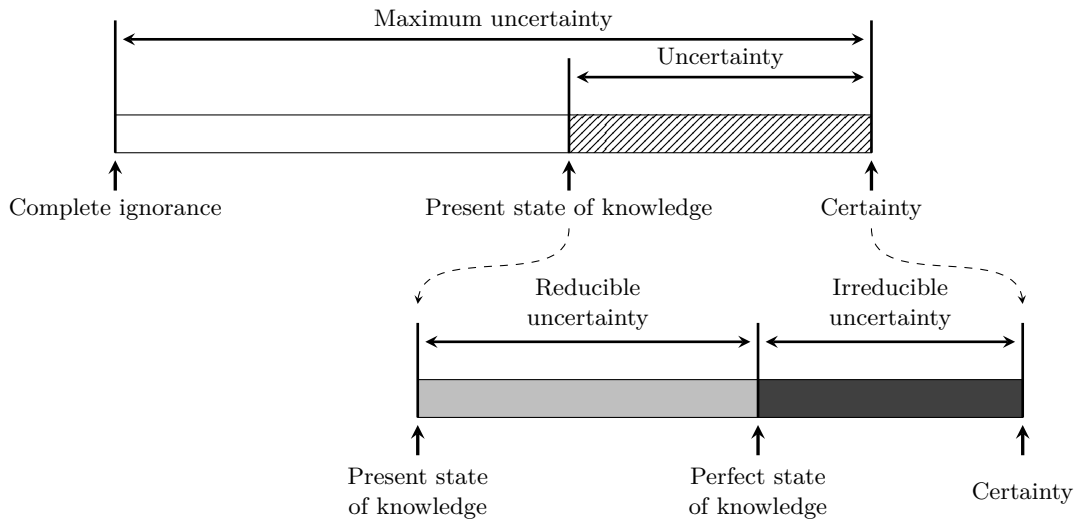


Figure 1.1: Interpretation of uncertainties in an engineering context, adapted from [4].

is accomplished, the perfect state of knowledge is reached. At this point, everything that can be known is known. A level of residual uncertainty, the irreducible uncertainties, remains and stands between the perfect state of knowledge and certainty.

Sources and Characterisation

Uncertainties have a variety of sources and causes that can lead to a wide range of phenomena and have a corresponding negative impact on simulation results [1, 5, 6]. While in [1] uncertainties are defined as parameter uncertainties, model uncertainties and uncertainties in basic variables, in [6] they are classified according to the stage of planning of the development of a model: model development, application or implementation of the model, results of the model. This provides a first understanding of the complexity and the different perspectives on how to classify and characterise uncertainties. A thorough analysis of the uncertainties and their causes can contribute to addressing them appropriately.

Every simulation is preceded by an abstraction of a real case. Accordingly, a structure is represented in simplified form as a model. An exact representation is often not possible and would result in an extremely high computational effort. Thus, simplification always induces uncertainty. Further assumptions and simplifications during the simulation, such as discretisation in space and time, will induce further uncertainty. While a fine discretisation leads to more accurate results but a higher computational effort, a coarser discretisation can deliver fast but less accurate outputs. A further simplification can be made, for example, by assumptions and simplifications in material laws, whether e.g. simple linear models are used or more complex non-linear models, which are closer to reality but result in a higher computational effort. In principle, every simulation model is subject to numerical errors and approximations, which also induce a certain degree of uncertainty. The examples presented correspond to epistemic un-

certainties, as these can be reduced by using more accurate models or a finer discretisation. Nevertheless, it is in the analyst's judgement how these assumptions are chosen and, above all, to the requirements for the corresponding simulation and its evaluation.

Uncertainties often occur in the data used. Measurement errors, i.e. incorrect recording of the real data, can also have a negative influence on the simulation and the incorrect data can lead to simulation results that do not correspond to the actual system behaviour. For example, during data acquisition, external influences can cause noise in the data that interferes with the actual signal to be recorded. A sensor may not be calibrated accurately enough, resulting in measurement uncertainties. Another source of uncertainty in the recorded data may be missing data due to damage or failure of a sensor, possibly due to the event itself that it is supposed to record. These uncertainties can usually be classified as epistemic, as the uncertainties can be reduced by obtaining more data, if possible.

However, it is not always clear whether uncertainties should be classified as epistemic or aleatory. To take up the previous example again: If a signal that is disturbed by measurement errors is uncertain, but this uncertainty cannot be reduced by obtaining further data because they are simply not available, it can be classified as aleatory and not epistemic. Another good example of such a case is given in [1]: A material property, in this case compressive strength of concrete, is classified as either aleatory or epistemic according to the circumstances. If the sample is a piece of concrete that can be taken from an existing building and tested for compressive strength (the method of testing does not matter here), it is considered epistemic. However, if this concrete exists only "virtually" in the design of a future building, and thus no testing can be performed, it is classified as aleatory, since no more accurate conclusions can be made than those known so far. This example shows that the classification is not always unambiguous and also depends on further circumstances and the state of planning.

A rigorous distinction between aleatory and epistemic uncertainties is of utmost importance and should always be conducted in the analysis. Such a distinction can lead to the characterisation of uncertainties and, in the best case, to their reduction through further analysis or the acquisition of additional information. This not only has an impact on the characterisation and quantification of the uncertainties, but also affects the system analysis and its results, as the system analysis with reduced uncertainties is, if applicable, more accurate and thus more reliable. For such a distinction, the following categories are given in [7]:

- Category I: Parameters without any uncertainty
- Category II: Parameters with only epistemic uncertainty
- Category III: Parameters with only aleatory uncertainty
- Category IV: Parameters with both aleatory and epistemic uncertainties

In this context, category I parameters are assumed to be constants. Since the parameters of category II only contain epistemic uncertainties, they can be bounded by known intervals, these

parameters are also called unknown-but-fixed constants. The aleatory uncertainties occurring in category III parameters are described as random variables that can be described by probability characteristics. These can be, for instance, mean, variance or distribution type. Since both uncertainties occur in category IV, they are modelled with imprecise probabilities. The characteristics of these uncertainties can only be roughly identified. Depending on which category an uncertain parameter falls into, it can be treated and quantified accordingly. Although modelling of uncertainties is suggested in this manner in [7], it is up to the analyst how to model a specific uncertainty. In some cases, it might be beneficial to choose another modelling scheme, depending on what type of uncertainty occur and what amount of information is available.

A more detailed review of epistemic uncertainties shows that these can be divided into different types according to [8], depending on how they are described. The three types are illustrated on the basis of the probability density function (PDF) and are as follows:

- Type I: Different shapes of the PDF with identical parameters
- Type II: Different PDFs with different parameters
- Type III: The same type of the PDF with different parameters

Type I describes the case where different PDFs can be used to approximate a data set even though the PDF parameters are identical. In [8] the example is given that, for instance, for a specific set of data both a lognormal and Weibull distribution can be used according to a goodness-of-fit test. In type II, an original PDF can be adjusted via procedures such as Bayesian Updating, which affects the shape of the PDF due to updated parameters. For type III, it can be assumed that a random variable follows a certain distribution. Due to uncertainties in the data (e.g. limited amount of samples), epistemic uncertainties arise in the estimation of the PDF parameters. Therefore, different parameters are possible, which can lead to various forms of the PDF.

The elaborations on the sources and characterisation of uncertainties demonstrate how complex the problem of a reliable quantification of uncertainties may be and that the identification of uncertainties might be challenging. Therefore, quantifying the uncertainties properly is of paramount importance in the context of reliability analysis. A rigorous assessment of uncertainties is important as it is crucial for the system analysis and results.

Modelling of uncertainties

The previously described characterisations of uncertainties can be described mathematically with specific models. There is a wide range of models available that can be selected depending on the characterisation of uncertainties in aleatory or epistemic. In the literature, a distinction is made between three groups of models. These are precise probability models, non-probabilistic models and imprecise probabilistic models. The three types of models are briefly described in the following and examples of such models are given.

The *precise probability models* [9, 10] are the most basic type of models and are based on classical probability theory. They can be used to model and quantify the aforementioned uncertainties for the utilisation in simulations. Such a probability model consists of a sample space Ω , which contains all possible outcomes ω of an experiment. A σ -field \mathcal{F} is a collection of subsets of Ω that is used to represent the results of the experiment. Due to the definition of \mathcal{F} , the empty set is a subset of the σ -field, such that $\emptyset \in \mathcal{F}$. In addition, if an event A is a subset of \mathcal{F} , it follows that $A \in \mathcal{F} \rightarrow A^c \in \mathcal{F}$, where A^c denotes the complementary event of A . The probability measure \mathcal{P} ensures that each event of \mathcal{F} is assigned a probability, such that $P : \mathcal{F} \rightarrow [0, 1]$ and $P(\Omega) = 1$. A precise probability model is thus described by a probability space $(\Omega, \mathcal{F}, \mathcal{P})$.

A cumulative distribution function (CDF) $F(x)$ is able to describe a random variable X uniquely, since F defines a value for each x :

$$F_X(x) = P_X(X \leq x). \quad (1.1)$$

Such a function has the properties that it is increasing and right continuous with $0 \leq F_X(x) \leq 1$. It can be described by $\lim_{x \rightarrow -\infty} F_X(x) = 0$ and $\lim_{x \rightarrow \infty} F_X(x) = 1$. The PDF is described by an integrable function $f_X(x)$ and is the derivative of the distribution function $F_X(x)$, such that

$$F_X(b) - F_X(a) = \int_a^b f_X(x) dx \quad a \leq b. \quad (1.2)$$

As $f_X(x)$ is directly related to $F_X(x)$ it has the property $\int_{-\infty}^{\infty} f_X(x) dx = 1$. The CDF and the PDF are commonly used functions in the classical probability theory, which is well established in the literature, see for instance [9–12].

Non-probabilistic models are a different way of modelling uncertainties, see for instance [13–16]. While in probabilistic models the parameters assume values with specific probabilities, in non-probabilistic models only fixed bounds are assumed and a parameter is characterised by whether it lies within these bounds. In this case the parameters do not follow any distribution. In the context of non-probabilistic models, uncertain parameters can be modelled with intervals. Intervals are a subset of the real numbers \mathbb{R} and are defined as

$$\bar{x} = [\underline{x}, \bar{x}] = \{x \leq x \leq \bar{x}\}. \quad (1.3)$$

Important quantities for describing the interval are the interval midpoint

$$\mu_{\bar{x}} = \frac{\underline{x} + \bar{x}}{2} \quad (1.4)$$

and the interval width

$$\xi = \frac{\bar{x} - \underline{x}}{2}. \quad (1.5)$$

In central notation an interval can be expressed as

$$\bar{x} = x + \xi\Delta, \quad (1.6)$$

where $\Delta = [-1, 1]$ is the unitary interval. Computations with intervals are made possible by interval arithmetic, which is an extension of the regular arithmetic to interval-valued parameters. Some of the basic arithmetic operations, such as addition, subtraction, multiplication and division are given here. The addition of two intervals is given by the addition of their respective lower bounds.

$$\bar{x} + \bar{y} = [\underline{x} + \underline{y}, \bar{x} + \bar{y}] \quad (1.7)$$

The subtraction is described by the subtraction of the respective lower bounds.

$$\bar{x} - \bar{y} = [\underline{x} - \underline{y}, \bar{x} - \bar{y}] \quad (1.8)$$

In multiplication, all combinations of upper and lower bounds of \bar{x} and \bar{y} are calculated. The respective minimum and maximum describe the interval product.

$$\bar{x} \cdot \bar{y} = \left[\min(\underline{x}\underline{y}, \bar{x}\underline{y}, \underline{x}\bar{y}, \bar{x}\bar{y}), \max(\underline{x}\underline{y}, \bar{x}\underline{y}, \underline{x}\bar{y}, \bar{x}\bar{y}) \right] \quad (1.9)$$

The division is described by dividing the interval \bar{x} by the respective upper and lower bounds of \bar{y} , provided 0 is not in y . If 0 is in y , the division is not defined.

$$\bar{x}/\bar{y} = \begin{cases} \bar{x} \cdot \left[\frac{1}{\bar{y}}, \frac{1}{\underline{y}} \right] & \text{if } 0 \notin \bar{y} \\ \text{undefined} & \text{if } 0 \in \bar{y} \end{cases} \quad (1.10)$$

These operations can also be extended to other cases, e.g. square or square root, or to the complex numbers \mathbb{C} . Not only simple numbers can be defined as intervals, but also vectors and matrices. In an interval vector or an interval matrix, each element is defined as an interval. Due to the assumption that these intervals are independent of each other, the interval vector $\bar{x} \in \mathbb{R}^n$ is described by an n -dimensional hypercube. The vertices of this hypercube are each described by the vertices of the intervals in the interval vector.

Another way of describing uncertainties in non-probabilistic models is the fuzzy set theory. This enables the description of the number x via a membership of an interval with a certain degree, the so-called membership level. The fuzzy set theory is therefore to be understood as a natural extension of the interval models and is often used to translate uncertain linguistic variables to a non-deterministic model. A value x is accordingly described by its membership in the fuzzy set \tilde{x} , or in other words, to what degree the value x belongs to the fuzzy set, such that

$$\tilde{x} = \{(x, \mu(x)) | x \in X\}, \quad (1.11)$$

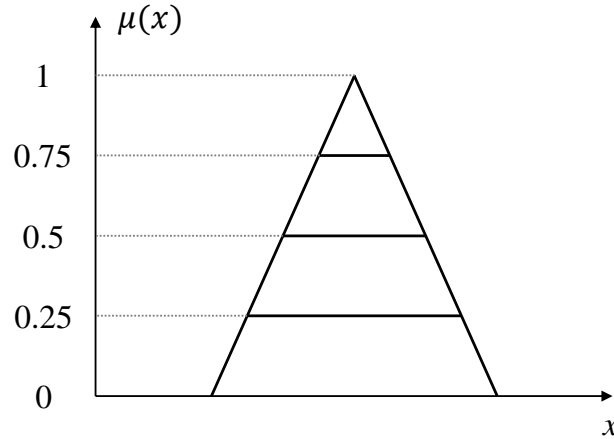


Figure 1.2: Example of a Fuzzy set with several α -level sets.

where $\mu(x) \in [0, 1]$ describes the membership function. For the sake of simplicity, a distinction can be made here between three cases. If $\mu(x) = 0$, then x does not belong to the fuzzy set, if $\mu(x) = 1$, then it belongs to the set completely, which only holds for one element x , and if $0 < \mu(x) < 1$, then it belongs to the set partially. The latter case can be described via the α -level sets, such as

$$x_\alpha = \{x \in X \mid \mu(x) \geq \alpha\}. \quad (1.12)$$

Therefore, the α -level set \tilde{x}_α describes a subset of the fuzzy set \tilde{x} . These sets have the property

$$x_{\alpha_m} \subseteq x_{\alpha_n} \quad \forall \alpha_n, \quad \alpha_m \in (0, 1] \quad \text{with } \alpha_n \leq \alpha_m. \quad (1.13)$$

An illustration of this concept is given in Fig. 1.2.

The possibility theory [14, 15] represents another approach for modelling uncertainties in a non-probabilistic way. It is closely related to the fuzzy set theory and consists of a pair (X_i, r_i) for a variable x_i . Here, X_i describes the set of all possible values that x_i can assume and r_i describes a function that is called the possibility distribution function for x_i . The possibility distribution function can take the values $0 \leq r_i(x_i) \leq 1$ and thus denotes the presence of a specific value $x_i \in X_i$. For a subset $\mathcal{A} \subseteq X_i$, the possibility $\text{Pos}(\mathcal{A})$ and necessity $\text{Nec}(\mathcal{A})$ can be specified, which are defined as follows

$$\text{Pos}(\mathcal{A}) = \sup\{r(x) : x \in \mathcal{A}\} \quad (1.14)$$

$$\text{Nec}(\mathcal{A}) = 1 - \sup\{r(x) : x \in \mathcal{A}^c\} = 1 - \text{Pos}(\mathcal{A}^c), \quad (1.15)$$

where \mathcal{A}^c denotes the complementary of \mathcal{A} .

Imprecise probabilistic models, see for example [16–18], are another way to model uncertainties effectively. Some examples are given here.

Evidence theory, or Dempster-Shafer theory, is to be understood as a generalisation of traditional probability theory and can be applied when the available information has a probabilistic

background but does not follow traditional probability theory, see [17]. It consists of the basic probability assignment m , the Belief function Bel and the Plausibility function Pl . In evidence theory, according to the basic probability assignment, weights are distributed among the sets A_i so that

$$\sum_i m(A_i) = 1. \quad (1.16)$$

By definition it is $m(\emptyset) = 0$. Moreover, $m(A_i)$ only makes statements about set A_i and not about any subsets. Thus, a subset $B \subseteq A_i$ requires its specific assignment $m(B)$. Although the basic idea of assigning a weight to each set is similar to probability theory, these weights cannot be understood as probabilities. According to the basis probability assignment, the so-called belief and plausibility functions can be employed to determine an upper and lower bound that a set of interest lies within. The lower bound is thus determined by the belief function

$$Bel(A) = \sum_{B|B \subseteq A} m(B) \quad (1.17)$$

and the plausibility function determines the upper bound

$$Pl(B) = \sum_{B|B \cap A \neq \emptyset} m(B). \quad (1.18)$$

For more details the reader is referred to [17, 19], for instance.

Interval probabilities [16, 17, 20] are another method of modelling uncertainties that result, for example, from limited data or vague information. The basic idea is to bound the probabilities with intervals in order to obtain a more precise statement about the result. Instead of a probability $P(A)$ from classical probability theory, an event A is bounded by a range of probabilities $[P1(A), P2(A)] \subseteq [0, 1]$. For two interval probabilities $P(A) = [a_1, a_2]$ and $P(B) = [b_1, b_2]$, the following rules apply for conjunction

$$P(A \cap B) = [\max(0, a_1 + b_1 - 1), \min(a_2, b_2)] \quad (1.19)$$

and disjunction

$$P(A \cup B) = [\max(a_1, b_1), \min(1, a_2 + b_2)]. \quad (1.20)$$

A frequently used method for modelling uncertainties in the context of imprecise probabilistic models is that of probability bounds or so-called p-box models. It represents a combination of classical probability theory with interval models. With p-boxes it is possible to bound a probability distribution if only little information is known about the input parameters. In general, only a few information or assumptions are necessary for the construction of a p-box. Depending on the amount and quality of the information, the bounds of the p-box can be very narrow or rather wider. Using a p-box, the CDF $F(x)$ of a random variable X is constrained

by an upper and lower CDF $\bar{F}(x)$ and $\underline{F}(x)$, respectively, such as

$$\underline{F}(x) \leq F(x) \leq \bar{F}(x). \quad (1.21)$$

That means that the precise but unknown CDF $F(x)$ lies within these CDF bounds. Such a p-box is called distribution-free if no further conclusions can be made about the distribution of the individual CDFs in the bounds and only the outer CDFs are known. Every conceivable CDF within these boundaries can be a possible CDF. With distributional p-boxes, on the other hand, the distribution is known but the parameters that determine this distribution are only known within certain bounds. For example, it may be known that a random variable follows a normal distribution, but for the parameters describing this distribution, i.e. mean and standard deviation, only upper and lower bound are known.

Uncertainty propagation

In the following section, a brief overview of different methods for propagating uncertainties in the context of reliability analyses in stochastic dynamics is given. Although propagation of uncertainties is not the main objective of the present work, it is nevertheless important to understand different methods, as they can be applied directly after uncertainty quantification in order to propagate uncertainties efficiently and to obtain the most accurate result possible. The overview in this section of the different methods and techniques is inspired by [3, 21].

Uncertainty propagation refers to the observation of an output variable y when an uncertain input parameter x is applied to a system such that $y = g(x)$, where $g(x)$ is known as the limit state function. It defines the failure domain in case $g(x) < 0$. Particularly in structural dynamics, the calculations of failure probabilities P_f are often in the focus in order to determine the reliability of a structure. In case of precise probabilistic uncertainties, the determination of the failure probability P_f can be expressed in a simple formulation

$$P_f = \int_{\mathcal{X}} I(x)f(x)dx \quad (1.22)$$

with $f(x)$ as PDF and $I(x)$ as indicator function

$$I(x) = \begin{cases} 1, & g(x) \leq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (1.23)$$

Several methods have been developed to calculate the failure probability for precise probabilistic models, see for instance [3, 21]. The first type are the stochastic simulation methods. One of the most well-known stochastic simulation methods is the Monte Carlo (MC) method [22]. MC

is known to be a robust sampling procedure by applying the following expression:

$$P_f^{MC} = \frac{1}{n} \sum_{i=1}^n I(x^{(i)}) \quad (1.24)$$

However, in particular for the efficient determination of small failure probabilities MC has its limitations. In this case, MC is impractical because an prohibitively high number of samples $N_{MC} \approx \frac{1}{P_f}$ may be required to determine the failure probability. For this purpose, advanced sampling techniques such as Subset-Simulation [23] or Directional Importance Sampling [24] were developed to overcome this issue. Nevertheless, the calculation of error probabilities in a detailed simulation model can still be computationally challenging. Asymptotic approximation methods, such as the first-order reliability method (FORM) [25–27] or the second-order reliability method (SORM) [28–30], are the second type of methods for uncertainty propagation and intend to approximate the limit state function by means of the first- and second-order Taylor expansion. The third type, moment methods, aim to adapt a probability distribution to the output variable of a limit state function using the information on the estimated moments as a basis. Integer moment-based methods [31], for instance, are typical examples of this type of model. Probability conservation based methods, such as the probability density evolution method [32, 33], form the fourth type of methods and are used to determine the reliability of a system as well as to model the time variation of the system. They are techniques for modelling time-dependent stochastic physical systems. The last types for precise probabilistic models are the so-called surrogates or metamodels. They are designed to imitate a system with a high computational cost in order to obtain a more cost-effective alternative and to calculate the corresponding probabilities of failure. These models range from neural networks [34, 35], support vector machines [36, 37], polynomial chaos expansions [38–40] and many more.

For the classification of non-probabilistic uncertainties, different models are also listed in [3]. A classical representative for propagating non-probabilistic uncertainties is interval arithmetic [41–44], which extends classical arithmetic to interval-valued parameters. Optimisation methods, such as in [3], are able to find the best possible ranges within the intervals in order to propagate uncertainties efficiently. Perturbation methods [45, 46] draw random samples from the distribution of the uncertain parameter and propagate them to the system, quantifying the uncertainty in the system response. Lastly, advanced sampling techniques for interval analysis are mentioned [47, 48].

Imprecise uncertainties are usually classified as a combination of the above-mentioned models, such as precise probabilistic uncertainties and non-probabilistic uncertainties. This is often due to the fact that aleatory and epistemic uncertainties cannot be separated from each other and must be represented in a single model. The propagation of these type of uncertainties is often carried out by double-loop approaches, see for instance [2, 49], or the non-intrusive stochastic simulation (NISS) [50, 51].

1.1.2 Stochastic processes and power spectral density estimation

Besides uncertainties, the estimation of power spectral density (PSD) functions from stochastic processes is the second topic of interest in this work. Therefore, the estimation of PSD functions is introduced in this section. This also includes an overview of stochastic processes, which occur in nature as environmental loads and form the basis of frequency analysis in this context.

Stochastic processes

A stochastic process is a process that varies over time and consists of stochastic variables or fluctuations at every point in time. This means that each value of said process is influenced or described by random phenomena or random variables and therefore cannot be described precisely or deterministically. If a random experiment is run several times under identical conditions, but is influenced by random variables at each point in time, different outcomes can be observed. Each of these outcomes is described as a realisation of the stochastic process, the collection of all outcomes as an ensemble. Environmental processes, such as earthquakes or wind loads, can be represented as stochastic processes [11].

The autocorrelation function is capable of defining the correlation of two points in time in the stochastic process. If this is calculated from two different stochastic processes, it is called cross-correlation. The autocorrelation function of a stochastic process $X(t)$ can be expressed as the mean value of the product of two values of said process, such that

$$R_X(\tau) = E[X(t)X(t + \tau)], \quad (1.25)$$

where τ describes the time lag. The autocorrelation function for a stationary stochastic process it only depending on the time lag τ and not on the time t itself. This also results in $R_X(\tau)$ being an even function: $R_X(\tau) = E[X(t)X(t + \tau)] = E[X(t)X(t - \tau)] = R_X(-\tau)$, see [12].

According to [11], a stationary stochastic process is described by the fact that “statistical equilibrium” prevails, meaning that “the statistical properties of the process do not change over time, i.e. they are the same at all time points”. A stationary process is strictly stationary if the finite dimensional distributions are time-invariant, described as

$$(X(t_1), X(t_2), \dots, X(t_n)) = (X(t_1 + \tau), X(t_2 + \tau), \dots, X(t_n + \tau)). \quad (1.26)$$

Such a process is said to be weakly stationary if this only holds for first- and second-order probabilities, so that the expected value is constant and the autocorrelation function is independent of the time t and only depends on the time lag τ . A process can additionally be ergodic if the ensemble mean value is identical to mean value of a single realisation of the ensemble [9, 12, 52].

An example of a stationary and non-stationary process is given in Fig. 1.3

To determine the PSD function of a stochastic process, the Fourier transform of the autocorre-

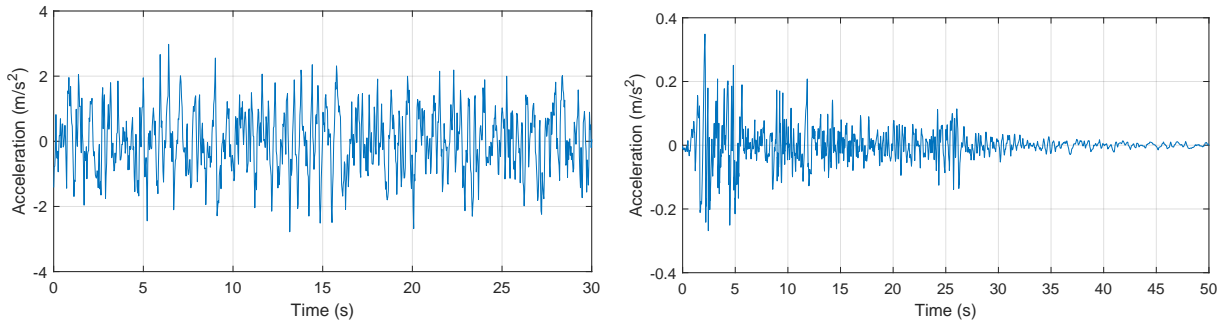


Figure 1.3: Example for a stationary and a non-stationary signal.

lation can be calculated

$$S_X(\omega) = \int_{-\infty}^{\infty} R_X(\tau) e^{-i\omega\tau} d\tau, \quad (1.27)$$

where $S_X(\omega)$ describes the PSD function. The inverse Fourier transform results in

$$R_X(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(\omega) e^{i\omega\tau} d\omega. \quad (1.28)$$

Eqs. 1.27 and 1.28 are the so-called Wiener-Khinchine theorem. The PSD function $S_X(\omega)$ is a real and even function, such that $S_X(\omega) = S_X(-\omega)$ and is described by its non-negativity $S_X(\omega) \geq 0$, see e.g. [52, 53].

Stationary PSD estimation

Since stochastic dynamic problems have been studied very efficiently in recent decades, different models have been developed in the modelling of stochastic processes. One of these models is the PSD function, as mentioned above, which can be utilised in applications such as earthquake, wind and ocean engineering, see for instance [52, 54, 55].

Although the rigorous mathematical relationships between stochastic processes and the PSD function were explained in the previous section, estimation techniques are often required. The relationships presented refer to continuous signals with infinite length, which only exist in theory. In the real case, the signals are discrete-valued and always have finite length, which is the reason for using estimators. Therefore, when estimating PSD functions, some problems occur. In the stationary case, the true PSD function of an ergodic process can only be calculated if the data set has infinite length. Therefore, it is assumed that the mean of multiple PSD estimates from shorter data sets of the same stochastic process generally tends to the true PSD function [12]. Another problem is the often too small sample size of the available data sets. In reality, in most cases, only a single record might be available to estimate the spectrum. In other cases, it can be reasonably argued that a small sample size can be used to estimate the spectrum. In any case, however, with a limited number of samples, it is difficult to average the true PSD function of ergodic processes, especially in the presence of spectral outliers that cannot be clearly identified as outliers.

Furthermore, the quality of a PSD estimate usually suffers from bias and variances [56, 57]. A PSD estimate thus has a high variance in the spectral direction and thus cannot reliably reflect the actual PSD.

In the non-stationary case, the time signal is usually divided into smaller segments and analysed to account for the change over time. The quality of the calculated non-stationary PSD function depends on the chosen length of the segments and leads to a compromise in the time-frequency resolution.

In order to solve the above problems, different estimators for PSD functions have been progressively derived, with different sensitivities to these problems. In general, it is convenient to have several options for estimating the PSD function in order to perform an estimation in the best possible way under the given conditions. Some selected concepts are presented below.

In earthquake engineering, for example, Housner [58] or Kanai [59] were the pioneers in utilisation of the PSD. In the following decades, the utilisation of the PSD in environmental processes was continuously improved, which enables to estimate and utilise non-stationary processes and PSD functions in the investigations of systems. A frequently used estimator of the stationary PSD function from time records is the periodogram, introduced in [60] and discussed in e.g. [12, 52, 61]. To transform a stationary stochastic process from time domain to frequency domain, the discrete Fourier transform (DFT) can be utilised. The periodogram is the squared absolute value of the DFT of the time signal x_n

$$S_X(\omega_k) = \lim_{T \rightarrow \infty} \frac{\Delta t^2}{T} \left| \sum_{n=0}^{N-1} x_n e^{-2\pi i k n / N} \right|^2, \quad (1.29)$$

where N is the total number of data points, T is the total length of the record, n describes the data point index in the record, Δt is the time discretisation and k is the integer frequency for $\omega_k = \frac{2\pi k}{T}$.

If the length of the time record x_n is infinite, i.e. $T \rightarrow \infty$, the periodogram is the true underlying PSD function. As in reality no time records of infinite length exist, the periodogram can be considered as a poor estimator with respect to the spectrum. Alternatively, the PSD function can be approximated by calculating the ensemble mean of all time records, if these are available [62].

$$S_X(\omega_k) = \mathbb{E} \left(\frac{\Delta t^2}{T} \left| \sum_{n=0}^{N-1} x_n e^{-2\pi i k n / N} \right|^2 \right) \quad (1.30)$$

An example of a signal transformed to frequency domain with the periodogram is given in Fig. 1.4.

Since the periodogram often leads to results with high variance, it is considered a poor estimator for PSD functions. Bartlett's method [63, 64] is an alternative method which can reduce this variance by splitting the time signal into several non-overlapping segments and calculating the periodogram for each of these segments. The individual periodograms are then averaged to

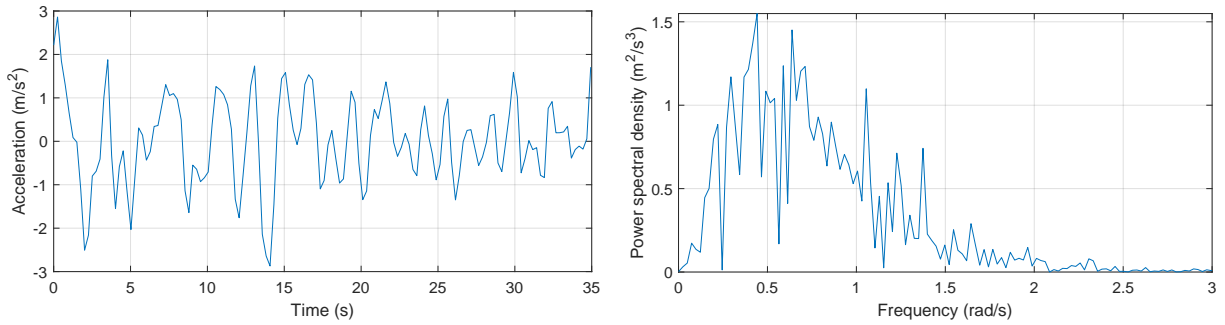


Figure 1.4: Example for a stationary signal transformed to frequency domain with the periodogram.

obtain a smoother estimated PSD function. However, this procedure results in a lower resolution in the frequency domain, depending on the length of the segments.

Another alternative is Welch's method [65], whose idea is similar to Bartlett's method. The aim is to calculate modified periodograms of segments of the time signal with length L instead of the periodogram of the entire time signal $x(t)$, which are then averaged. The segments can overlap in order to maintain continuity. This will result in a smoother spectral estimate. The entire signal is accordingly divided into K segments

$$\begin{aligned}
 x_1(t) &= x(t^*) \\
 x_2(t) &= x(t^* + D) \\
 &\vdots \\
 x_K(t) &= x(t^* + (K - 1)D)
 \end{aligned} \tag{1.31}$$

with $t^* = 0, 1, \dots, L - 1$ and D as a parameter that determines the spacing for the starting points of the segments, respectively. Each of these time segments is multiplied by a window function $W(t)$ before the modified periodograms are formed:

$$P_k(\omega_m) = \frac{1}{L} \left| \sum_{t^*=0}^{L-1} x_k(t^*) W(t^*) e^{-2\pi i m t^* / L} \right|^2 \tag{1.32}$$

with $k = 1, \dots, K$. These resulting modified periodograms are averaged to obtain the estimated PSD function:

$$S_x^W(\omega_k) = \frac{1}{K} \sum_{k=1}^K P_k(\omega_m) \tag{1.33}$$

The selection of the window function can be chosen according to the PSD estimation requirements. Two window functions are suggested in [65], which are $W_1(t) = 1 - t^2$ and $W_2(t) = 1 - |t|$. Both window functions ensure that the values in the middle of the segment are weighted more heavily than the outer values. This results in further smoothing of the data through the estimation process, which can be appreciated from Fig. 1.6, where the same signal as for the periodogram is transformed.

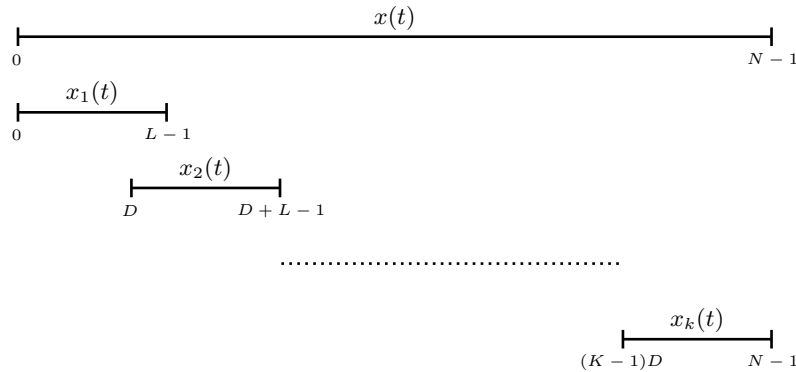


Figure 1.5: Overlapping windows of the time signal, adapted from [65].

Some other methods are the Blackman-Tukey method [66, 67], the Lomb-Scargle periodogram [68, 69], multi-taper methods and approaches such as zero-padding and many more, see for instance [56, 61]. These PSD estimators come with advantages and disadvantages. Each of these methods deserves to be presented, but it is beyond the scope of this work, so for the sake of simplicity, these methods are only referred to.

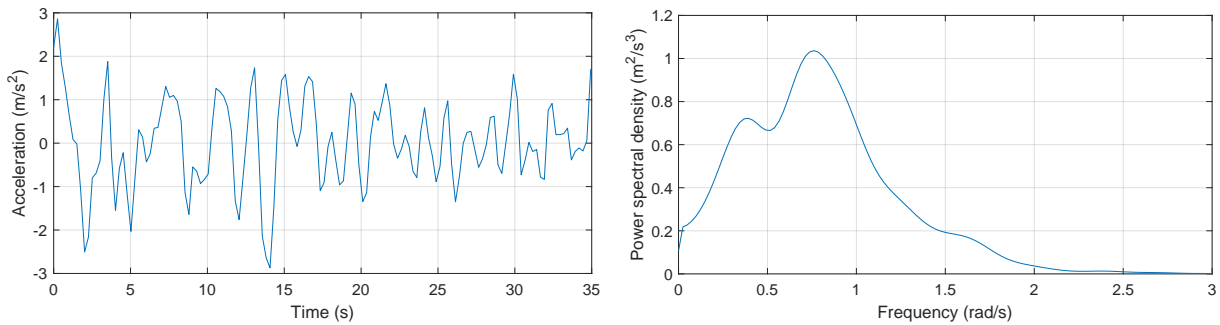


Figure 1.6: Example for a stationary signal transformed to frequency domain with Welch’s method. The signal used is identical to the one in Fig. 1.4.

Non-stationary PSD estimation

Further possibilities in the investigation of environmental processes with the development of evolutionary PSD functions were developed in [70, 71], while in [72–74] important contributions to the estimation such spectra were made. The assumption that a signal has a constant frequency distribution often does not reflect reality. For example, an earthquake is a non-stationary signal with a frequency composition that varies over time. The estimators in the previous section are not able to represent this time-varying character of a signal in frequency-domain. Therefore, methods other than those presented previously must be used to derive a representation of the frequency components in time-frequency domain. Two of them, the short-time Fourier transform and the wavelet transform, are presented here.

An alternative to the stationary estimators is the short-time Fourier transform (STFT) [12, 75].

A window function $w(m)$ is shifted over the entire signal and the DFT is calculated for each segment. The window function is utilised in order to extract the spectral characteristics for the local area, because it is assumed that a signal is stationary if it is sufficiently short. This procedure of shifting the window function and computing the DFT is repeated until the end of the signal is reached. The result is a time-frequency representation of the signal, by which the time-varying frequency component is determined.

$$X_m(\omega_k) = \sum_{m=0}^{K-1} x(n-m)w(m)e^{-2\pi i k m / K} \quad (1.34)$$

By calculating the absolute squared value of the STFT, i.e. $|X_m(\omega_k)|^2$, the representation of the PSD function for the non-stationary case, the so-called spectrogram, is obtained. In this, the time-varying frequencies can be identified directly. However, one disadvantage of this approach is that there is a trade-off between resolution in the frequency domain and resolution in the time domain. A high resolution in both domains cannot be obtained. The resolution is significantly determined by the length of the window function, which means that a short window provides good resolution in the frequency domain and poor resolution in the time domain and conversely. The analyst must therefore find the best compromise of time-frequency resolution for the modelling requirements.

Wavelets are another way of describing the time-varying character of the frequencies of a time signal. The elaborations on wavelets are inspired by the explanation in [76–78]. As for the stationary case, a process representation for a non-stationary process is required. In this context, a new representation for a non-stationary process was introduced in [79, 80]. Instead of a Fourier basis, a time/frequency-localised wavelet basis is utilised and refers to the theory of locally stationary processes as in the STFT before. The process representation is

$$X(t) = \sum_j \sum_k w_{j,k} \psi_{j,k}(t) \xi_{j,k} \quad (1.35)$$

where $\xi_{j,k}$ is as stochastic orthonormal increment sequence, $\psi_{j,k}(t)$ is a family of wavelets, and j and k representing the different scales and translation levels, respectively. The term $|w_{j,k}|^2$ contributes locally to the variance of the process.

Since a wavelet basis is utilised for the aforementioned model, these are discussed further. The generalised harmonic wavelets have a band-limited, box-shaped frequency spectrum [72, 73]. They are defined for scale (m, n) and position (k) and can be represented in frequency domain in the form

$$\Psi_{(m,n),k}^G(\omega) = \begin{cases} \frac{1}{(n-m)\Delta\omega} e^{(-i\omega \frac{kT_0}{n-m})}, & m\Delta\omega \leq \omega < n\Delta\omega \\ 0, & \text{otherwise.} \end{cases} \quad (1.36)$$

The variables m , n and k are considered to be positive integers and $\Delta\omega = \frac{2\pi}{T_0}$ with T_0 as the total time duration of the signal.

In order to get the time-domain representation of the wavelet in Eq. 1.36, the inverse Fourier

transform is applied and is equal to

$$\psi_{(m,n),k}^G(t) = \frac{\exp\left(in\Delta\omega\left(t - \frac{kT_0}{n-m}\right)\right) - \exp\left(im\Delta\omega\left(t - \frac{kT_0}{n-m}\right)\right)}{i(n-m)\Delta\omega\left(t - \frac{kT_0}{n-m}\right)}, \quad (1.37)$$

which is in general complex-valued. For completeness the magnitude

$$|\psi_{(m,n),k}^G(t)| = \frac{\sin\left(\pi(n-m)\left(\frac{t}{T_0} - \frac{k}{n-m}\right)\right)}{\pi(n-m)\left(\frac{t}{T_0} - \frac{k}{n-m}\right)} \quad (1.38)$$

and phase

$$\varphi_{(m,n),k}^G(t) = \pi(n+m)\left(\frac{t}{T_0} - \frac{k}{n-m}\right) \quad (1.39)$$

are given [81]. A collection of harmonic wavelets spans adjacent non-overlapping intervals at different scales along the frequency axis. The continuous generalised harmonic wavelet transform (GHWT) is defined as [73]

$$W_{(m,n),k}^G = \frac{n-m}{kT_0} \int_{-\infty}^{\infty} f(t) \overline{\psi_{(m,n),k}^G(t)} dt, \quad (1.40)$$

where $\overline{\psi_{(m,n),k}^G(t)}$ is the complex conjugation of $\psi_{(m,n),k}^G(t)$. This allows the function $f(t)$ to be projected onto the wavelet basis. When utilising the generalised harmonic wavelets, Eq. 1.35 becomes [82]

$$X(t) = \sum_{(m,n)} \sum_k \left(X_{(m,n),k}(t) \right), \quad (1.41)$$

with

$$X_{(m,n),k}(t) = \sqrt{S_{(m,n),k}(n-m)\Delta\omega} \psi_{(m,n),k}(t) \xi_{(m,n),k}. \quad (1.42)$$

This represents a localised process at scale (m, n) and translation (k) defined in the intervals $[m\Delta\omega, n\Delta\omega]$ and $\left[\frac{kT_0}{n-m}, \frac{(k+1)T_0}{n-m}\right]$. $S_{(m,n),k}$ is the evolutionary PSD function $S_X(\omega, t)$ at scale (m, n) and translation (k) . This equation can be written in the form [82]

$$X_{(m,n),k}(t) = \int_{m\Delta\omega}^{n\Delta\omega} e^{i\omega\left(t - \frac{kT_0}{n-m}\right)} dZ_{(m,n),k}(\omega) \quad (1.43)$$

with the properties

$$E(dZ_{(m,n),k}(\omega)) = 0 \quad (1.44)$$

and

$$E\left(\left|dZ_{(m,n),k}(\omega)\right|^2\right) = S_{(m,n),k}(n-m)\Delta\omega. \quad (1.45)$$

Eventually, the evolutionary PSD function $S_X(\omega, t)$ of a non-stationary process $X(t)$ is estimated

by a wavelet process based estimation approach (see [82, 83])

$$S_X(\omega, t) = S_{(m,n),k}^X = \frac{E\left(\left|W_{(m,n),k}^G[X]\right|^2\right)}{(n-m)\Delta\omega}, \quad (1.46)$$

$$m\Delta\omega \leq \omega < n\Delta\omega, \quad \frac{kT_0}{n-m} \leq t < \frac{(k+1)T_0}{n-m}$$

where $S_{(m,n),k}^X$ represents the evolutionary PSD function of the non-stationary process $X(t)$. It is assumed to have constant value in the intervals $[m\Delta\omega, n\Delta\omega]$ and $\left[\frac{kT_0}{n-m}, \frac{(k+1)T_0}{n-m}\right]$. An example of an estimated non-stationary PSD function is given in Fig. 1.7.

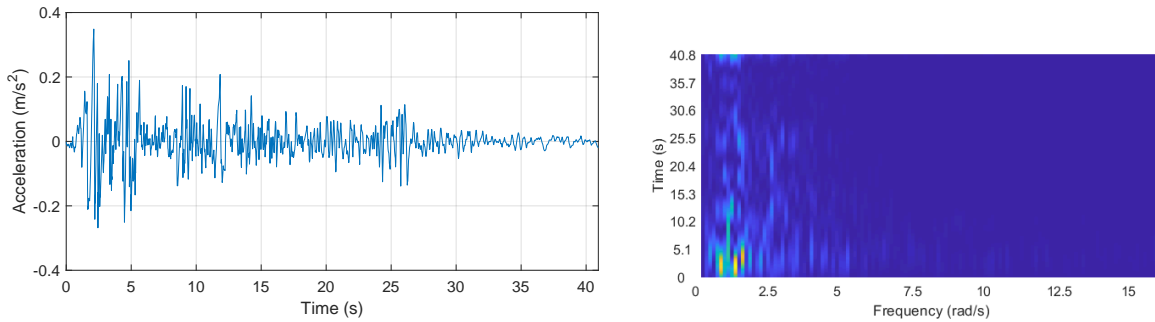


Figure 1.7: Example for a non-stationary signal transformed to the time-frequency domain utilising the wavelet transform.

Analytical PSD functions

The modelling of stochastic processes is an important aspect in simulations. Such a process in time domain can be adapted to specific needs and requirements to a structure or it can be planned against a worst-case scenario, for instance. Therefore, it might be useful to generate artificial time series in order to perform specific simulations. For this generation, an analytical (or estimated) expression of a PSD function is required. A variety of these (semi-)empirical PSD models have been derived for specific applications, such as earthquakes, and serve to describe certain processes. They offer the possibility of adjusting the process via various parameters so that intensity or soil conditions are reflected in said process. Some of the most commonly used PSD models are presented here.

The first PSD function presented here is taken from [84, 85] and represents an artificially generated PSD function of an earthquake.

$$S_{SD}(\omega) = \frac{1}{4}\sigma^2 b^3 \omega^2 e^{-b|\omega|} \quad (1.47)$$

In this equation, σ represents the standard deviation and b is a parameter proportional to the correlation distance of the stochastic process. It is mostly used in modelling purposes and was proposed in line with the so-called Spectral Representation Method (SRM), which will be

explained later in this section. An example of the PSD function with parameters $\sigma = 1$ and $b = 1$ is given in Fig. 1.8. The corresponding autocorrelation to Eq. 1.47 function appears to be

$$R_{SD}(\tau) = \sigma^2 \frac{b^4(b^2 - 3\tau^2)}{(b^2 - \tau^2)^3}. \quad (1.48)$$

The Kanai-Tajimi PSD function [59, 86] is a semi-empirical PSD function for the characterisation of seismic ground motions. It has been observed from multiple real earthquake ground motions that the spectrum has constant characteristics in the bedrock. On the surface, however, it is multiplied by the vibration properties of the ground layer. The model reads as follows

$$S_{KT}(\omega) = S_0 \frac{1 + 4\zeta_g^2 \frac{\omega^2}{\omega_g^2}}{\left(1 - \frac{\omega^2}{\omega_g^2}\right)^2 + 4\zeta_g^2 \frac{\omega^2}{\omega_g^2}}, \quad (1.49)$$

where ω_g is the dominant frequency of the process and the parameter ζ_g describes the sharpness of this peak frequency. The intensity of the excitation is described by S_0 , which is characterised by the seismic waves. This model has a clear physical motivation and is often used to model earthquake ground motions subject to engineering structures due to its simplicity, see for example [53].

A modified version, the Clough-Penzien spectrum [87], which extends the Kanai-Tajimi PSD by a linear filter is given by

$$S_{CP}(\omega) = S_0 \cdot \frac{\omega_g^4 + 4\zeta_g^2 \omega_g^2 \omega^2}{(\omega_g^2 - \omega^2)^2 + 4\zeta_g^2 \omega_g^2 \omega^2} \cdot \frac{\omega^4}{(\omega_f^2 - \omega^2)^2 + 4\zeta_f^2 \omega_f^2 \omega^2}. \quad (1.50)$$

In this equation, ω_g describes the dominant frequency with shape parameter ζ_g as in the Kanai-Tajimi spectrum. The parameters ω_f and ζ_f are the natural frequency and damping of the linear filter, respectively.

In Fig. 1.8 an overview of the presented PSD functions is given.

Waves in the ocean often affect offshore structures through their dynamic forces. In [53] it is shown that the spectral density of the wave force is linearly related to the spectral density of the sea surface elevation. It is therefore sufficient to model the sea surface elevation. A well-known model is the Pierson-Moskowitz spectrum [88]

$$S_{PM}(\omega) = \frac{\alpha g^2}{\omega^5} e^{-\beta \left(\frac{g u_0}{\omega}\right)^4}. \quad (1.51)$$

Here, α and β describe empirical parameters, g is the gravity acceleration and the parameter u_0 the average wind speed.

As part of the Joint North Sea Wave Observation Project (JONSWAP) [89], a PSD function was derived which is an extension of the Pierson-Moskowitz PSD function and extended by a peak enhancement factor. To obtain a realistic representation of this stochastic process, a large

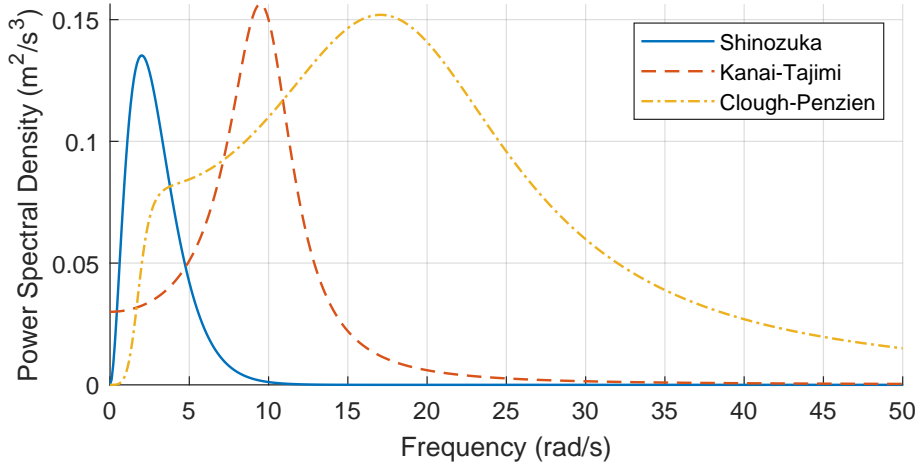


Figure 1.8: Examples for earthquake power spectra.

amount of real data was analysed for the derivation of this model, which reads as follows

$$S_J(\omega) = \frac{\alpha g^2}{\omega^5} \exp\left(-\frac{5}{4} \left(\frac{\omega_p}{\omega}\right)^2\right) \gamma^r \quad (1.52)$$

with the peak enhancement factor

$$r = \exp\left(\frac{-(\omega - \omega_p)^2}{2\sigma^2\omega_p^2}\right). \quad (1.53)$$

In these equations, α describes a spectral energy parameter, g is the gravity acceleration, ω_p describes the peak frequency, γ^r is the peak enhancement factor and σ the spectral width parameter.

A more realistic representation of a stochastic process in the frequency domain can be determined with non-stationary or evolutionary PSD functions. Non-stationary PSD functions are characterised by the fact that they not only represent the distribution of the spectral density over the individual frequencies, but also represent their temporal change. It is therefore a function dependent on frequency and time. Non-stationary PSD functions can be distinguished between separable and non-separable PSD functions. While separable PSD functions consist of two separable parts for frequency and time, these are nested in a non-separable function.

A separable PSD function can be generated by multiplying one of the above mentioned stationary PSD functions by means of a so-called envelope or time modulation function, such as $S_N(\omega, t) = S(\omega)g(t)$ or $S_N(\omega, t) = S(\omega)|g(t)|^2$, with $S(\omega)$ as a stationary PSD function and $g(t)$ as an envelope function. A stationary PSD function can thus be synthesised into an evolutionary PSD function, which leads to a two-dimensional function. Some selected examples ([81, 90]) of such envelope functions are listed below and illustrated in Fig. 1.9.

The envelope function $g_1(t)$ depends on the parameters k , a and b . In Fig. 1.9 these have been chosen to be $k = 4$, $a = 0.25$ and $b = 0.5$. Of course, other parameters are conceivable,

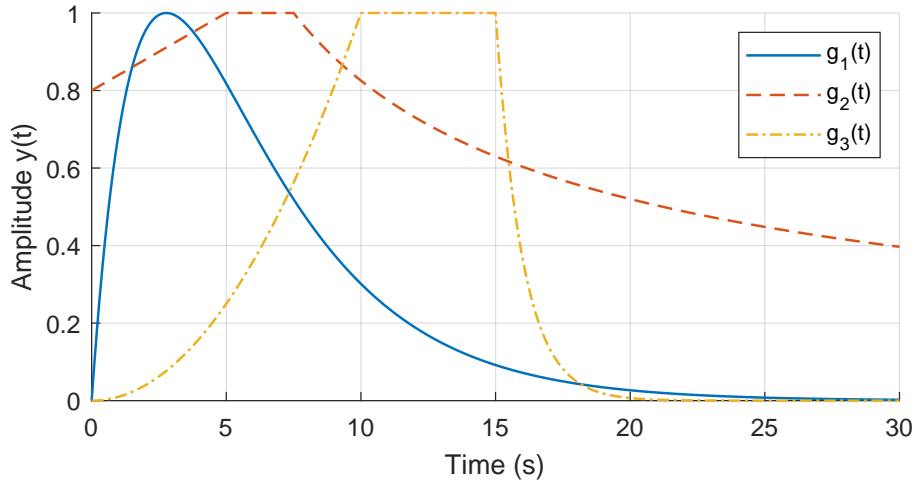


Figure 1.9: Examples of envelope functions for a generating a time-modulated PSD function.

depending on the requirements on the stochastic process.

$$g_1(t) = k \left(e^{-at} - e^{-bt} \right) \quad (1.54)$$

The envelope function $g_2(t)$ consists of different cases that are applied depending on the time. Again, the choice of timing is variable and can be adapted to the process. For the function shown in Fig. 1.9, the parameters $t_a = 5$ and $t_b = 7.5$ have been chosen.

$$g_2(t) = \begin{cases} 0.8 + 0.2\frac{t}{2}, & t < t_a \\ 1, & t_a \leq t \leq t_b \\ \left(\frac{t_b}{t}\right)^{2/3}, & t > t_b \end{cases} \quad (1.55)$$

The third presented envelope function $g_3(t)$ also consists of different cases that can be adapted to the stochastic process depending on the requirements. For the envelope function shown in Fig. 1.9, $\alpha = 1$, $t_a = 10$ and $t_b = 15$ have been chosen.

$$g_3(t) = \begin{cases} \left(\frac{t}{t_a}\right)^2, & t < t_a \\ 1, & t_a \leq t \leq t_b \\ e^{-\alpha(t-t_b)}, & t > t_b \end{cases} \quad (1.56)$$

However, if frequency and time parameters are inseparable in one expression, the term non-separable PSD function is used. An example of such a function is

$$S_E(\omega, t) = S_0 \left(\frac{\omega}{5\pi} \right)^2 e^{-ct} t^2 e^{-\left(\frac{\omega}{5\pi}\right)^2 t} \quad (1.57)$$

with $c = 0.15$ for instance, see e.g. [82]. Such a function is depicted in Fig. 1.10.

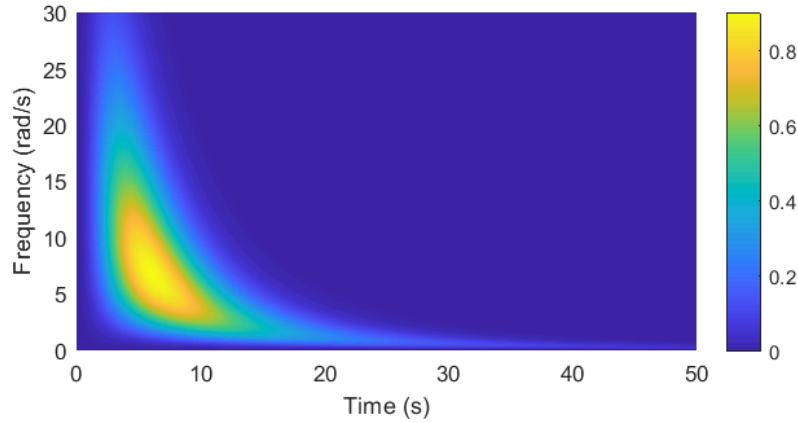


Figure 1.10: Example for the non-stationary non-separable PSD Function $S_E(\omega, t)$ (Eq. 1.57).

Generation of stochastic processes

While in the previous section the focus was on analytical or (semi-)empirical expressions of PSD functions for certain stochastic processes, the generation of time signals with the characteristics of such a PSD function is described. To obtain a stochastic process in the time domain, the analytical PSD functions mentioned above can be used. This can be useful to determine the reliability of a structure in the context of e.g. Monte Carlo (MC) simulations. A major advantage of this approach is that the processes can be adapted to specific requirements in order to evaluate different possible scenarios. For the generation of a stochastic process, random variables are often used, which leads to a certain variability in the processes and thus also in the simulation results. It is therefore necessary either to generate a sufficiently large number of MC samples to minimise the influence of the random variables, or to keep the number of random variables small, if possible. Some selected methods for generating stochastic processes are presented here. A widely used method is the Spectral Representation Method (SRM) [84]. It provides a suitable approach for generating compatible stochastic processes X_t^{SRM} derived from an underlying PSD function S_X . Due to the form of the method, the generated time signals have the characteristics and properties of the underlying PSD function. The method reads as follows

$$X_t^{SRM} = \sqrt{2} \sum_{n=0}^{N-1} A_n \cos(\omega_n t + \varphi_n), \quad (1.58)$$

with amplitude

$$A_n = \sqrt{2S_X(\omega_n)\Delta\omega}, \quad (1.59)$$

where $n = 0, 1, 2, \dots, N-1$. N is the number of frequency points in S_X , $\omega_n = n\Delta\omega$ is the frequency vector with $\Delta\omega$ as frequency discretisation and t as time coordinate. The random variables φ_n are uniformly distributed random phase angles in the range $[0, 2\pi]$. The strength of the method is that a large number of time signals representing the PSD in the time domain can be generated with little effort. The generated time signals also exhibit a low autocorrelation. A

disadvantage, however, is that a large number of random variables are required, since for each frequency component an additional random variable is needed.

To reduce the number of random variables, a method based on SRM was proposed in [91], the so-called stochastic harmonic functions (SHF). The general idea is to subdivide the frequencies into smaller intervals $[\omega_{i-1}, \omega_i]$ for $i = 1, 2, \dots, N$ from which a random frequency is sampled and the corresponding PSD value is plugged into the expression of the amplitude. A stochastic process can subsequently be generated using Eq. 1.58 by substituting the amplitude according to SHF. Two types, the Stochastic Harmonic Functions of the First Kind (SHF-I) and the Stochastic Harmonic Functions of the Second Kind (SHF-II) are proposed.

For SHF-I the amplitude is

$$A^{SHF-I}(\omega_i) = \sqrt{\frac{1}{\pi} \int_{\omega_{i-1}}^{\omega_i} G(\omega) d\omega}, \quad (1.60)$$

where $f_i(\omega)$ is the standardised PSD over the interval $[\omega_{i-1}, \omega_i]$, such that

$$f_i(\omega) = \frac{G(\omega)}{\int_{\omega_{i-1}}^{\omega_i} G(\omega)} \quad \text{for } \omega \in [\omega_{i-1}, \omega_i], \quad (1.61)$$

with $G(\omega)$ as target PSD function.

For SHF-II the amplitude appears to be

$$A^{SHF-II}(\omega_i) = \sqrt{\frac{1}{\pi} G(\tilde{\omega}_i) \Delta\omega_i}, \quad (1.62)$$

where $\Delta\omega_i = \omega_i - \omega_{i-1}$ is the length of the subdivided interval. In this case, the ω_i 's are uniformly distributed over the said interval, such that

$$f_i(\omega) = \frac{1}{\omega_i - \omega_{i-1}} = \frac{1}{\Delta\omega_i}. \quad (1.63)$$

In [91] it is also proposed, that the intervals may overlap. Furthermore, it was shown that the proposed methods generating stochastic processes with either SRM or SHF have similar properties, such as low autocorrelation. However, by utilising SHF it can be achieved to reduce the number of random variables. SHF was further extended to generate non-stationary time signals [92].

Since a high number of random variables for a simulation of a non-linear structure can result in a high computational effort, a further extension of SHF is proposed in [93], which introduces dependencies between the frequencies in order to reduce the number of random variables. For the dependence between frequencies, only one uniformly distributed random variable λ in the

range $[0, 1]$ is introduced. The random frequencies ω_j can be determined via this random variable

$$\omega_j = f_j(\lambda), \quad (1.64)$$

where $f_j(0) = \omega_{j-1}$ and $f_j(1) = \omega_j$. A convenient way to determine the frequencies according to λ is

$$f_j(\lambda) = \omega_{j-1} + (\omega_j - \omega_{j-1}) \lambda. \quad (1.65)$$

Therefore, the amplitudes are

$$A(f_j(\lambda)) = \sqrt{4S(f_j(\lambda))(\omega_j - \omega_{j-1})}. \quad (1.66)$$

The fundamental equation of the SRM (1.58) can thus be modified to generate artificial stochastic processes

$$X_t^{SHF-R} = \sum_{j=1}^N A(f_j(\lambda)) \cos(f_j(\omega)t + \varphi_n). \quad (1.67)$$

It has been shown, that realisations of non-stationary stochastic processes can be generated by utilising a generalisation of SRM (Eq. 1.58) with

$$X(t) = \sum_{j=0}^{N-1} \sqrt{4S_X(\omega_j, t)\Delta\omega} \cos(\omega_j t + \varphi_j). \quad (1.68)$$

Note, that this equation is similar to Eq. 1.58 except that the stationary PSD function $S_X(\omega)$ is replaced by a compatible non-stationary equivalent $S_X(\omega, t)$, see [94]. Therefore, the generated stochastic process has a non-stationary character, which is passed by the evolutionary PSD $S_X(\omega_j, t)$.

Many derived methods are based on SRM, which has been studied extensively in the literature, see for instance [95–97]. Nevertheless, the needs and requirements for certain simulations of stochastic processes are constantly changing or other developed methods provide new ways to efficiently generate stochastic processes.

For example, in [98] a method is presented to generate multivariate stationary stochastic processes from the Fourier-Stieltjes integral of the spectral representation, leading to ergodic stochastic processes. In [99], an enhanced SRM was developed to generate ergodic multivariate stochastic processes via shifting a frequency increment. In [100], 500 real earthquake ground motions were analysed and divided into groups, of which the standards time-domain enhancement attenuation process is described as an edge envelope curve. A method for non-stationary zero-mean processes whose target acceleration is one of its own samples is presented in [101]. The variety of methods for generating stationary and non-stationary stochastic processes, especially the recently published results, shows that it is a thriving research topic and that there is a constant need for improvement as the requirements are constantly changing as well. It should be noted that this is not an exhaustive list and only serves to give an impression of the extent of the

research area.

To conclude this section, some approaches are presented that have already dealt with the problem of uncertainties in the data or the estimation of PSD functions. In these works, the uncertainties arise from different sources, such as incorrect recording, suffering from noise or missing data. The problem of missing data is in particular an important area of research, as it has a big impact on estimating PSD functions. The DFT is very sensitive to changes in the time signal, which have a strong impact on the PSD functions. An incorrectly reconstructed signal can therefore lead to highly erroneous results, such as the peak frequency being determined incorrectly or with an incorrect magnitude. Especially missing data reconstruction plays an important role. In [102], individual segments of signals separated by long gaps were investigated. These gaps were filled using a maximum entropy method, which was followed by a prediction based on the observed data. In [103], an algorithm is proposed that examines two cases: signals with missing data and signals with completely randomly sampled data points. That means, that the signal is not uniformly sampled. The algorithm performs an iterative deconvolution of the time window in the frequency domain, removing artefacts due to the finite nature of signals. Another algorithm that estimates PSD functions in missing data problems with high accuracy is proposed in [104], which is also applicable to practical problems. For the reconstruction of missing data, artificial neural networks (ANN) can be used [76]. An ANN is able to detect the stochastic pattern in the available data and thus reconstruct the missing data. From the reconstructed signal, the PSD function can be estimated using the established methods. Other methods ([77, 105]) fill the gaps in signals with missing data with random variables to propagate this signal, then consisting of fixed values and random variables through the DFT in a closed-form expression. The result is a PSD function described by a probability density function at each frequency. For spectra that can only be described with a few frequency components, compressive sensing approaches can be used to fill gaps in the signals [78, 106]. For this purpose, an appropriate basis, such as the Fourier basis, is chosen in the time domain and an L1 minimisation is performed to obtain the sparsest representation of the process. Then the signal can be reconstructed or the PSD function can be obtained directly. It is shown that signals with up to 65% missing data can be efficiently reconstructed. In [107], an approach is presented that uses genetic algorithm to reconstruct missing data and noisy signals.

Other approaches quantify uncertainties in the data differently from missing data problems. Metrics that can be used to assess uncertainty in a PSD are presented in [108]. PSD functions that are consistent with a given range of values for the estimated statistic represent an uncertainty set over the underlying PSD function. With this uncertainty set, upper and lower bounds are derived to quantify the underlying PSD function with confidence. In [109], a large set of accelerograms is used to determine interval parameters for a semi-empirical PSD function. Thus, different representations of the PSD functions are resulting, depending on the bounds used for the derived interval parameters. To evaluate the structural safety, the PSD function of the ground motion acceleration was embedded with different interval parameters in classical

first-passage problems in order to perform a reliability analysis to derive the bounds for the failure probability.

1.2 Aims and objectives

Since uncertainties occur in the time signal, e.g. due to measurement errors, poor quality due to inaccurate calibration of the sensor, discretisation of the signal, etc., and further uncertainties are induced by the estimation process of the PSD - as mentioned, the periodogram is a poor estimator, Welch's method averages the data - the true underlying PSD cannot be determined unambiguously and with absolute certainty. Therefore, approaches are proposed to capture the uncertainties simultaneously and in one step directly in the frequency domain. The main goal of this work is to derive non-deterministic PSD representations that are able to quantify the uncertainties mentioned. These models are then able to be used in the context of reliability analyses and to assess the reliability of a structure that takes these uncertainties into account, for instance.

To derive these models, a set of available data from environmental processes such as earthquakes or wind loads is used. The precondition for the PSD models to be derived is that these data have similar spectral characteristics in the frequency domain, such as peak frequency or general distribution of frequencies and amplitudes. Once this spectral similarity is verified, the PSD models can be derived in two different ways, which should be chosen according to how much data is available. If multiple data is available, reliable statistics such as mean and standard deviation can be obtained, so a probabilistic model is derived. If data is limited, it is not possible to derive reliable statistics, so in this case an interval-based approach is chosen that determines an upper and lower bound on the data set. In some cases, when only a single but uncertain signal may be available, for example when only reliable bounds of the signal are known, an approach is used which propagates this interval signal through the DFT to generate an interval-based PSD model with an upper and lower bound.

The main objective of quantifying and modelling uncertainties in the PSD function can therefore be refined into the following specific objectives:

1. Development of a probabilistic PSD model based on the assumption that multiple similar data records are available. In such a case, statistics can be reliably derived and therefore confident statements about the distribution of the data can be made. Therefore, a probabilistic approach is chosen.
2. Development of an interval-based PSD model based on the assumption that only a few similar data records are available. In such a case, no reliable statements can be made about the distribution of the data. In this case, an interval-based approach is chosen.
3. The development of an approach to determine the spectral similarity of a data set, which can be used to choose appropriate data for the derivation of the two non-deterministic

PSD models mentioned above. The approach can also be used to assess a large data set for similarity, which can be used in a reliability analysis to reduce variability in the data and simulation results.

4. An extension of the DFT, which allows the calculation of exact bounds for the Fourier amplitude and the PSD function of interval-valued signals. The repeated variables problem is taken into account to counteract artificially inflated bounds.

1.3 Original contributions

The contributions of this work in the context of uncertainty quantification of PSD functions are mainly in the development of different non-deterministic load models of PSD functions. These models are able to quantify uncertainties in the data in the time domain and eventually in the frequency domain and aim to achieve robust simulation results in the context of reliability analyses. The simulation results are presented in a form identical to the non-deterministic input data. The main developments are as follows:

First, a model is proposed that represents the PSD function in the frequency domain as a probabilistic model. A variety of data with similar characteristics in the frequency domain, such as peak frequency and shape, are used to derive reliable statistics. From these statistics, a probability density function is derived for each individual frequency, modelling the spectral density at that frequency as a random variable. The resulting *relaxed PSD* thus forms a non-stationary stochastic process itself. In this work, a truncated normal distribution is proposed for the probability density function of the individual frequencies, which has been proven to fit with various statistical tests. For other data sets, other distributions are also conceivable. In any case, it must be ensured that the function cannot assume negative values, since negative values are physically impossible within the framework of PSD functions. Due to the probabilistic character of the PSD function, individual PSD functions can be sampled and stochastic processes can be generated from them, which enables the use in the context of Monte Carlo simulations or similar schemes to analyse the system behaviour of a structure. In a further work, different methods for generating stochastic processes from the relaxed PSD were investigated. The aim of this work is to find alternative methods that generate a stochastic process with as few random variables as possible. In particular, the variances of the transformed signals in the frequency domain have been investigated under different numbers of samples. In addition, the autocorrelation function of the generated stochastic processes was analysed.

A second development in this thesis also deals with the estimation of an alternative non-deterministic PSD model. In contrast to the previous model, the assumption is made that additional uncertainties are induced due to limited data. Since no reliable statements about the distribution of the data can be made due to the limited data set, this non-deterministic model is defined by an upper and lower bound that encloses the data. A Radial Basis Function (RBF) network is used to approximate a basic power spectrum derived from the data set, which roughly

pre-defines the shape of the PSD and reflects its physics. The resulting basis functions from the RBF network are manipulated via the associated weights so that the bounds are optimised taking into account dependencies between the frequencies, resulting in the *imprecise PSD*. The advantage of this approach is that reliable data-enclosing bounds are derived and the analyst has control over the number of basis functions to be used, which is an advantage for subsequent interval propagation. This approach is applicable to both artificially generated data and real data sets and thus has a practical use. In an earlier development, confidence intervals based on the data were used to derive the bounds, which were calculated at different confidence levels. From the different confidence levels, fuzzy memberships can be subjectively assigned to generate a fuzzy number for each individual frequency.

The third development is mainly concerned with the classification of data in the frequency domain, especially if they have a high spectral variance. For simulations, it is important to keep the variability in the data and in the results low, as this high variance data sets cannot provide reliable results. If the variance in the data is too high, this also has a negative effect on the results and they become less reliable. Therefore, an approach is proposed that determines the *spectral similarity* of the data entirely in the frequency domain and divides them into groups accordingly. To get a first measure of similarity, the Bhattacharyya distance is used. The distance values are divided into groups according to the k -means algorithm, from which new load models can be generated. For reliability analyses, all classified load models of a data set can be averaged and a reliability analysis based on the probability of occurrence of each load model can be obtained. The result of this approach is more accurate simulation results and the reduction of variability in the data. Furthermore, this approach can be used to determine the spectral similarity to derive a relaxed PSD or imprecise PSD, as their precondition is that only similar data can be used.

For the last development, unlike before, no data set is used, but only a single signal is considered. This signal is assumed to suffer from poor precision and may only be known in reliable bounds. It is therefore represented as an interval signal. Since the DFT is required for the transformation of a signal into the frequency domain, but is not capable of transforming interval-valued signals, an extension is proposed. This novel *interval DFT algorithm* is able to transform such interval signals into the frequency domain. This is done by taking into account the repeated variables problem, which in interval arithmetic leads to the fact that calculated bounds are artificially inflated when a variable occurs several times. The interval DFT algorithm is able to capture this problem and thus provide exact bounds for each frequency. The interval signal can be transformed into the frequency domain and the exact bounds for the Fourier amplitude and PSD function can be obtained. Further work evaluates the missing data problem assuming that missing data can be represented as intervals with some confidence. This signal is then transformed into the frequency domain and reliable bounds can be derived even under missing data. The work is mainly concerned with different cases such as the length of the gap and the size of the interval uncertainty. It is shown that reliable results can be obtained even with a

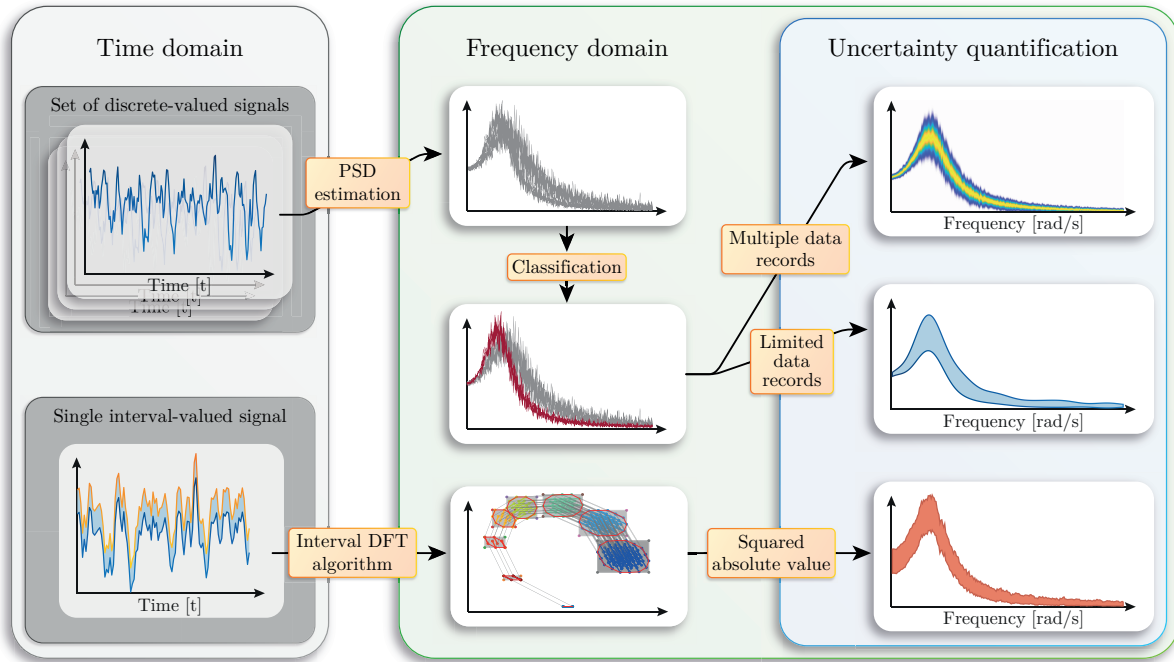


Figure 1.11: Concept of the proposed methods

high number of missing data under an appropriate reconstruction of the data.

The developments mentioned above result in a framework for quantifying uncertainties in the estimation of PSD functions. Each development contributes successfully to a specific case. All models have proven useful in practical simulations and have contributed to an improvement in quantifying uncertainties effectively. An overview of the framework in this work, i.e. the individual contributions and its interaction, is shown in Fig. 1.11. The starting point in each case is the time domain, in which a set of discrete-valued time signals is given. These are individually transformed into the frequency domain and the so-called ensemble of PSD functions can be obtained. In the case that these have a high spectral variance, the developed classification approach is applied to identify a group of PSD functions that have a spectral similarity. This ensemble of similar PSD functions can then, depending on how much data is available, be transformed into the relaxed PSD if multiple records are available, or into the imprecise PSD if only limited records are available. If only a single uncertain signal is available, i.e. an interval-valued signal, it is first transformed into the frequency domain using the interval DFT by propagating the intervals as complex numbers in the Fourier domain. Then, an endpoint analysis is carried out until the absolute value is formed, which is done individually for each frequency and results in an interval PSD model. The derived models, shown on the right hand side in Fig. 1.11, can then be used for simulation, for example in reliability analysis, via various methods of uncertainty propagation. The basic idea is that the respective model is propagated through the system in the way it exists, i.e. probabilistically or interval-valued, and the system response can also be obtained probabilistically or interval-valued, respectively. This scheme is

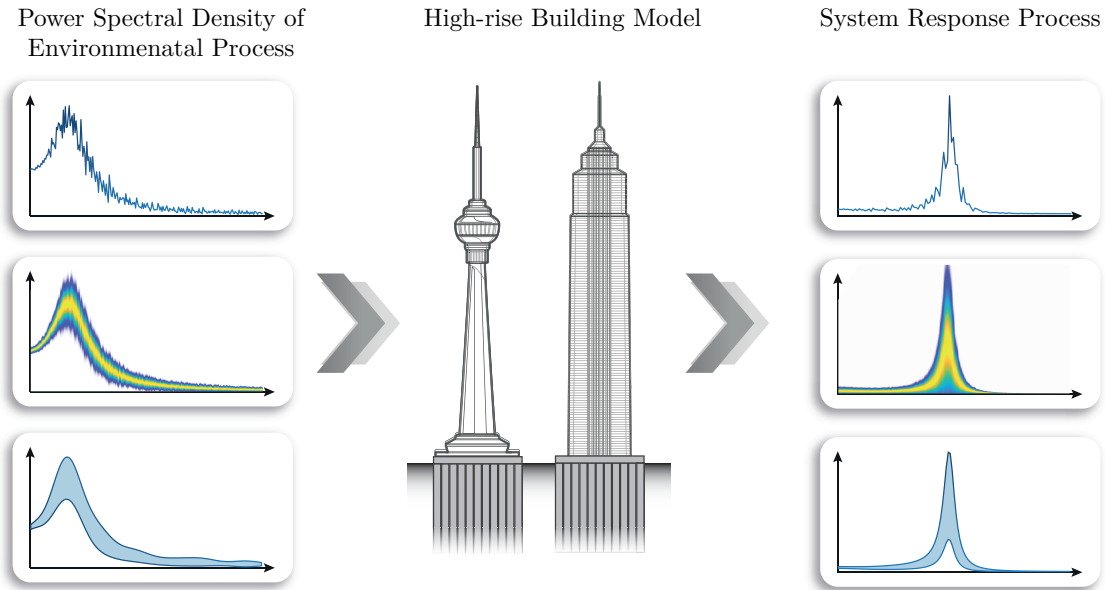


Figure 1.12: Uncertainty propagation for identifying the system behaviour

shown in Fig. 1.12. The standard approach is represented as a discrete-valued PSD function which will lead to a discrete-valued response after simulation. The probabilistic model of the relaxed PSD yields a probabilistic response after the propagation, while the imprecise PSD or interval PSD from the interval DFT algorithm yield an interval-valued system response.

1.4 Structure of the thesis

This dissertation is comprised of four journal articles and three conference papers. The journal articles provide the main developments of this work, while the conference papers can be seen as further developments or elaborations of the findings in the journal articles. The structure is based on the specific research results in the framework of uncertainty quantification in PSD estimation, i.e. the relaxed PSD estimation, the imprecise PSD estimation, the definition of the spectral similarity and the interval DFT algorithm.

The first journal article is concerned with the development of the relaxed PSD derived from multiple data records, which is a probabilistic representation of an ensemble of PSD functions, see Chapter 2. The results of this journal article are examined with regard to the generation of stochastic processes, see Chapter 3.

The second journal article addresses the same problem, but assuming that only limited data records are available. Therefore, the imprecise PSD is derived by using an RBF network to obtain an upper and lower bound for the data set. This is explained in more detail in Chapter 4. An alternative and earlier development of the imprecise PSD based on the determination of confidence intervals in combination with Fuzzy sets is discussed in Chapter 5.

Since the development of a load model, such as the relaxed PSD and the imprecise PSD, requires the definition of the similarity of the data, the determination of the spectral similarity of these data is explained in Chapter 6.

The last article considers the uncertainty in the time signal itself, leading to the development of the interval DFT algorithm and an additional bounded PSD function, which is discussed in Chapter 7. The results presented in Chapter 8 are a further work focusing in particular on missing data problems using the interval DFT algorithm.

2 | Relaxed power spectrum estimation from multiple data records utilising subjective probabilities

The first development in this thesis introduces a probabilistic representation of an ensemble of PSD functions with similar properties to quantify the epistemic uncertainties. These uncertainties arise from various problems in data acquisition, such as measurement errors, sensor failures, inaccurately calibrated sensors or incorrectly recorded extreme values. In addition, the process of estimating PSD functions induces further uncertainties, as these are often based on the assumption that a continuous infinite signal is present. Since this is never the case in reality, and due to the aforementioned problems in data acquisition, a certain degree of uncertainty in the recorded signal can always be assumed in the real case. As these uncertainties are inherent in the signal in the time domain, they are also propagated into the frequency domain when estimating the PSD function. The goal is therefore to find a way to quantify these uncertainties in the frequency domain and thus provide a basis for making simulation analyses more reliable. Reliable statistics, such as mean and standard deviation, can be derived mainly in the presence of multiple data. Hence, a probabilistic model based on probability density functions can be derived. The basic principle is to transform a set of data in the time domain, e.g. of earthquakes or wind loads, into the frequency domain. If these data have similar characteristics, such as the peak frequency or the shape of the PSD functions, mean and standard deviation can be determined for each frequency component, from which a probability density function is then generated. Since negative values are not physically possible in the context of PSD functions, a truncated normal distribution is suggested. However, due to the shape of the data, other distributions are also conceivable. Furthermore, the truncation bounds can be fitted individually to the given data. Statistical tests were performed to validate the choice of a truncated normal distribution. For a convergence study, individual PSD functions are sampled from the generated relaxed PSD, a time signal is generated from it and then transformed back into the frequency domain. Thus, the validity of such a probabilistic model is checked by means of statistical tests and a convergence study. Furthermore, the influence of such a model on the generated time signals and their autocorrelation function is examined, as these could be used in Monte Carlo simulations. These results also show valid results.

In order to demonstrate the practical applicability of the presented relaxed PSD, two numerical examples were carried out. Reasonable results can be observed for both examples. The relaxed PSD thus enables the quantification of uncertainties efficiently in the frequency domain. Instead of discrete values, a range of possible values with different probabilities can be calculated, which is particularly useful in the area of reliability analyses in order to determine a probabilistic response in the frequency domain.

Relaxed power spectrum estimation from multiple data records utilising subjective probabilities

Marco Behrendt^{a,b}, Marius Bittner^a, Liam Comerford^b, Michael Beer^{a,b,c}, Jianbing Chen^d

^aInstitute for Risk and Reliability, Leibniz Universität Hannover, Germany

^bInstitute for Risk and Uncertainty, University of Liverpool, United Kingdom

^cInternational Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University, Shanghai, China

^dState Key Laboratory of Disaster Reduction in Civil Engineering & College of Civil Engineering, Tongji University, Shanghai, China

Published in *Mechanical Systems and Signal Processing* on 15 February 2022

Abstract

In structural dynamics, the consideration of statistical uncertainties is imperative to ensure a realistic modelling of loading and material parameters. It is well-known that any deterministic analysis only constitutes a single result for the given input parameters. Because of aleatoric or epistemic uncertainties, many factors must be considered either in certain intervals or with subjective probabilities. Especially for environmental processes, such as earthquakes or wind loads, a reliable prediction of future event characteristics is important for the design of safe structures. This work attends to the statistical procedure of simulating the response behaviour of a dynamic system under an excitation described by a stochastic process. A versatile option for this procedure is the estimation of the Power Spectral Density (PSD) function from real data records. The PSD function determines dominant frequencies and their magnitude of influence on the stochastic process. There are numerous methods for estimating the PSD function from source data, but usually these estimators do not account for uncertainties inherent in data records as they have a rigorous mathematical relationship between data and estimated PSD function. To address this issue, an approach for a stochastic load model that captures epistemic uncertainties by encompassing inherent statistical differences that exist across real data sets is proposed. Due to an increase in available data, reliable statistical information can be extracted from an ensemble of similar PSD functions that differ, for instance, only slightly in shape and peak frequency. Based on these statistics, a PSD function model is derived utilising subjective probabilities to capture the epistemic uncertainties and represent this information effectively. The spectral densities are characterised as random variables instead of employing discrete values, and thus the PSD function itself represents a non-stationary random process comprising a range of possible valid PSD functions for a given data set. This novel representation is useable for producing non-ergodic process realisations immediately applicable for Monte Carlo simulation analyses. The strengths and advantages are demonstrated by means of numerical examples.

Keywords: Power spectral density, Relaxed power spectrum, Random vibrations, Stochastic processes, Stochastic dynamics, Uncertainty quantification.

2.1 Introduction

In the vast field of stochastic dynamics, enormous progress has been made in recent years, which has contributed to improving simulation models or to better understanding the underlying physics [52, 55]. For the simulation of buildings and other structures in stochastic dynamics, models with excitation and response processes are required. These buildings are often subject

to excitations consisting of random vibrations [12, 54]. Stochastic processes are suitable to characterise these excitations, especially when the underlying excitation is an environmental process. Buildings can, for example, be subjected to permanent loads, such as wind loads, or in certain regions to very short-term loads, such as earthquakes. In order to analyse the influence of these processes on structures, whether to design safe structures for the future or to make predictions about the durability of existing buildings, these environmental processes are recorded and applied to building models. In many cases, a direct application of safety codes of civil engineering practice such as Eurocode 8 is not possible (e.g., for high-rise buildings, critical infrastructure such as bridges, or abnormal structures such as oil platforms or dams); instead, detailed time analyses and simulations must be carried out to ensure robust design [5]. For this purpose Monte Carlo (MC) simulations will be applied. It has been shown that for complex non-linear relations, MC simulations can compute probabilistic characteristics of the response [110].

A suitable and widely used tool in the modelling of stochastic processes is the Power Spectral Density (PSD) function, through which Housner proposed the white noise model [58]. Later, earthquakes were described via PSD functions, which have been continuously improved [59, 86]. With the help of PSD functions, environmental processes can be examined for their relevant frequencies and the corresponding amplitudes. In addition, the PSD function can be used to derive a relationship between excitation and response, at least in the linear case [111]. Furthermore, adequate stochastic processes can be generated via a descriptive PSD function [84], for example for extensive MC simulations. In this area, some progress has been made recently to reduce the number of random variables for the generation of time histories of signals, namely the Stochastic Harmonic Function (SHF) [91–93].

Thus, when estimating such a model by statistical analysis of source data, it is assumed that the governing process coincides with a single power spectrum. Without having significant confidence that this is the case, for any set of source data where the power spectrum is of interest, traditional statistical power spectrum estimation could result in a highly unrepresentative model of the process.

Thus, when estimating such a model by statistical analysis of source data, it is assumed that the governing process coincides with a single power spectrum. Without having significant confidence that this is the case, for any set of source data where the power spectrum is of interest, traditional statistical power spectrum estimation could result in a highly unrepresentative model of the process. Especially in the field of environmental stochastic load modelling, when estimating any spectral model from multiple source data sets, the assumption that each realisation, if it exists in the limit, corresponds to the same power spectrum is highly unlikely. Uncertainties introduced by estimating the PSD function from averaging the periodograms of multiple real data sets can be substantial and at least an attempt should be made to quantify them. An overview of efficient methods is given in [17] and the references therein.

Nowadays, for the purpose of simulations, databases with a large variety of environmental pro-

cesses are accessible, e.g. [112]. Nevertheless, it should be noted that even these data can often be subject to uncertainties caused by equipment failure, damaged sensors or measurement errors, to name a few examples. Missing data problems also occur frequently. Different approaches to deal with this issue and reconstruct data appropriately are given in [76–78] and have been successfully applied to real-world problems [106]. Epistemic uncertainties in the data sets can lead to incorrect or even dangerous results, for example if simulations with an uncertain data set indicates that a structure is not endangered, but the structure would actually be damaged by the excitation. Many of the well-known and commonly used PSD function estimators, such as periodogram, Bartlett’s method, Welch’s method or Blackman-Tukey method (see for example [61], [56]) do not take uncertainties in the data into account. Other models, such as the Super Asperity Model [113], take uncertainties into account, but are derived and valid only for specific regions and particular parameters. Moreover, it is usually time-consuming and costly to set up these kinds of models.

This paper proposes a stochastic load representation that takes epistemic model uncertainties into account and aims to be universally valid and independent of site-specific parameters. Attempts to represent the PSD function as a probabilistic model for missing data problems have already been made [77, 105]. However, in this work the excitation spectrum is estimated from multiple data sets with same characteristics to obtain a reliable probabilistic representation. The individual discrete frequencies will be treated as probability density functions and thus random variables, resulting in a model controlled by subjective probabilities. For this purpose, meaningful statistical variables of the ensemble are extracted. The main novelty of this load model is the accounting for epistemic uncertainties in data sets to derive a probabilistic description of the input data in the frequency domain. The structural response in the frequency domain can also be represented probabilistically, which enables likely and unlikely spectral density values to be identified by probability density functions, instead of using discrete and thus less accurate values. This facilitates the assessment of the risk to structures from environmental processes. The load model is thus directly useable for MC simulations and is referred to as the *relaxed* power spectral density function. The main difference between the traditional PSD function and the relaxed PSD function is that random variables are now used to describe the input PSD function, whereas previously discrete values were used. This allows uncertainties in the data to be better captured. Furthermore, the relaxed PSD function can be incorporated directly into recently developed methods for reliability assessment, such as the Probability Density Evolution Method [52, 114] and their extensions, see for instance [8, 115].

This work is structured as follows: A brief overview of stochastic processes, power spectral density functions and their relationship is given in section 2.2. Section 2.3 describes the estimation of the relaxed PSD function from an ensemble of given data utilising subjective probabilities. In addition, the differences to traditional methods are presented. The strengths and advantages of the relaxed PSD function are demonstrated in Section 2.4 by means of two numerical examples. The final conclusion is given in Section 2.5.

2.2 Stochastic process representation and power spectrum estimation

In this section a brief overview of the theory of stationary stochastic processes as well as the power spectrum estimation are presented.

For a given stationary power spectral density function $S_X(\omega)$, a stochastic process $X(t)$ can be generated using

$$X(t) = \sum_{n=0}^{N-1} \sqrt{4S_X(\omega_n)\Delta\omega} \cos(\omega_n t + \varphi_n), \quad (2.1)$$

where

$$\begin{aligned} \omega_n &= n\Delta\omega, \quad n = 0, 1, 2, \dots, N-1 \\ \Delta\omega &= \frac{\omega_u}{N}, \end{aligned} \quad (2.2)$$

with $N \rightarrow \infty$, ω_u as the upper cut-off frequency and φ_n as uniformly distributed random phase angles in the range $[0, 2\pi]$. This method is referred to as the Spectral Representation Method (SRM) [84] and has been widely used to generate artificial earthquake time histories, as in e.g. [116].

In general, this means that for every real-valued process $X(t)$ a corresponding orthogonal process $Z(\omega)$ exists and can be written in the form [11]

$$X(t) = \int_0^{\infty} e^{i\omega t} dZ(\omega), \quad (2.3)$$

where the process $Z(\omega)$ has the properties

$$\begin{aligned} E\left(\left|dZ^2(\omega)\right|\right) &= 4S_X(\omega)d\omega \\ E(dZ(\omega)) &= 0. \end{aligned} \quad (2.4)$$

To transform a stochastic process from time domain to frequency domain, the discrete Fourier transform is applied. A frequently used estimator of the power spectrum is the periodogram [12] which is the squared absolute value of the discrete Fourier transform of the time signal $X(t)$

$$\hat{S}_X(\omega_k) = \lim_{T \rightarrow \infty} \frac{\Delta t^2}{T} \left| \sum_{t=0}^{T-1} X(t) e^{-2\pi i k t / T} \right|^2, \quad (2.5)$$

with T as the total length of the record, Δt as the time increment, t describes the data point index in the record and k is the integer frequency for $\omega_k = \frac{2\pi k}{T}$.

2.3 Relaxed power spectrum and relaxed spectral representation method

In this section, the estimation of a relaxed PSD function from an ensemble of periodograms using subjective probabilities is given. This novel model combines the calculation of statistical

information from process record ensembles with probability density functions, which represent the individual frequencies. The differences in the estimation of stationary processes from the relaxed model are explained using short examples.

The strength of this novel method can be exploited when a large amount of data, for instance environmental processes such as earthquake ground motions, are available. Nowadays, a large number of time histories of earthquakes around the globe are provided among others in the PEER Ground Motion Database [112], which can be used to calibrate the input ensemble. The ensemble consists of a number of different periodograms with cardinality R , i.e. ensemble = $\{\hat{S}_{X_i}\}, i = 1, 2, \dots, R$. Here R is determined by the number of periodograms available, estimated either from real seismic ground motions or from artificially generated stochastic processes. Each periodogram in the set can be evaluated and yield a value $s_{i,\omega_n} = \hat{S}_{X_i}(\omega_n)$. The ensemble is an empirical collection of periodograms that determines the possible range in the probability input space for the relaxed PSD function described below.

Throughout this work artificially generated stochastic processes are used by utilising the following definition of the underlying analytical power spectral density function, describing an environmental process in the frequency domain, given in [84],

$$S_X(\omega) = \frac{1}{4}\sigma^2 b^3 \omega^2 e^{-b|\omega|}, \quad -\infty < \omega < \infty. \quad (2.6)$$

Here, $\sigma = 1$ is the standard deviation of the stochastic process and $b = 1$ is a parameter proportional to the correlation distance of the stochastic process [85]. For the generation of the time signals, a MC simulation with 50 samples is carried out using SRM (2.1), where the analytical PSD function in Eq. 2.6 is used as the source spectrum. The aim of this procedure is to simulate data sets and their corresponding estimated periodograms and to set up the novel relaxed PSD function. For the MC simulation, the following parameters are used in the SRM: Total length of time record $T = 64$ s, time step size $\Delta t = 0.25$ s, cut-off frequency $\omega_U = 12.5664$ rad/s, frequency step size $\Delta\omega = 0.0974$ rad/s and a total number of frequencies $N_\omega = 129$, see [84]. Since SRM (2.1) is influenced by random phase angles φ , different time signals are generated despite the identical source spectrum. These signals are used as input signals and are individually transformed to the frequency domain by Eq. 2.5, resulting in the so-called ensemble of periodograms, which is a collection of data in the frequency domain with similar characteristics, see Fig. 2.1. As mentioned before, for the ensemble presented here, a total of 50 artificially generated time signals and the corresponding estimated periodograms are employed. Although the individual periodograms in the ensemble show the same characteristics, such as peak frequency and shape, there is still variance among them. In order to estimate the relaxed PSD function from the ensemble, the mean μ_{ω_n}

$$\mu_{\omega_n} = \frac{1}{N} \sum_{i=0}^{N-1} s_{i,\omega_n} \quad (2.7)$$

and standard deviation σ_{ω_n}

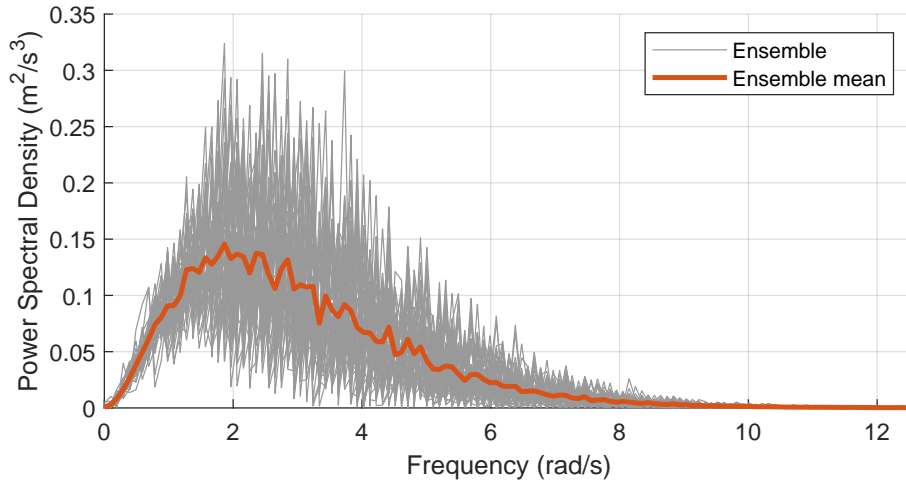


Figure 2.1: Ensemble $\{\hat{S}_{X_i}\}, i = 1, 2, \dots, 50$ with corresponding mean

$$\sigma_{\omega_n} = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} (s_{i,\omega_n} - \mu_{\omega_n})^2}, \quad (2.8)$$

with $\omega_n = n\Delta\omega$ and $n = 0, 1, 2, \dots, N - 1$, must be calculated for each individual frequency. By extracting this statistical information from the ensemble, a probability distribution function can be derived for each frequency. Here, a truncated normal distribution is utilised and given by

$$f_{\omega_n}(s; \mu, \sigma, a, b) = \frac{1}{\sigma} \frac{\Phi\left(\frac{s-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}, \quad (2.9)$$

with $\Phi(\eta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\eta^2\right)$ as standard normal distribution and $\Phi(\zeta) = \frac{1}{2} \left(1 + \operatorname{erf}\left(\zeta/\sqrt{2}\right)\right)$ as the corresponding cumulative distribution function. The truncation bounds are described with a and b . The subscripts in Eq. 2.9 are omitted for simplicity. The probability density function must be determined separately for each frequency ω_n , since dependencies and correlations between the frequencies are not taken into account. As negative values are not possible in terms of power spectral density functions, the truncation bound a must not be smaller than 0. The utilisation of a truncated normal distribution may result in a shift of the mean value. Therefore, to highlight the influence of the truncation bounds, the bounds $[a_{\omega_n} = 0, b_{\omega_n} = \infty]$ and $[a_{\omega_n} = 0, b_{\omega_n} = 2\mu_{\omega_n}]$ are used in this work.

Using the frequency $\omega = 2.16$ rad/s as an example, the spectral densities from the ensemble are depicted with the fitted truncated normal distribution in Fig. 2.2. Both the histogram and the corresponding empirical cumulative distribution function of spectral densities are presented in blue. The fitted probability density function with truncation bounds $[0, \infty]$ and the corresponding cumulative distribution function are also shown in red.

From Fig. 2.2, it can be obtained that the truncated normal distribution fits the data well. This is also confirmed by various statistical tests in which the null hypothesis could not be rejected.

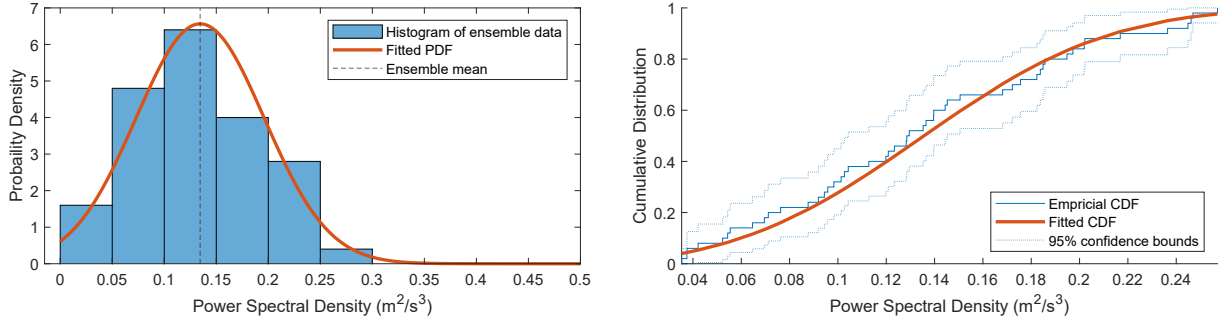


Figure 2.2: Histogram and empirical CDF of spectral densities at frequency $\omega = 2.16$ rad/s with fitted truncated normal distribution

Table 2.1: Statistical tests

	Kolmogorov-Smirnov	Chi-square	Anderson-Darling
p -value	0.9494	0.3676	0.3933

The corresponding p -values can be found in Table 2.1.

The resulting relaxed PSD functions estimated from the ensemble of periodograms is depicted in Fig. 2.3. Note that the probability density is normalised here.

For instance, the probability density functions for the three selected frequencies $\omega = 1.08$ rad/s, $\omega = 2.65$ rad/s and $\omega = 4.22$ rad/s are depicted in Fig. 2.4. The solid lines show the probability density functions for the relaxed PSD function with truncation bounds $[0, \infty]$, while the dashed lines represent the same for a relaxed PSD function with truncation bounds $[0, 2\mu]$. These lines describe “slices” of the relaxed PSD functions (Fig. 2.3) for selected frequencies and exemplify the probability density for those frequencies assigned to the PSD values.

The derived relaxed PSD function can be utilised for producing non-ergodic stochastic process realisations immediately applicable for MC simulations. In order to generate adequate time signals for MC simulations, SRM (Eq. 2.1) can be reformulated to

$$X(t, \theta) = \sum_{n=0}^{N-1} \sqrt{4S_{X, \omega_n}(\omega, \theta) \Delta\omega} \cos(\omega_n t + \varphi_n(\theta)) \quad (2.10)$$

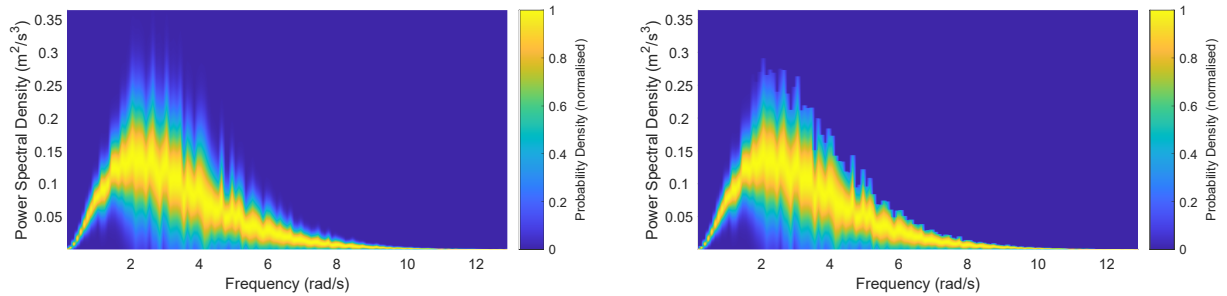


Figure 2.3: Normalised relaxed PSD functions corresponding to Eq. 2.9 with truncation bounds $[0, \infty]$ (left) and $[0, 2\mu]$ (right)

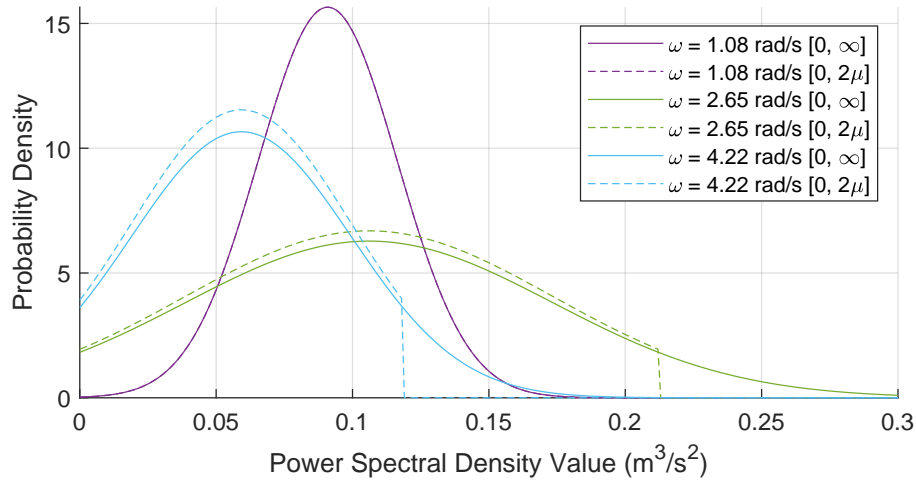


Figure 2.4: Probability density functions for selected frequencies ω_n in Eq. 2.9 of the relaxed PSD functions

and shall be called the Relaxed Spectral Representation Method (RSRM). In this equation, the spectral density $S_X(\omega_n)$ is replaced by the sampled spectral density value $S_{X,\omega_n}(\omega, \theta)$ from the relaxed PSD function of the corresponding frequency.

Since the difference between Eq. 2.10 and Eq. 2.1 is only that a sampled PSD function from the relaxed PSD function is used instead of a deterministic/analytical PSD function (such as Eq. 2.6), the conditions and requirements on the stochastic process are not changed and still apply to both equations, see [84]. Therefore, the generated time signals in Eq. 2.10 have identical properties to those in Eq. 2.1, such as stationarity and non-ergodicity. Furthermore, in both cases zero-mean signals are generated.

As the spectral densities for individual frequencies are treated as random variables controlled by subjective probabilities, the PSD function itself is a non-stationary random process. The relaxed PSD functions derived here will be used for further investigations and simulations throughout this work.

2.3.1 Convergence of the relaxed PSD function

In this section, a brief convergence study is conducted to show that the PSD functions sampled from the relaxed PSD function converge towards the target relaxed PSD function. For this purpose, 100,000 PSD functions were sampled from the relaxed PSD function and adequate time signals were generated using Eq. 2.10. The time signals were transformed back to the frequency domain by Eq. 2.5. From the resulting ensemble, the relaxed PSD function was calculated again. Fig. 2.5 shows the deviation of the estimated PSD function and the original PSD function for the mean value and the standard deviation, as well as the convergence towards the target relaxed PSD function, calculated via the Euclidean norm.

With increasing sample size, the mean PSD function converges to the target mean PSD function. However, the standard deviation is slightly higher than in the target function. This can be attributed to the sampling process. With a high number of samples, higher, respectively lower

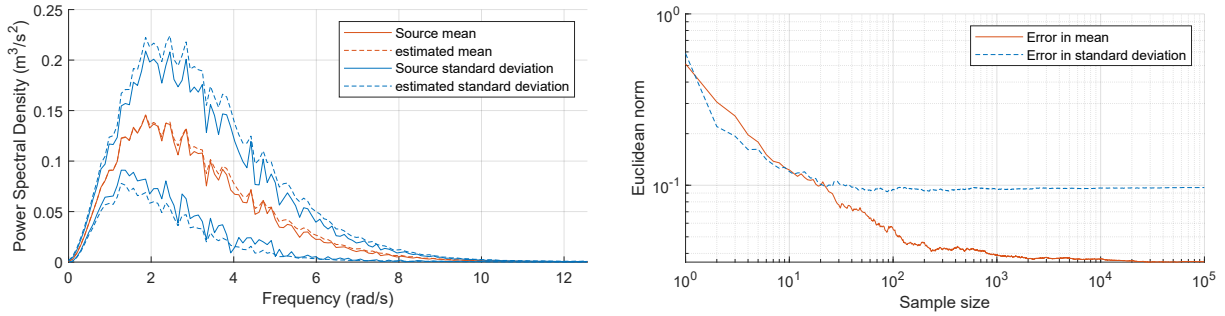


Figure 2.5: Target and converged mean and standard deviation of the relaxed PSD function with corresponding error

spectral densities are sampled for the individual frequencies that do not occur in the original ensemble, which leads to higher standard deviations being calculated. This fact is not critical for the simulation as the relaxed PSD function is still able to account for epistemic uncertainties, for which this new representation of PSD functions is intended.

Convergence was also assessed with different input ensembles from which different relaxed PSD functions were derived. Therefore, it can be stated that convergence holds for arbitrary relaxed PSD functions.

2.3.2 Influence on Generated Time Signals

To investigate the influence of the relaxed PSD function on the generated time signals, different simulations were carried out. In addition, to obtain comparative values, the standard and commonly used SRM was utilised in combination with the ensemble mean to generate time signals.

First, an example for two sampled power spectral densities from the relaxed PSD function and the corresponding generated time signals are depicted in Fig. 2.6. It should be noted, that correlations or dependencies between frequencies are not considered here. Similarities such as peak frequency and shape can be recognised, but differences can be seen in the specific values of the spectral density, of course. These differences are propagated to the generated signals. However, using the exact same random phase angles φ in Eq. 2.10, which neglects the randomness in the phase shift and therefore provides a look on the amplitude changes only, slight differences in the generated time signals can be seen. Shape and amplitudes of the time signals show similarities, but still differ which may cause a difference in the application to a system, depending on the system's properties.

Since for the simulation of buildings and structures certain limit values in the excitations are often decisive, e.g. when regarding first passage failure conditions or reliability over time estimates, the attention is directed to the maximum amplitudes in the time signals. For this purpose, 10,000 time signals are generated with SRM (Eq. 2.1) and RSRM (Eq. 2.10) and the respective absolute maximum values are plotted as a histogram. In addition, the empirical cumulative distribution functions are presented, see Fig. 2.7. Histograms and CDFs are presented

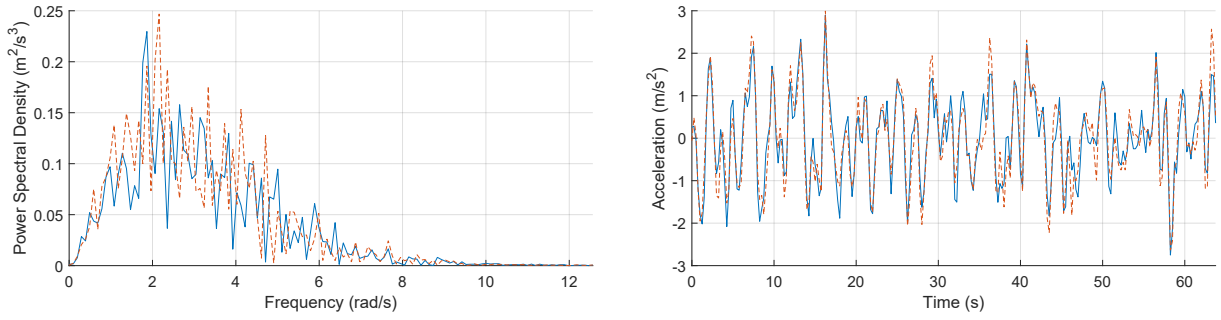


Figure 2.6: Example for two sampled PSD functions and the corresponding generated stochastic processes (Eq. 2.10) by utilising identical random phase angles φ

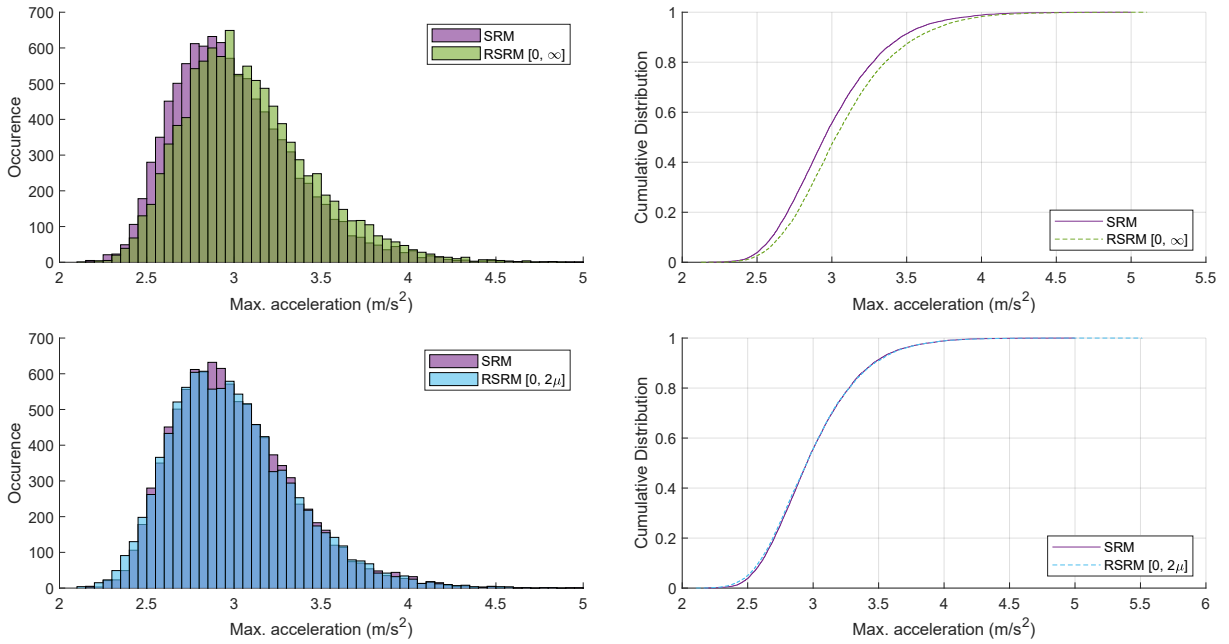


Figure 2.7: Histogram and empirical CDF of the maximum acceleration $\max_{t \in [0, T]} (|\dot{x}(t)|)$ of generated time signals by SRM and RSRM with truncation bounds $[0, \infty]$ (top) and $[0, 2\mu]$ (bottom)

for both relaxed PSD functions with truncation bounds $[0, \infty]$ as well as for $[0, 2\mu]$.

The results clearly show that the truncation bounds of $[0, \infty]$ lead to a shift of the mean in the direction of higher maximum values, which leads to a higher probability of failure of buildings and structures. When using the truncation bounds of $[0, 2\mu]$, the histogram and CDF are almost identical to those of conventional SRM, since only a smaller range is thus possible for sampling the spectral densities.

As a final investigation, the autocorrelation functions of SRM and RSRM are compared. An analytical expression for the autocorrelation function is given by

$$R_{XX}(\tau) = \sigma^2 \frac{b^4(b^2 - 3\tau^2)}{(b^2 + \tau^2)^3}, \quad -\infty < \tau < \infty, \quad (2.11)$$

with $\sigma = 1$ and $b = 1$ identical as in Eq. 2.6. The parameter τ describes the time lag [84, 85].

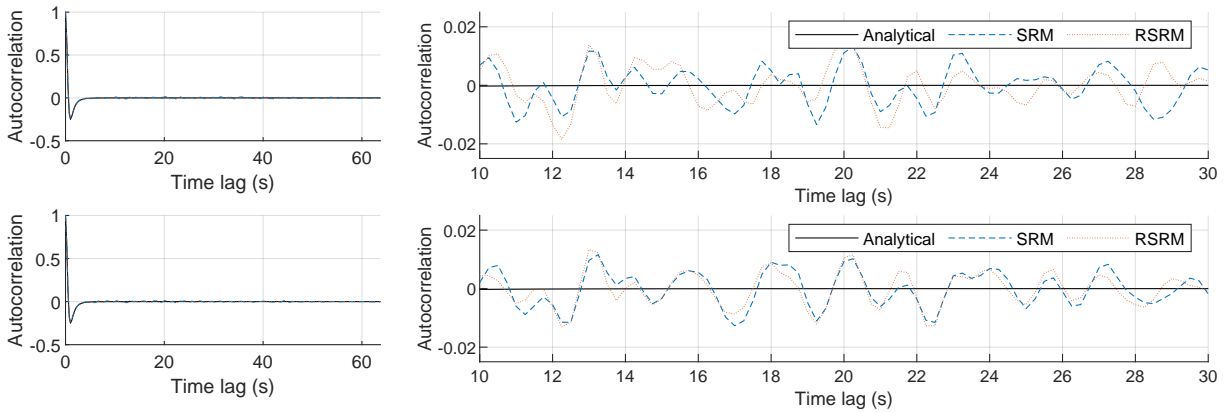


Figure 2.8: Mean autocorrelation for 100 (top) and 1,000 (bottom) time signals

Two examples are presented. It is important to note that the same random phase angles φ were used for the generated time signals in Eq. 2.1 and Eq. 2.10 in order to exclude differences due to the randomness of the phase angles. For the first example, 100 time signals were generated and the mean value of the autocorrelation was calculated. The result is depicted in Fig. 2.8. It can already be seen in the enlarged area that the autocorrelation does not exceed the value $|0.02|$. For the second example, 1000 samples were used. The results are also depicted in Fig. 2.8. An improvement can be obtained compared to the example with 100 samples, but no significant improvement in the autocorrelation was detected by further increasing the number of samples.

2.4 Numerical examples

To show the strengths and advantages of the relaxed PSD function, it is applied in the context of two numerical examples. The two derived models of a relaxed PSD function with truncation bounds $[0, \infty]$ and $[0, 2\mu]$ are utilised as excitation (see Fig. 2.3). A total of 10,000 MC samples were generated for each of the simulations. In order to obtain a relaxed response spectrum in the frequency domain of the systems, the same procedure to generate the relaxed excitation from the ensemble (Eq. 2.7-2.9) is applied to the individual responses obtained by MC simulation.

2.4.1 Linear SDOF oscillator

For the first example, a simple Single Degree-of-Freedom (SDOF) linear oscillator in the form

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = F(t), \quad (2.12)$$

with mass $m = 50$ kg, damping coefficient $c = 37.2 \frac{\text{Ns}}{\text{m}}$ and spring constant $k = 1922 \frac{\text{N}}{\text{m}}$ is utilised. 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 process by utilising the RSRM (Eq. 2.10). An explicit Runge-Kutta scheme according to [117] is used to solve Eq. 2.12. After applying the MC samples of both relaxed PSD functions to the system, the relaxed response

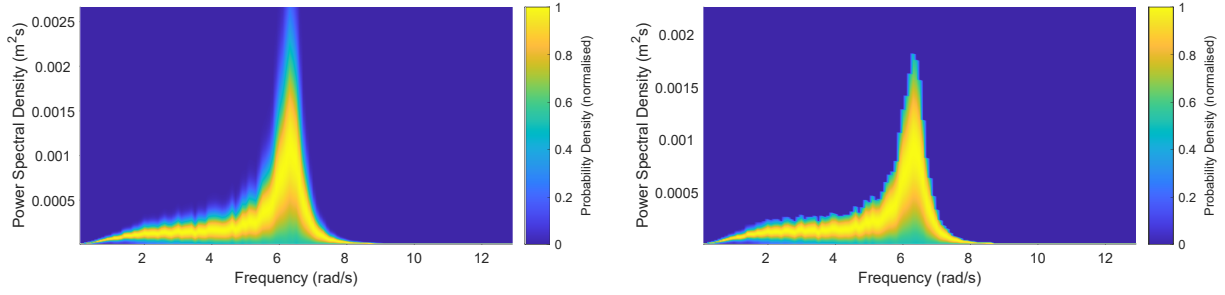


Figure 2.9: Normalised relaxed response spectra of the linear oscillator corresponding to Eq. 2.9 with truncation bounds $[0, \text{inf}]$ (left) and $[0, 2\mu]$ (right)

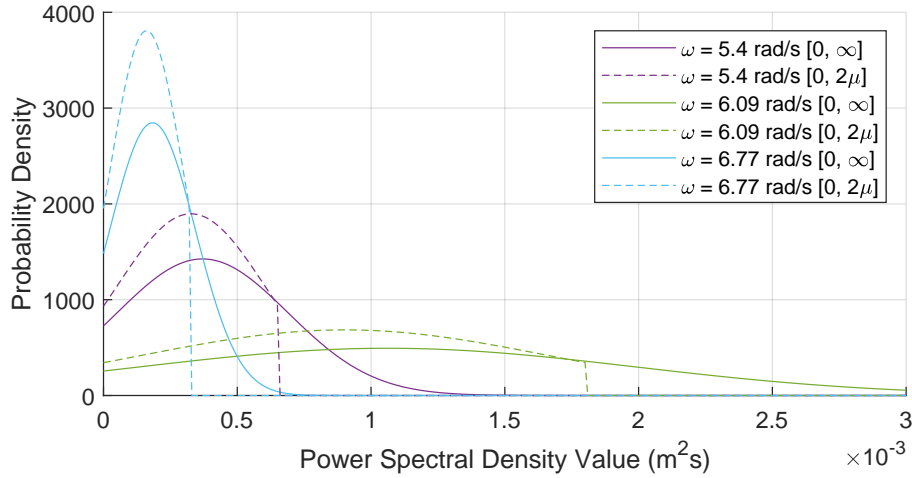


Figure 2.10: Probability density functions for selected frequencies ω_n in Eq. 2.9 of the relaxed response spectra of the linear oscillator

spectra of the system's displacement were obtained from the individual responses. These results can be seen for both excitations in Fig. 2.9.

As expected, the responses of the two relaxed excitation spectra look similar, as the excitation also shows only minor differences. However, the excitation with larger truncation bounds also leads to a clearly more relaxed behaviour in the response. This also shifts the mean value slightly in the direction of higher spectral densities. This is also confirmed when looking at the probability density functions for three frequencies around the natural frequency of the system (see Fig. 2.10). The PDFs of the same frequencies have similar shapes, but differ in mean value and probability density, which is due to the bounds.

For a simple system like the one used here, analytical results are often available. Here, the analytical solution can be calculated via the transfer function

$$H(\omega) = \frac{1}{\omega_0^2 - \omega^2 + 2\xi\omega_0\omega i} \quad (2.13)$$

and the frequency response function

$$S_Y(\omega) = S_X(\omega)|H(\omega)|^2. \quad (2.14)$$

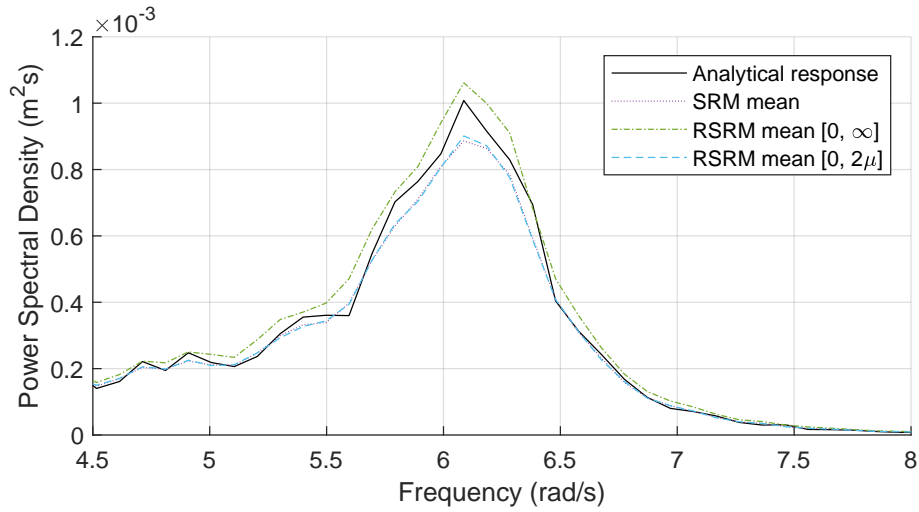


Figure 2.11: Comparison of the numerical and analytical mean response spectra

In Eq. 2.13, $\omega_0 = \sqrt{\frac{k}{m}}$ describes the natural angular frequency of the system, $\xi = \frac{c}{2\sqrt{km}}$ is the damping ratio and i is the imaginary unit. In Eq. 2.14, $S_X(\omega)$ is the excitation spectrum and $S_Y(\omega)$ describes the response spectrum. For better comparability, the analytical system response for the ensemble mean (see Fig. 2.1) is therefore calculated. The resulting analytical response spectrum and the mean spectrum of the SRM as well as for the two relaxed PSD functions are depicted in Fig. 2.11.

In addition to the results in the frequency domain, those in the time domain are also important, e.g. for reliability analysis. Therefore, the focus will now be on the resulting system displacement, more precisely on the maximum displacement of the system obtained by the MC samples generated from the relaxed PSD functions. Fig. 2.12 shows the histograms of the resulting maximum displacements of the simulations, each from the simulation with the standard SRM and with the RSRM. The empirical cumulative distribution functions are shown to the right. Here again, similar to the behaviour in frequency domain, a shift of the mean value towards higher maximum displacements can be obtained with the relaxed power spectrum with truncation bounds $[0, \infty]$. This behaviour becomes particularly clear when looking at the cumulative distribution functions. The entire function is shifted towards higher displacements yielding a higher failure probability of the system. Utilising the truncation bounds of $[0, 2\mu]$, the mean value is almost identical to that of the SRM. However, it can be seen here, especially in the CDF, that both smaller and larger maximum system displacements can be noticed, which can be explained by the relaxed nature of the model.

2.4.2 Non-linear Bouc-Wen-Baber-Noori model

Another simulation was carried out with a more complex Multiple Degree-of-Freedom (MDOF) model. A non-linear eight-storey shear frame structure was utilised, which was simulated adopting the Bouc-Wen-Baber-Noori model. The model is an extension of the equation of motion

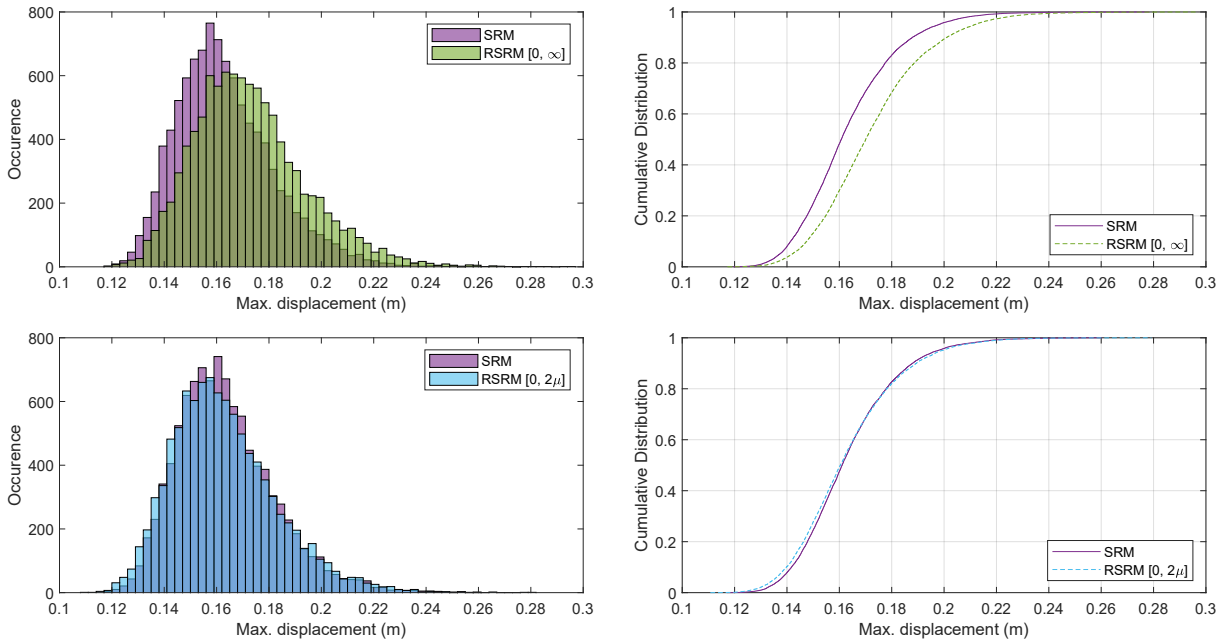


Figure 2.12: Histogram and empirical CDF of the maximum displacement $\max_{t \in [0, T]} (|x(t)|)$ of the response of the linear oscillator by utilising SRM and RSRM with truncation bounds $[0, \infty]$ (top) and $[0, 2\mu]$ (bottom)

(Eq. 2.12) and reads as follows

$$\mathbf{M}\ddot{x}(t) + \mathbf{C}\dot{x}(t) + \alpha\mathbf{K}x(t) + (1 - \alpha)\mathbf{H}z(t) = \mathbf{F}(t), \quad (2.15)$$

where \mathbf{M} , \mathbf{C} , \mathbf{K} and \mathbf{H} are mass matrix, damping matrix, stiffness matrix and hysteretic matrix, respectively. The parameter α controls the "degree" of non-linearity and z is the hysteretic variable, a pseudo-displacement. For a detailed description of the model see e.g. [118, 119] and [120–122] for its extension. The model parameters given in Table 2.2 are adopted from [123].

The mass and stiffness given in Table 2.3 are utilised for the structure.

To show the non-linear behaviour of the building, the restoring force versus the displacement of the first storey is plotted in Fig. 2.13.

As in the linear example before, the two relaxed PSD functions with truncation bounds $[0, \infty]$ and $[0, 2\mu]$ are used as the excitation (see Fig. 2.3). Fig. 2.14 shows the response in the frequency domain for the excitation with truncation bounds $[0, \infty]$ on the left. On the right is the response of the excitation with bounds $[0, 2\mu]$ depicted. The responses were calculated from the displacements of the top storey of the structure. In Fig. 2.14, the areas apart from the natural frequency of the building have been omitted as they are close to zero.

A similar behaviour as in the previous linear example results can be obtained. For both relaxed excitations, a relaxed response can also be determined, which can be used for uncertainty quantification. It can also be seen that the truncation bounds $[0, \infty]$ again lead to a more relaxed response. However, the difference between the two relaxed responses is not as high as in the linear example, which is particularly clear in Fig. 2.15, where the “slices” of the relaxed PSD

Table 2.2: Parameter for the Bouc-Wen-Baber-Noori model [123]

Parameter	Value	Description
α	0.01	Ratio of linear to non-linear response
A	1	Basic hysteresis shape control
β	1.4	Basic hysteresis shape control
γ	0.2	Basic hysteresis shape control
n	1	Sharpness of yield
δ_v	0.002	Strength degradation
δ_η	0.001	Stiffness degradation
ζ_s	0.95	Measure of total slip
q	0.25	Pinching initiation
p	2	Pinching slope
Ψ	0.2	Pinching magnitude
δ_ψ	0.005	Pinching rate
λ	0.1	Pinching severity/rate interaction

Table 2.3: Structural parameters of the eight-storey shear frame structure

Storey	1	2	3	4	5	6	7	8
Mass ($\times 10^5$ kg)	3.5	3.3	3.0	3.0	3.0	3.0	3.0	2.7
Stiffness ($\times 10^{10}$ N/m)	1.47	1.63	1.62	1.60	1.60	1.92	1.85	0.96

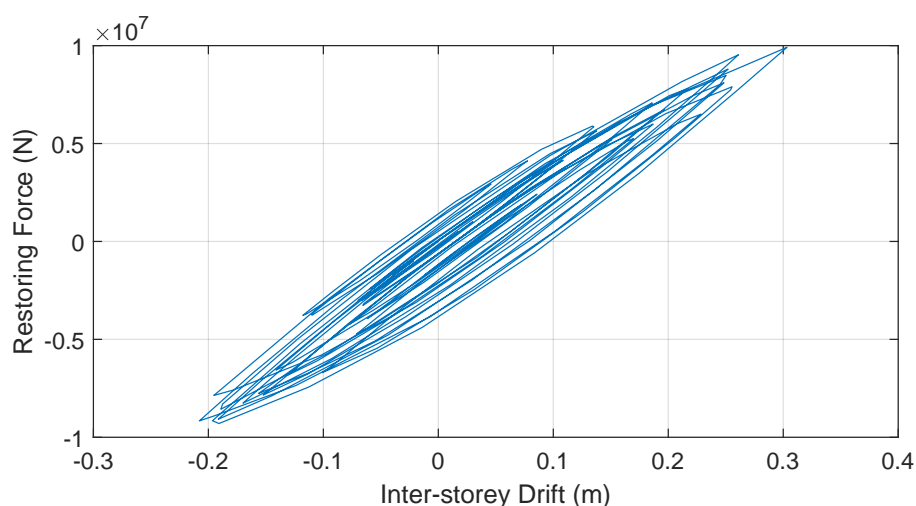


Figure 2.13: Restoring force versus inter-storey drift for the first storey of the non-linear Bouc-Wen-Baber-Noori model

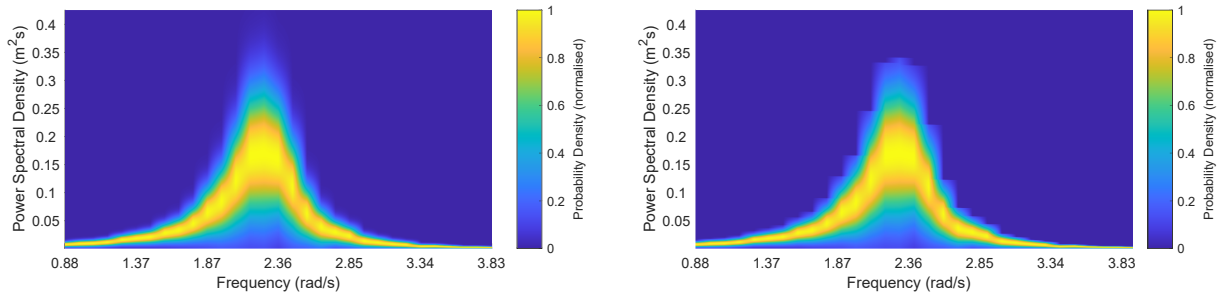


Figure 2.14: Normalised relaxed response spectra of the Bouc-Wen-Baber-Noori model corresponding to Eq. 2.9 with truncation bounds $[0, \infty]$ (left) and $[0, 2\mu]$ (right)

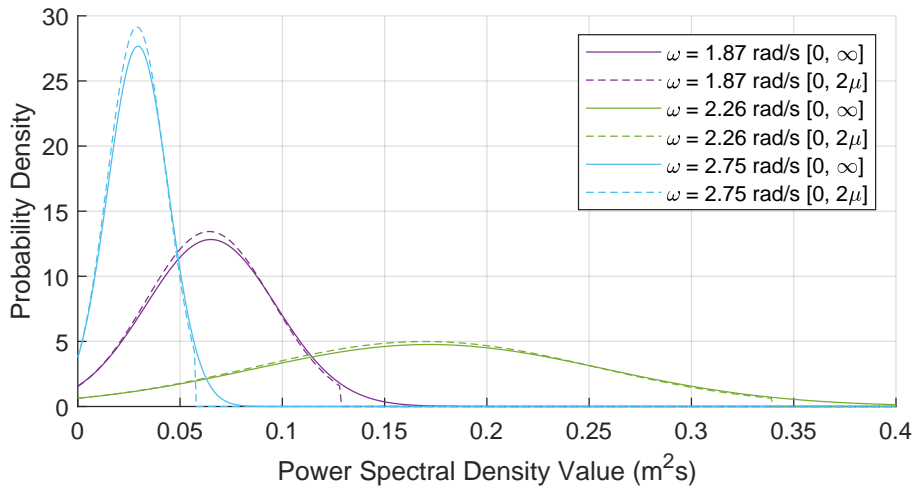


Figure 2.15: Probability density functions for selected frequencies ω_n in Eq. 2.9 of the relaxed response spectra of the Bouc-Wen-Baber-Noori model

functions in Fig. 2.14 are given. The influence thus seems to be smaller in the case of a non-linear MDOF system. Nevertheless, a slight shift of the mean value towards higher spectral densities due to the wider truncation bounds compared to the excitation with smaller truncation bounds can be seen here as well.

Considering the results in the time domain, the behaviour is similar to the linear system, although not as strong. The left side of Fig. 2.16 shows the histograms of the maximum displacements from the top storey of the building resulting from the MC simulation of the two relaxed excitation spectra and the traditional SRM. Time histories, generated from the sampled PSD functions of the relaxed PSD function, were applied to the structure. From the histograms it can be seen that the mean is almost identical to that of SRM. However, lower and higher maximum system displacements can also be determined. This becomes clear when looking at the right side of Fig. 2.16. There, the empirical cumulative distribution functions corresponding to the histograms are shown. For both cases it can be identified that the CDFs intersect at a certain point. Here, the relaxed excitations lead to a higher dispersion of the maximum values into higher and lower ranges. This also allows the determination of different failure probabilities of buildings and structures.

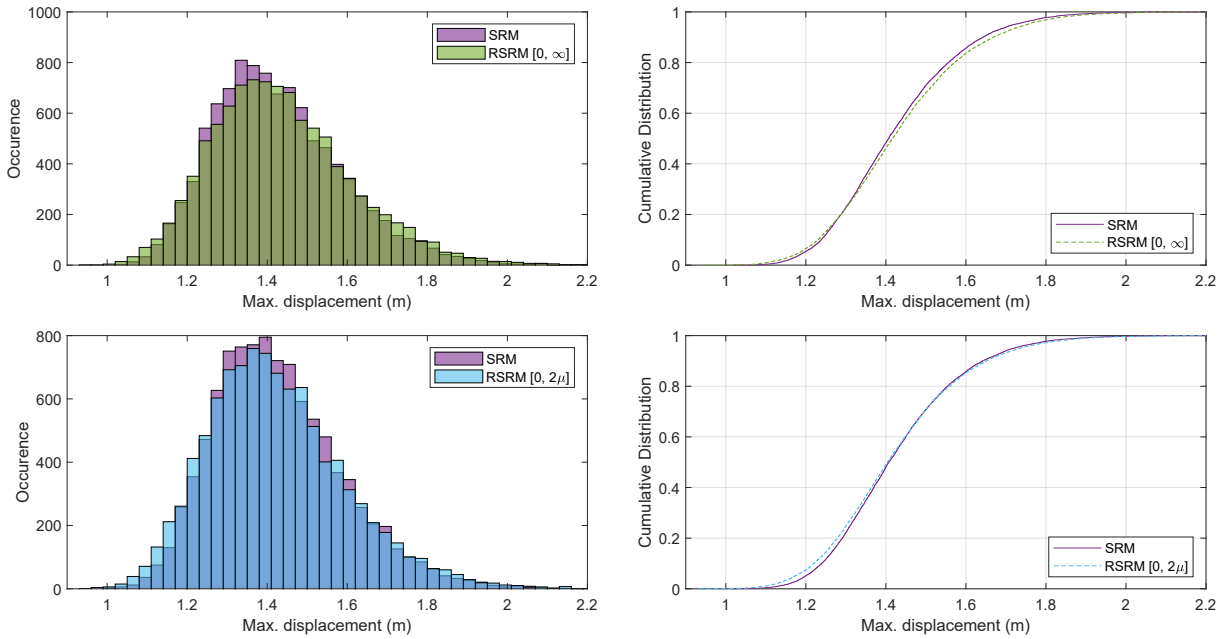


Figure 2.16: Histogram and empirical CDF of the maximum displacement $\max_{t \in [0, T]} (|x(t)|)$ of the response of the Bouc-Wen-Baber-Noori model by utilising SRM and RSRM with truncation bounds $[0, \infty]$ (top) and $[0, 2\mu]$ (bottom)

2.5 Conclusion

In this work, a new model of a PSD function incorporated in the generation of stochastic processes by spectral representation is proposed: the relaxed PSD function. Compared to traditional methods, where the spectral densities are described with discrete values, the relaxed PSD function offers several advantages. By describing the spectral densities using probability density functions, different probabilities can be assigned to the densities. Especially in the case where a large amount of data is available, an approach with subjective probabilities is useful, as it is possible to obtain reliable statistical information from the ensemble of data. By using a probabilistic approach, the relaxed PSD function is more robust in the presence of outliers and captures epistemic uncertainties.

Although correlation and dependence between frequencies is not considered, the variance changes throughout the spectrum. This is simply due to the non-negative nature of the PSD function, i.e. smaller mean densities must lead to smaller variances. Hence, the relaxed PSD function is defined as a non-stationary random process with mean values and uncorrelated variances indexed over a finite frequency space.

An important aspect in the development of the relaxed PSD function is the choice of probability density function used to represent the spectral densities. Here, a truncated normal distribution is suggested and utilised, but not limited to this. Other PDFs are also possible, as this depends mainly on the underlying data, such as its shape and appearance. In addition, it is possible to control the probability density functions, e.g. by changing the truncation bounds to vary the

width of the spectrum. One option, for instance, is to make them dependent on the mean and standard deviation instead of using the entire positive range. Both cases were demonstrated in this work. However, it must be ensured that the bounds are chosen large enough, as too narrow bounds are very restrictive and the relaxed PSD function hardly differs from traditional methods in the simulation results. Manipulating the parameters of the probability density functions will yield in a new shape and will of course also influence the simulation results. It must therefore be ascertained that a change in the input data does not falsify the output data.

Another important point for developing a relaxed PSD function is to guarantee that only data that has similarities is used. For example, the data must have similar peak frequencies, shape and spectral densities. If the data sets vary too much from each other, it is almost impossible to determine a relaxed PSD function that represents the entire ensemble. A robust and reliable relaxed PSD function can therefore only be derived from similar data records.

The developed load model is suitable for producing non-ergodic process realisations that are directly useable for MC simulation analyses. Due to the sampling approach used, the system response in the frequency domain can also be represented as a relaxed model. Instead of a discrete but uncertain response of the system, the spectral density can thus be determined in a probabilistic representation. Based on the determined probability densities, a range can be identified with some certainty in which the actual response is located. With discrete spectral densities, on the other hand, only a specific value is given without safety margins.

The simulation results in the frequency domain exhibit, depending on the chosen truncation bounds, a shift of the mean value towards higher spectral densities, as well as a shift towards higher system displacements in the time domain. This can be attributed to the fact that the probability density functions used cannot sample values smaller than zero due to the non-negative nature of the PSD function, but there is no limitation in the direction of higher densities when utilising the truncation bounds $[0, \infty]$. This slightly shifts the mean value to higher spectral densities and results in an imbalance and in a non-symmetric sampling of the spectral densities. However, this does not falsify the results, which is confirmed by the fact that the mean value of the SRM is also in the range of high probability densities of the relaxed response. The effect of the shifted mean values is amplified by a higher standard deviation in the PDFs. Therefore, the standard deviation of the relaxed PSD function should not be too high, as the results may then no longer be representative. Again, it should be noted that this can be circumvented if the data from which the PSD function is estimated are similar, thus excluding very high standard deviations. Due to the shifted mean, the relaxed PSD function with high standard deviation can be considered a more conservative method compared to SRM.

Probabilistic consideration of spectral densities by means of normally distributed random variables increases the dimension of random variables. In general, methods with a small number of random variables are to be preferred in order to avoid stochastic fluctuations that are strongly influenced by this randomness. Therefore, every attempt should be made to reduce the random variables. The use of Stochastic Harmonic Functions (SHF) or bivariate probability distributions

provide a good basis.

In summary, the relaxed PSD function offers a useful and valuable extension to traditional methods such as the Spectral Representation Method.

Acknowledgement

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, Grants BE 2570/4-1, CO 1849/1-1) and the National Natural Science Foundation of China (NSFC, Grant 11761131014) under the Sino-German research project 'Uncertainty modelling in power spectrum estimation of environmental processes with applications in high-rise building performance evaluation'.

3 | Stochastic process generation from relaxed power spectra utilising stochastic harmonic functions

This part of the thesis deals with the generation of stochastic processes from the relaxed PSD introduced in Chapter 2. The stochastic processes employed reflect the characteristics of a PSD function in the time domain. However, since random variables are required to generate these processes, an investigation of their influence is highly important. As a large number of random variables affects the results and can therefore only be averaged out by a large number of samples and thus a high simulation effort, it is advisable to use methods that keep the number of random variables low. In the context of this work, an investigation is carried out with different methods, in particular the Spectral Representation Method (SRM), as well as the method of Stochastic Harmonic Functions (SHF).

The SRM is a traditional method for generating time signals using a known PSD and has been used in Chapter 2 to generate stochastic processes from the relaxed PSD. Although SRM yields robust results and also has low autocorrelation in the time domain, the number of random variables required is relatively high, as one random variable is needed per each discrete frequency of the PSD. SHF, on the other hand, offers a promising alternative, as this method defines intervals from which a spectral density is sampled. The sampled spectral density is then used to generate a time signal. Thus, the number of random variables can be significantly reduced. The disadvantage of this method, however, may be a higher autocorrelation of the time signal. Another difficulty is the extension of the random dimension by using PDFs in the relaxed PSD. Using SRM, the number of random variables would thus be doubled, since for each frequency the randomly sampled spectral density would also be added. SHF offers a good alternative to reduce the number of random variables.

However, using SHF to generate stochastic processes from the relaxed PSD is still a compromise, as the number of random variables can be reduced, but will increase the autocorrelation. It can be shown that the new method yields better results, with a slight decrease in the number of random variables required.

Stochastic process generation from relaxed power spectra utilising stochastic harmonic functions

Marco Behrendt^{a,b}, Marius Bittner^{a,c}, Michael Beer^{a,b,d}

^aInstitute for Risk and Reliability, Leibniz Universität Hannover, Germany

^bInstitute for Risk and Uncertainty, University of Liverpool, United Kingdom

^cInternational Research Training Group 2657 – Computational Mechanics Techniques in High Dimensions,
Leibniz Universität Hannover, Germany

^dInternational Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University,
China

Published in *Proceedings of the 8th International Symposium on Reliability Engineering and Risk Management (ISRERM 2022)* in September 2022

Abstract

In order to design and safely construct buildings and structures that are exposed to environmental processes such as earthquakes and wind loads, simulations are essential in advance. Although simulations are an approximation of reality, they are still dominated by uncertainties, that must be taken into account. These uncertainties can arise for various reasons, such as incorrectly recorded data or inaccurate simulation models. One widely used approach for generating and simulating environmental processes is the power spectral density (PSD) function. It establishes a relationship between the time and frequency domains and determines the relevant frequencies and their magnitude of the transformed signals. Since the model of the PSD function provides discrete values of the density for each frequency, the idea arises to model this density uncertain. For this purpose, statistical values are extracted from an ensemble of similar PSD functions that differ slightly in shape and peak frequency, for example. These values are used to derive a relaxed model of a PSD function. By using such a model, the response processes of buildings and structures are also uncertain, resulting in a range of possible values instead of discrete values. However, a disadvantage of this model is a higher number of random variables that can affect the simulation results. In this work, the stochastic harmonic functions (SHF) are used instead of the spectral representation method (SRM). Stochastic processes generated with both SRM and SHF are compared in a benchmark simulation to assess the advantages.

Keywords: Power spectral density function, Stochastic dynamics, Stochastic processes, Spectral representation, Uncertainty quantification.

3.1 Introduction

Many problems in engineering sciences are subject to random vibrations and thus lead to stochastic dynamic problems, for example, where environmental processes have an influence. These examples could be high-rise buildings, bridges or critical infrastructure subjected to earthquake, wind loads, or offshore platforms excited by waves. The fields of random vibrations [12, 53, 54, 111, 124] and stochastic dynamics [52, 55, 125, 126] are well described in the literature. Problems arise due to the structural complexity, incomplete information of the system or hardly predictable natural processes. Aleatory and epistemic uncertainties exist. Therefore, for these aforementioned facilities often a direct application of safety codes of civil engineering practice

is not possible, which demands detailed time analyses and simulations to ensure a robust design [5]. To determine the influence of environmental processes on structures and to determine their response behaviour, simulations with either artificial or recorded data are imperative.

Not only the model design and natural influences but also real data records for the system analysis are often subject to uncertainties. These uncertainties can arise due to various reasons, such as a limited amount of samples, damaged sensors, device failure, perhaps due to the environmental process itself, sensor threshold limitations and measurement errors. Additionally, the sensor may capture the data incorrectly, e.g., extreme features. Other causes such as sensor maintenance, bandwidth limitation or data acquisition restrictions could lead to poor quality of the data records as well. Due to uncertainties in real data records incorrect estimated system responses can be obtained. Therefore, uncertainties must be mitigated as much as possible.

Since stochastic dynamics have been studied very efficiently in recent decades, different models have been developed to describe seismic ground motion. One of these tools to capture information of phenomena involving random fluctuations in time and space is the power spectral density (PSD) function, which is widely used in the modelling of stochastic processes, especially in applications such as earthquake, wind and ocean engineering [52, 54, 55]. In earthquake engineering, for example, the use of the PSD function dates back to Housner [58] or Kanai [59]. The available information of the probabilistic models for the input is often not enough to directly assess the resulting randomness of a structural response. For seismic ground motions with critical peaks, the structural response is generally non-linear and a direct relation of the probabilistic input models to the probabilistic output is not available. For this purpose the Monte Carlo (MC) simulation will be applied. It has been shown that for complex non-linear relations MC can compute probabilistic characteristics of the response [110].

For the development of a load model that uses such a random excitation process, the following applies: The more real data sets are available, the better, since the numerical results are statistically more accurate for a large amount of data. In addition, it is useful if the underlying physics is understood well in order to make further assumptions. Since there is often not enough data available and too little is known about the physics of the structures, other approaches must be found to develop a load model that represents the data in the best possible way [106]. When estimating the PSD function from multiple real data records by averaging their periodograms, it is assumed that the process defined by the records may be represented by a single PSD function. When dealing with real data, the uncertainty introduced by this assumption could be significant and at the very least, an attempt should be made to quantify it. Such a load model, the so-called relaxed PSD function, was developed by the authors of this work in [127]. Meaningful statistical variables of the ensemble are extracted for each frequency separately. Based on these variables a probability density function for each single frequency is calculated, which results in a probabilistic model of the PSD function. In [127] the classical spectral representation method (SRM) is utilised, see e.g., [84, 128]. In this work, the stochastic process generation based on the relaxed PSD function is extended by using stochastic harmonic functions (SHF) [91]. In [92]

SHF was successfully expanded to non-stationary stochastic processes and in [129] a data-driven description using SHF to generate wind induced vibration on a high-rise building was applied. Continued efforts led to a further reduction in the number of random variables [93].

This work is organised as follows: An overview of environmental processes and PSD function estimation is given in Section 3.2. A brief overview of SHF is given in Section 3.3. Section 3.4 introduces the relaxed PSD function that uses SRM and the relaxed PSD function that replaces SRM with SHF. A benchmark simulation and the corresponding results are given in Section 3.5. The final conclusion is given in Section 3.6.

3.2 Spectral representation and PSD function estimation

To generate a stationary stochastic process, the following model, formulated by Shinozuka, is considered in this work [84]

$$X(t, \theta) = \sum_{n=0}^{N-1} \sqrt{4S_X(\omega_n)\Delta\omega} \cos(\omega_n t + \varphi_n(\theta)), \quad (3.1)$$

where $\omega_n = n\Delta\omega$ with $n = 0, 1, 2, \dots, N-1$, $\Delta\omega = \omega_u/N$ and $\varphi_n(\theta)$ with $n_\varphi = N$ are uniform distributed phase angles in the range $[0, 2\pi]$. This method is referred to as the SRM and has been widely used to generate artificial earthquake time histories, as in e.g., [116]. For the PSD function used to generate the sample functions in Eq. 3.1, $S_X(\omega)$ is given as the two-sided PSD function

$$S_X(\omega) = \frac{1}{4} \sigma^2 b^3 \omega^2 e^{-b|\omega|}, \quad (3.2)$$

in which $b = 1$ and $\sigma = 1$. This specific PSD function in combination with Eq. 3.1 resembles a stochastic process $X(t)$ with zero mean, standard deviation σ equal to one and autocorrelation function $R_{XX}(\tau)$

$$R_{XX}(\tau) = \sigma^2 \frac{b^4(b^2 - 3\tau^2)}{(b^2 + \tau^2)^3}. \quad (3.3)$$

In this equation τ denotes the time lag. To transform a stochastic process from time domain to frequency domain, the discrete Fourier transform is applied more precisely the discrete Fourier transform (DFT) algorithm. A frequently used estimator of the PSD function is the periodogram [12] which can be understood as the squared absolute value of the discrete Fourier transform of the time signal x_n

$$\hat{S}_X(\omega_k) = \frac{\Delta t^2}{T} \left| \sum_{n=0}^{N-1} x_n e^{-2\pi i k t / N} \right|^2 \quad (3.4)$$

In this equation, T is the total length of the record in time, Δt is the time increment, m describes the data point index in the record and k is the integer frequency for $\omega_k = \frac{2\pi k}{T}$.

An example for a generated stochastic process using Eq. 3.1 from the given analytical two-sided PSD function (Eq. 3.2) is depicted in Fig. 3.1.

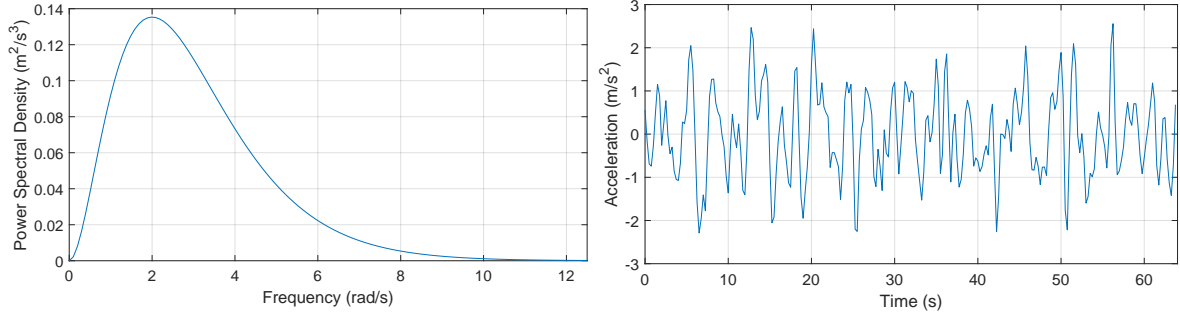


Figure 3.1: Analytical PSD function (left) and generated stochastic process (right).

3.3 Stochastic harmonic function representation

A major drawback of SRM is that when dealing with long time intervals a high number of summation terms in Eq. 3.1 are necessary to avoid a high correlation of realised signals, this leads to a high number of random variables needed. To overcome this drawback in [91] the SHF representation was developed by introducing a second random variable, $\tilde{\omega}_n(\theta)$, to generate sample functions $X_{SHF}(t, \theta)$. The frequency domain of the PSD function is therefore distributed into N_{SHF} equally spaced intervals $[\omega_{n-1}, \omega_n]$ with $\omega_0 = \omega_L$ and $\omega_{N_{SHF}} = \omega_U$, where ω_L and ω_U are the lower and upper bound cut-off frequency, respectively.

The generation of an example function with the following content describes the reformulation of the SRM

$$X_{SHF}(t, \theta) = \sum_{n=0}^{N_{SHF}-1} \sqrt{2}A(\tilde{\omega}_n(\theta)) \cos(\tilde{\omega}_n(\theta)t + \varphi_n(\theta)), \quad (3.5)$$

where the amplitude factor A is defined as

$$A(\tilde{\omega}_n(\theta)) = \sqrt{2S_X(\tilde{\omega}_n(\theta))\Delta\tilde{\omega}_n} \quad (3.6)$$

and $\Delta\tilde{\omega}_n = \omega_n - \omega_{n-1}$ is the length of the interval. In addition to the random variable $\varphi_n(\theta)$, the frequency component $\tilde{\omega}_n(\theta)$ is drawn from a uniform distribution over the interval $[\omega_{n-1}, \omega_n]$. In [91] it has been shown that the total number of needed random variables, i.e. $N_\varphi + N_{\tilde{\omega}} = 2N_{SHF}$, to resemble specific input PSD functions was smaller than using only one random dimension in SRM.

The procedure of introducing random distributed ω_n 's over specific intervals is also possible by describing a different underlying distribution function than the uniform distribution. The method presented here is in fact called stochastic harmonic function of the second kind (SHF-II) because it utilises a uniform distribution of the frequency realisations over their pre defined support intervals.

3.4 Relaxed PSD function

The relaxed PSD function is a probabilistic representation of the set of power spectra. This approach was derived and presented in detail in [127], whereas a brief overview is given here. The strength of this novel method can be utilised when a large amount of data, specifically earthquake time histories are available. Nowadays for example on the PEER Ground Motion Database [130], a large amount of time histories of earthquakes around the globe is provided which can be used to calibrate the input ensemble. The ensemble is a set of different PSD functions with cardinality R , i.e. ensemble = $\{\hat{S}_{X_i}\}$, $i = 1, 2, \dots, R$. Here, R is determined by the number of available PSD functions estimated either from real seismic ground motions or from artificially created stochastic processes. Each PSD function in the set can be evaluated and yield a value $s_{i,\omega_n} = \hat{S}_{X_i}(\omega_n)$. The ensemble is an empirical collection of PSD functions that determine the possible range in the probability input space for the relaxed PSD function described below. However, before merging these data into an ensemble, it is necessary to verify that the spectral similarity is high enough to derive a load model based on these, as it is counter-intuitive to generate a load model from significantly dissimilar data. An approach for determining their similarity is presented in [131].

3.4.1 Relaxed PSD function utilising a truncated normal distribution

For the calculation of a relaxed PSD function representation of an ensemble of PSD functions a truncated normal distribution is utilised. In order to define the probabilistic representation, the mean

$$\mu_{\omega_n} = \frac{1}{R} \sum_{i=1}^R s_{i,\omega_n} \quad (3.7)$$

and the standard deviation

$$\sigma_{\omega_n} = \sqrt{\frac{1}{R} \sum_{i=1}^R (s_{i,\omega_n} - \mu_{\omega_n})^2}, \quad (3.8)$$

are calculated for each frequency component $\omega_n = n\Delta\omega$, $n = 0, 1, 2, \dots, N - 1$. By extracting these statistical information from the ensemble, a truncated normal distribution (Eq. 3.9) can be defined for each frequency component in order to develop the relaxed PSD function. The probability density function is given by

$$f_{\omega_n}(s; \mu, \sigma, a = 0, b = \infty) = \frac{1}{\sigma} \frac{\varphi\left(\frac{s-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}, \quad (3.9)$$

and has to be determined for each frequency ω_n separately. In Eq. 3.9 $a = \max(0, \min(0, \mu_{\omega_n} - 4\sigma_{\omega_n}))$ and $b = \max(\mu_{\omega_n} + 4\sigma_{\omega_n}, 2\mu_{\omega_n})$ defining the interval, $\varphi(\xi) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\xi^2\right)$ is the standard normal distribution and $\Phi(s) = \frac{1}{2} \left(1 + \operatorname{erf}(s/\sqrt{2})\right)$ is the corresponding cumulative distribution function. As negative values are not possible in terms of PSD functions, the interval

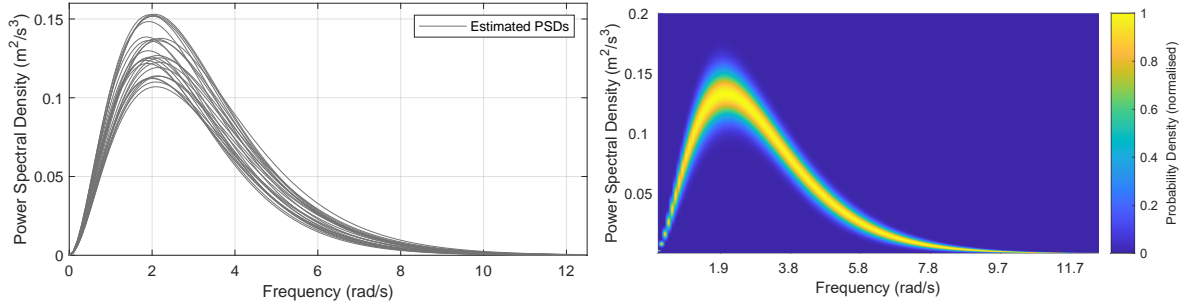


Figure 3.2: Ensemble of PSD functions (left) and relaxed PSD function (right).

$[a_{\omega_n}; b_{\omega_n}]$ is defined beforehand for each frequency.

3.4.2 Relaxed spectral representation method

Using Eq. 3.9 as the shaping stochastic process of the relaxed PSD function, the SRM method can be rewritten and shall be called the relaxed spectral representation method (RSRM)

$$X(t, \theta) = \sum_{n=0}^{N-1} \sqrt{4S_{X, \omega_n}(\theta) \Delta\omega} \cos(\omega_n t + \varphi_n(\theta)), \quad (3.10)$$

where $S_{X, \omega_n}(\theta)$ denotes that corresponding to each ω_n a PSD function value can be drawn from the distribution function f_{ω_n} (Eq. 3.9) utilising a sampling technique, such as inverse sampling. This leads likewise as for $\varphi_n(\theta), n_\varphi$ to a number of n_S new random distributions, which are now distinctively for each ω_n defined, i.e. $S_{X, \omega_n}(\theta), n_S$. This extension random dimension wise is not ideal. However, for a first accurate approximation this approach was chosen.

It is possible to say that the probability density function with changing bounds in Eq. 3.9 from which a sample for each frequency is drawn, describes a non-stationary stochastic process itself in the frequency domain. Here the relation lies between the value of the PSD function amplitude S_X and the frequency ω .

3.4.3 Relaxed stochastic harmonic function representation

To derive an alternative approach to RSRM, the classical SRM is replaced by SHF-II. Instead of determining a truncated normal distribution corresponding to (Eq. 3.9) for each frequency, this will be done only for each interval defined by SHF. The probability density function is calculated at each interval's midpoint. This provides the best possible approximation of each interval's PSD function value in a probabilistic representation. The RSRM method Eq. 3.10 can be rewritten and shall be called relaxed stochastic harmonic function method (RSHF-II). This novel description of stochastic processes reads as follows

$$X_{SHF}(t, \theta) = \sum_{n=0}^{N_{SHF}-1} \sqrt{4S_{X, \tilde{\omega}_n \theta}(\theta) \Delta\tilde{\omega}_n} \cos(\tilde{\omega}_n(\theta)t + \varphi_n(\theta)), \quad (3.11)$$

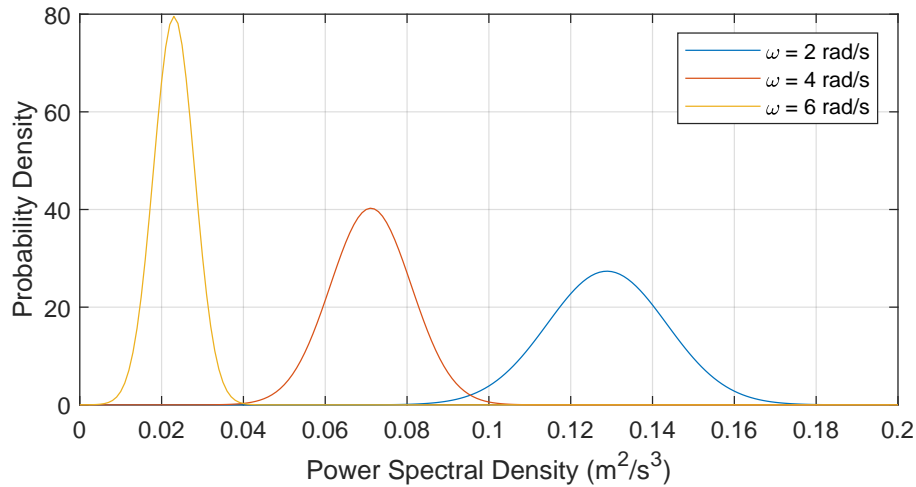


Figure 3.3: Probability density functions for selected frequencies.

where $S_{X, \tilde{\omega}_n \theta}(\theta)$ describes the PSD function value drawn from a truncated normal distribution based on the ensemble and sampled $\tilde{\omega}_n(\theta)$.

Although replacing SRM by SHF-II and thus introducing the intervals for frequency domain increases the number of random dimensions, the total number of random variables can be reduced, which will be demonstrated in the following benchmark simulation.

3.5 Benchmark setup and results

The primary goal of the benchmark is to assess the accuracy and performance of all presented methods to generate a stationary stochastic process with underlying PSD function given in Eq. 3.2. For the relaxed PSD function the parameters b and σ vary in a small range, 25 different PSD functions are used to build the ensemble as discussed in Section 3.4 and shown in Fig. 3.2. For instance, the probability density functions for the frequencies 2 rad/s, 4 rad/s and 6 rad/s for the relaxed PSD function are depicted in Fig. 3.3. The resulting sample functions of $X(t)$ from Eq. 3.1 are chosen to be the reference since the accuracy of SRM for this specific setup has been shown in [84] to be very good. Different criteria will be analysed and shown in this section.

First, just to visualise the different signals for each method one realisation for each method is shown in Fig. 3.4. Here the number of sample functions used is indicated by n_θ , on the right-hand of each generated signal. As discussed in the earlier sections, due to the different sampling techniques and random information weaved into the sample function generation, the random dimension n_θ differs. SRM with $n_\theta = 128$ is chosen to be the benchmark, since this configuration has been shown in [84] to deliver a very good approximation. From Fig. 3.4 it is possible to see that the amplitude of all signals seems to vary approximately between -2 and 2 , other than that not much of information can be drawn from one signal itself.

To overcome the lack of information when regarding and comparing single signals to one and

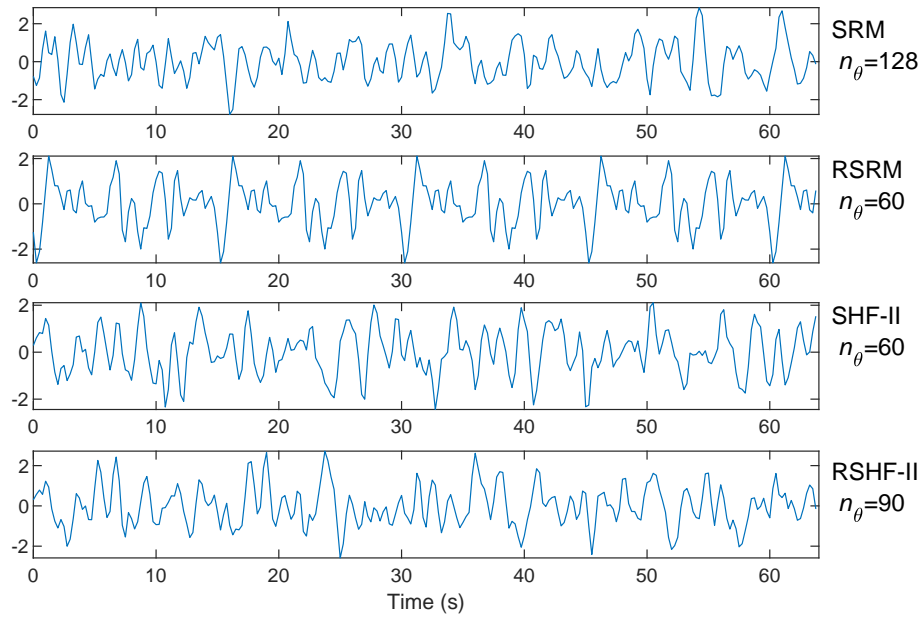


Figure 3.4: Realisations of the sample functions for SRM: Eq. 3.1, SHF-II: Eq. 3.5, RSRM: Eq. 3.10, RSHF-II: Eq. 3.11

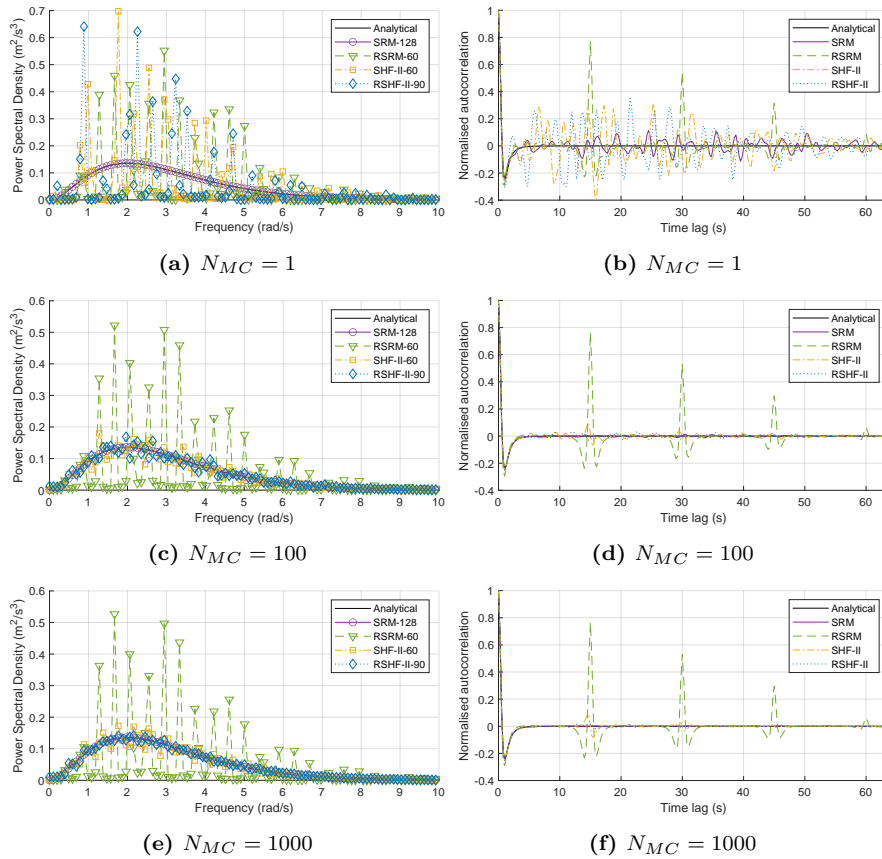


Figure 3.5: Mean PSD after applying the DFT of MC output (left) and mean autocorrelation for different MC simulations (right) according to Eq. 3.13.

another, MC simulation is applied. It is fair to say at this point that this indicates the problem that usually a single generated signal does not carry enough information to resemble the input characteristics e.g., the PSD function out of which the signal is generated. This procedure has also been applied in [91] and physically makes sense since the goal is to describe random processes in nature and each one of them is different and varies in a range. It is sufficient to say that a total ensemble resembles the information given to the process generation procedure.

Second, in Fig. 3.5 the results of the MC simulation can be seen, here on the *left side* the mean PSD function approximation for different Number of MC samples N_{MC} is shown, on the *right side* the corresponding mean autocorrelation of all signals generated during the MC simulation. For the *left side* it can be said that by increasing the number of samples N_{MC} , the accuracy to resemble the analytical PSD function (Eq. 3.2) of SHF-II and RSHF-II is improving. When comparing (c) and (e) it is also possible to see that the accuracy of RSHF-II is better than SHF-II. As already mentioned, for this specific example setup SRM is performing nearly perfect while RSRM yields in poor results.

On the *right side* the autocorrelation for a sole signal (b) and the mean autocorrelation for N_{MC} signals (d), (f) are shown. Here in (b) can be seen that even though SRM is performing the best given the aligned setup of the example it does show a not neglectable deviation. In (d) and (f), analogous to the findings of the results on the *right side*, by increasing N_{MC} the autocorrelation is decreasing and therefore the ulterior method is performing better (SRM, SHF-II, RSHF-II). The exception here is RSRM which shows large correlations that indicate a repeating pattern in the signal after a time lag $\tau \approx 15$ s. This can already be seen with a precise look on the RSRM result in Fig. 3.4.

Finally, to assess the overall performance of the presented methods, the results of reproducing the PSD function and the characteristics of the autocorrelation, are analysed in a more extensive MC simulation concerning the error in the Euclidean norm of the respective signal's PSD functions and autocorrelation. The error ϵ^S is calculated compared to the analytical PSD function in Eq. 3.2. ϵ^A compares the numerical mean autocorrelation with the analytical autocorrelation given in Eq. 3.13. Suppose \hat{X} are N_{MC} artificially generated signals of any presented method. Let $R_{\hat{X}\hat{X}}$ be the autocorrelation of a deterministic signal calculated by

$$R_{\hat{X}\hat{X}}(\tau) = \sum_{m=1}^{\frac{T_E}{\Delta t}} \hat{X}(m\Delta t + \tau)\hat{X}^*(m\Delta t), \quad (3.12)$$

where the asterisk flags the complex conjugate. In a normalised fashion this is rewritten to be $\hat{R}(\tau) = \frac{1}{\sqrt{R_{\hat{X}\hat{X}}(0)^2}}R_{\hat{X}\hat{X}}(\tau)$. With this normalised autocorrelation vector values the mean over all MC simulations is calculated by using

$$\bar{\hat{R}}(\tau) = \bar{\hat{R}}_{\hat{X}\hat{X}}(\tau) = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \hat{R}_i(\tau). \quad (3.13)$$

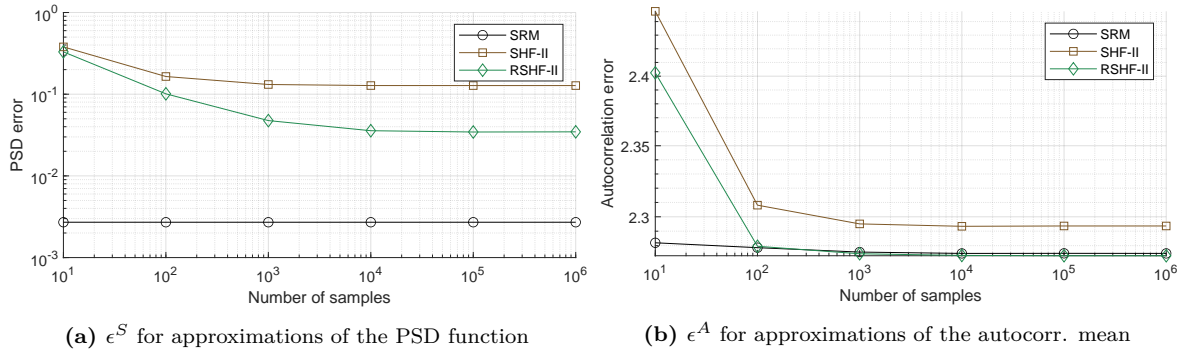


Figure 3.6: Error in Euclidean norm for PSD function approximation and autocorrelation for different numbers of MC samples.

Table 3.1: Error in PSD function for 10⁶ MC samples.

Error	Value
ϵ_{RSRM}^S	0.9747
ϵ_{SHF-II}^S	0.1273
$\epsilon_{RSHF-II}^S$	0.0346
ϵ_{SRM}^S	0.0027

The errors are always defined comparing two vectors in the Euclidean norm using following expression

$$\epsilon = \epsilon(V, \hat{V}) = \sqrt{\sum_{k=1}^K |V(k) - \hat{V}(k)|^2}, \quad (3.14)$$

here \hat{V} can either be a vector of PSD function values (Eq. 3.4) or a vector of normalised mean autocorrelation values (Eq. 3.13).

In Fig. 3.6 the resulting values ranging from one sample to one million samples are presented in a log-log scale. The key information that can be drawn from this figure is that in both cases RSHF-II is converging faster to a smaller error than SHF-II. The corresponding values are given in Table 3.1 for the PSD function approximation and in Table 3.2 for the autocorrelation. For the latter case RSHF-II is even performing better in the Euclidean norm than SRM. Since RSRM converges to the highest errors and these errors are significantly larger compared to the others, RSRM is omitted in Fig. 3.6, however, the corresponding values are given in Table 3.1 and Table 3.2 for comparison.

Table 3.2: Error in autocorrelation function for 10⁶ MC samples.

Error	Value
ϵ_{RSRM}^A	3.1118
ϵ_{SHF-II}^A	2.2935
$\epsilon_{RSHF-II}^A$	2.2730
ϵ_{SRM}^A	2.2746

3.6 Conclusion

Based on a benchmark model it has been shown that the novel method RSHF-II used to generate stationary stochastic processes is performing better than SHF-II, by only a slight increase in the number of random variables used during the process. A better approximation with less MC samples can be achieved. However, this is still a trade-off since RSHF is introducing another random dimension and for each frequency interval, a new probability distribution function was introduced to handle the uncertainty in the PSD functions amplitude. Nonetheless, this poses a challenge and at the same time offers a chance. The future challenge lies in reducing the total number of random variables used in the whole process of generating these relaxed PSD functions. This might be achievable by using multivariate distribution functions for each frequency interval with a specific correlation dependant on the PSD functions slope/gradient or by introducing hard dependencies between the random dimensions, based on data-driven information. The chance for the future application of this method is to incorporate data of natural influences such as wind, earthquake or vibrations where it is possible to assume that the signals somehow must base on a similar PSD function. But in most cases it is impossible to define a crisp PSD function for these signals. So gathered data can be transformed into an ensemble which is the foundation of the relaxed PSD function.

At a first glance at the results, the performance of SRM is always outstanding but it has to be mentioned that for all calculations the total number of random variables for SRM is always 128, this is considerably larger than the number of random variables used in SHF-II and also larger than the number of random variables used in RSHF-II. This fact should be kept in mind, especially when evaluating the error graphs in Fig. 3.6. That means, by changing only one of the parameters for this example represented in [128], the results yielded by SRM are getting worse. Since the methods of RSHF-II shall be included in the practical analysis where the ensemble information to define the relaxed PSD function is gathered from data, a more robust method to parameter and setup changes is mandatory. Different time-increments, different frequency discretisations or a different number of random variables yield very different results. For these scenarios SRM seems to be not applicable, therefore the usage of RSHF-II can yield a faster and more accurate representation of pre-defined signals with specific data driven information.

A crucial point is the variation of the signals time step size, for complex structural analysis an adaptive time step size control is crucial to perform a stable analysis, especially for large structures analysed using e.g., Finite Element models. A change of parameters in this regard for the signal generation sets the correct representation of a signal's PSD function for SRM out of sync, meanwhile methods like SHF-II and the relaxed PSD functions still perform robust.

Lastly, the computational performance needs to be considered, as SHF-II and RSRM are introducing new random variables interval-wise, which is kind of the description of a stochastic process but only by generalising information into intervals, which are in fact more distinct. Due to this fact for each interval new random variables needs to be generated and because of the RSFH-II dependence on the ensemble, for each MC calculation new distribution functions needs

to be defined. For now they are defined within the generation of the sample functions and therefore reduces the performance of the signal generation greatly, this should be considered to be done previously before the signal generation. That means for each sample function the random realisations should be made in advance to increase the computational performance.

4 | Estimation of an imprecise power spectral density function with optimised bounds from scarce data for epistemic uncertainty quantification

The following chapter deals with the derivation of an imprecise PSD model for the quantification of uncertainties. The initial problem is identical to that explained in Chapter 2. However, while the assumption was made that a large data set of similar PSD functions is available, this chapter deals with the same issue under the assumption that only limited data are available. A probabilistic model, as before, cannot be reliably derived because no safe statistical values such as mean and standard deviation can be extracted due to a lack of data. A probabilistic model generated from a limited amount of data can therefore be highly erroneous and falsify the simulation results. A reliability analysis based on such a model would have to be considered very critically. For this reason, an interval approach is used in this work.

The idea is to compute an interval-based PSD model from a limited data set, generating an upper and lower bound based on the data. For this purpose, a radial basis function (RBF) network is used to approximate the so-called basis spectrum, which is determined to get a first approximation of the shape of the PSD functions. The basis functions resulting from the RBF network are manipulated via the associated weights in such a way that a reliable upper and lower bound can be calculated via these basis functions. This is done in the context of an optimisation with the constraint that all PSD functions are included in the bounds. This approach also reflects the physics of the data set and takes into account dependencies between frequencies. This results in the derivation of reliable bounds for the corresponding data set.

The presented model is directly applicable to interval propagation due to the definition via basis functions only. An advantage of this approach is that the user has direct control over the number of basis functions and can thus plan for the simulation. The number of basis functions can also determine how the resulting bounds look like. The analyst thus has some flexibility in deriving the imprecise PSD. The model itself is able to quantify epistemic uncertainties that occur in the data set and thus make the simulation results more reliable, which is a great advantage in the context of reliability analyses. Therefore, this work contributes significantly to the quantification of uncertainties with a limited amount of available data.

Estimation of an imprecise power spectral density function with optimised bounds from scarce data for epistemic uncertainty quantification

Marco Behrendt^a, Matthias G.R. Faes^b, Marcos A. Valdebenito^b, Michael Beer^{a,c,d}

^aInstitute for Risk and Reliability, Leibniz Universität Hannover, Callinstraße 34, 30167 Hannover, Germany

^bChair for Reliability Engineering, TU Dortmund University, Leonhard-Euler-Strasse 5, 44227 Dortmund, Germany

^cInstitute for Risk and Uncertainty, University of Liverpool, Liverpool L69 7ZF, United Kingdom

^dInternational Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University, Shanghai 200092, China

Submitted to *Mechanical Systems and Signal Processing* on 9 September 2022

Abstract

In engineering and especially in stochastic dynamics, the modelling of environmental processes is indispensable in order to design structures safely or to determine the reliability of existing structures. Earthquakes or wind loads are examples of such environmental processes and can be described by stochastic processes. Such a process can be characterised by the power spectral density (PSD) function in the frequency domain. The PSD function determines the relevant frequencies and their amplitudes of a given time signal. For the reliable generation of a load model described by a PSD function, uncertainties that occur in time signals must be taken into account. This work mainly deals with the case where data is limited and it is infeasible to derive reliable statistics from the data. In such a case, it may be useful to identify bounds that characterise the data set. The proposed approach is to employ a radial basis function network to generate basis functions whose weights are optimised to obtain data-enclosing bounds. This results in an interval-based PSD function. No assumptions are required about the distribution of the data within those bounds. Thus, the spectral densities at each frequency are described by optimised bounds instead of relying on discrete values. The applicability of the imprecise PSD model is illustrated with recorded earthquake ground motions, demonstrating that it can be utilised for real world problems.

Keywords: Power spectral density function, Random vibrations, Stochastic processes, Stochastic dynamics, Imprecise probabilities, Uncertainty quantification.

4.1 Introduction

The robust determination of the reliability of buildings and structures in engineering and especially in the field of stochastic dynamics is of utmost importance [9, 52, 55, 125]. Buildings and structures are subject to random vibrations induced, for example, by environmental processes such as earthquakes or wind loads [53, 54, 111]. These loads initiate a dynamic system behaviour of the structures. To determine whether this can lead to critical system behaviour, simulations can be carried out as part of a reliability analysis. Simulations are an important part of engineering, especially to determine failure probabilities of such structures. This can be done for existing structures or for the design of new structures in the future.

Within the framework of spectral analysis, a signal can be decomposed into its harmonic com-

ponents via the Fourier transform, which allows it to be examined for dominant frequencies and their amplitude by means of the power spectral density (PSD) function [12, 61]. The PSD function is an important tool for determining whether the governing frequencies of the excitation interfere with those of the structure under investigation, which can lead to dangerous system behaviour. For linear systems, a relationship between input and output PSD can be derived, while for non-linear systems, a time signal analysis must be conducted. Various methods can be used to generate time signals that intrinsically reflect the characteristics of the PSD and thus represent it in the time domain. Such artificially generated time signals can be used to perform reliability analyses, e.g. in the context of Monte Carlo simulations [22, 110] and other advanced sampling techniques such as subset sampling [23], line sampling [132], directional importance sampling [24] or others.

In general, data records are subject to uncertainties, which may stem, for example, from measurement errors, damaged or inaccurately calibrated sensors or from a limited number of available data, see for instance [4, 6]. Transformations based on estimations, such as a PSD estimation, can introduce additional uncertainties, as some of these estimators may provide results of poor quality [61]. To obtain reliable simulation results, these uncertainties must be considered in the representation of the physical process. If these uncertainties are not taken into account or are incorrectly quantified, this can lead to fatal misinterpretations of the results. For example, a building may be classified as safe under a certain load, when in reality it has a high risk of damage or collapse. The consideration of uncertainties in data sets is therefore of utmost importance to obtain reliable simulation results. Typically, uncertainties can be divided into aleatory and epistemic uncertainties [1]. While aleatory uncertainties are irreducible, epistemic uncertainties can be reduced, for example, by obtaining further information. There are different general approaches available to quantify these uncertainties depending on their source and occurrence, such as probabilistic models [9, 10], non-probabilistic models [13] or imprecise probabilistic models [17]. Specific methods are, for instance, p-boxes [133], which are used to bound the cumulative distribution function of an uncertain parameter, sliced-normal [134, 135] or sliced-exponential [136] approaches can be utilised to derive probability distributions of multivariate data sets, interval predictor models are able to capture reliable bounds on a data set when information is limited [137, 138] which can also be combined with interval neural networks [139]. A framework for uncertainty quantification with limited information is given in [140]. Other works use operator norm theory to reliably determine first passage problems under imprecise loads [2, 49, 141].

Some approaches to estimate the PSD functions that account for uncertainties in the data have already been presented. For example, in [77, 105] the problem of missing data is addressed. These missing data are reconstructed and assumed to be normally distributed. The probability distributions of the reconstructed missing data are then propagated through the discrete Fourier transform to quantify the uncertainties in the frequency domain. In [109], a large set of accelerograms is used to determine interval parameters for a semi-empirical PSD function. Thus,

different representations of the PSD functions result, depending on the bounds used for the derived interval parameters. A relaxed PSD function, based on a large data set of similar signals transformed into the frequency domain, is derived in [127]. Since it is possible to extract robust statistical information from a large amount of data, the relaxed PSD provides a probabilistic representation of the data in the frequency domain. Although these are different approaches, they all have in common that the PSD functions are not treated as purely deterministic and discrete-valued functions, as it is usually the case.

In this work, specifically uncertainties that stem from a limited amount of available data are considered. If not sufficient data are available, the actual underlying PSD function cannot be estimated with certainty from the data records. Commonly used estimators of the PSD function, such as the periodogram, could lead to a highly unrepresentative model under scarce data, so that the simulation results may not reflect the actual response behaviour of the system under investigation.

Since reliable statistical information can not be derived from a small amount of data, this paper proposes an interval approach to define optimal bounds without considering the distribution within these bounds. The estimation of the proposed imprecise PSD is carried out entirely in the frequency domain, using a radial basis function (RBF) network [142] in order to approximate a basis power spectrum and to obtain basis functions representing such basis power spectrum. The individual weights of the basis functions will be optimised to obtain reasonable bounds considering the actual minimum and maximum of the data set. These bounds reflect the physics of the data and also account for dependencies between frequency components. The premise for this approach is data similarity. A method for determining the spectral similarity for such a data set is given in [131]. To illustrate the strength of the imprecise PSD, different data sets are utilised to derive optimal bounds for those. In particular, two artificially generated data sets are utilised and one estimated from real earthquake ground motions is used to show the feasibility of this approach for real world cases.

This paper is structured as follows: A brief overview of PSD estimation, stochastic processes and RBF networks is given in Section 4.2. The proposed imprecise PSD model is described in Section 4.3. This approach is illustrated by means of two academic examples in Section 4.4 and a set of real data records in Section 4.5. The paper concludes with Section 4.6.

4.2 Preliminaries

This section introduces some basic theoretical concepts that are relevant for the derivation and understanding of the imprecise PSD model introduced later in this work.

4.2.1 PSD estimation and stochastic processes

A stochastic process is affected by random occurrences. Therefore it cannot be described in a purely deterministic way, but has to be modelled as a stochastic process. The resulting stochastic

process at any time is determined by random variables, see e.g., [11].

If no data are available or if the data do not meet the requirements for the simulation, artificially generated stochastic processes can be used for the simulations as an approximation to real stochastic processes. Such a process can be generated using the Spectral Representation Method (SRM) [84]. SRM requires an analytical or empirical function of a PSD S_X to construct a stochastic process X_t with their underlying characteristics. SRM reads as follows

$$X_t = \sum_{n=0}^{N_\omega-1} \sqrt{4S_X(\omega_n) \Delta\omega} \cos(\omega_n t + \varphi_n), \quad (4.1)$$

where

$$\omega_n = n\Delta\omega, \quad n = 0, 1, 2, \dots, N_\omega - 1, \quad (4.2)$$

with N_ω as the total number of frequency points considered in the analysis ω_n as the frequency vector, $\Delta\omega$ as frequency step size, φ_n as uniformly distributed random phase angles in the range $[0, 2\pi]$ and t as time coordinate. Note that $\Delta\omega$ and N_ω are selected according to the properties of the problem at hand. For instance, the frequency step size can be defined as $\Delta\omega = 2\pi/T$, with T as total length of the record, and the number of frequency points N_ω can be chosen according to a cut-off frequency around 99% or more of the total energy of the PSD function [84]. This provides a suitable method for generating compatible time signals derived from and carrying the characteristics of the underlying PSD function S_X .

The estimation of the PSD function of a stationary stochastic process can be obtained by the periodogram [12, 52], which is formed by the squared absolute value of the discrete Fourier transform of the signal $x(t)$. The periodogram reads as follows

$$\hat{S}_X(\omega_k) = \frac{1}{N_t} \left| \sum_{j=0}^{N_t-1} x(j) e^{-\frac{i2\pi}{N_t} kj} \right|^2, \quad (4.3)$$

where N_t is the total number of data points in the time record, $x(j)$ represents the value of the time signal at the j -th time instant, where $j = 0, \dots, N_t - 1$, i is the imaginary unit and k is the integer frequency for $\omega_k = \frac{2\pi k}{T}$ with T as the total length of the record.

However, the periodogram is considered a poor estimator for PSD functions because it may exhibit a high variation in the frequency domain. Even small perturbations or noise in the data can lead to a high variability in the estimated PSDs, which does not correspond to reality. An alternative approach is Welch's method [65]. It is based on forming overlapping segments of the time signal and uses a periodogram modified via a window function to estimate the PSD. The individual estimates are then averaged to obtain a smoother PSD function in trade-off to a lower resolution in the frequency domain.

In Welch's method, the signal $x(t)$ is divided into K segments, such that $x_1(t) = x(t^*)$, $x_2(t) = x(t^* + D)$, \dots , $x_K(t) = x(t^* + (K - 1)D)$ with $t^* = 0, 1, \dots, L - 1$, L as the length of the individual segments and D as a parameter that determines the spacing for the starting points

of the segments, respectively. It is noted that D determines the degree of overlap between the segments. For example, when $D = L/2$, there is a 50% of overlap. Each segment is multiplied by a window function $W(t^*)$ before the modified periodograms are calculated as:

$$P_k(\omega_m) = \frac{1}{L} \left| \sum_{t^*=0}^{L-1} x_k(t^*) W(t^*) e^{-2\pi i m t^* / L} \right|^2 \quad (4.4)$$

with $k = 1, \dots, K$ and ω_m analogous to ω_k in Eq. 4.3. The resulting modified periodograms are averaged to obtain the estimated smoother PSD function.

$$\hat{S}_x^W(\omega_m) = \frac{1}{K} \sum_{k=1}^K P_k(\omega_m) \quad (4.5)$$

The selection of the window function can be chosen according to the PSD estimation requirements. Two window functions are suggested in [65], which are

$$W_1(j) = 1 - \left(\frac{j - \frac{L-1}{2}}{\frac{L+1}{2}} \right)^2 \quad (4.6)$$

and

$$W_1(j) = 1 - \left| \frac{j - \frac{L-1}{2}}{\frac{L+1}{2}} \right|, \quad (4.7)$$

with $j = 0, 1, \dots, L - 1$. Since both window functions ensure that the values in the middle of the signal segment are weighted more heavily than the outer values. This results in further smoothing of the data through the estimation process.

4.2.2 Radial basis function networks

A radial basis function (RBF) network is a class of artificial neural networks [142]. It typically consists of three layers, namely the input layer, the hidden layer and the output layer. It is used to interpolate or approximate functions from a given (and possibly multidimensional) input space to the scalar output space but can be extended to a multi-output network. Thus, in this work the RBF network is a mapping of $y : \mathbb{R}^{N_\omega} \rightarrow \mathbb{R}$.

The input layer of an RBF network passes the input data to the hidden layer. The hidden layer consists of a number of N_B neurons whose activation functions are RBFs, which are characterised by the fact that they are symmetrical around their assigned centre c_i . In this work, the RBF

$$\varphi_i(x) = e^{-(\|x - c_i\| \cdot b_{\varphi_i})^2} \quad (4.8)$$

is used, where $\|x - c_i\| \cdot b_{\varphi_i}$ describes the Euclidean distance from the input x to the designated centre c_i multiplied with a scale factor $b_{\varphi_i} = \sqrt{-\log(0.5)}/s_B$, where s_B denotes the basis function spread.

The function values of the RBFs based on the input data are propagated to the output layer,

where a weighted linear combination of all neurons takes place. The weights w_i of all neurons can be determined with a linear least squares method. In addition, to manipulate the sensitivity of a neuron, a bias b_0 can be employed. Thus, the RBF network results in

$$y(x) = \sum_{i=1}^{N_B} w_i \varphi_i(\|x - c_i\| \cdot b_{\varphi_i}) + b_0 \quad x \in \mathbb{R}^{N_\omega}. \quad (4.9)$$

For an exact interpolation of a function, the number of RBFs N_B must be equal to the number of data points N_ω . In general, however, exact function interpolation is not necessary. Often, the input data are noisy. Therefore, it is advisable to approximate a smoother function and thus average out the noise. In addition, for an exact interpolation the number of neurons can be prohibitively high, which leads to a significantly higher computational effort. In the case of an approximation, the number of RBFs N_B is usually less than the number of data points N_ω . For more information on RBF networks, such as training and validation of the network, the reader is referred to [143–146] and the references therein.

4.3 Method development

For robust simulation results considering uncertainties introduced by the limited number of available data and the PSD estimation processes in general, it is proposed to derive an imprecise PSD function, i.e., an interval-valued PSD function determined by an optimal upper and lower bound with respect to the data set used and parameters chosen. The estimation process is carried out entirely in the frequency domain. The data, for example earthquake ground motions, are usually given in the time domain. After transforming these data into the frequency domain, an ensemble of PSD functions is obtained. Based on such an ensemble, the imprecise PSD function can be derived performing the following steps:

1. Identification of the basis power spectrum of the ensemble
2. Fitting an RBF network to the basis power spectrum
3. Optimisation of the weights of the basis functions with a constrained optimisation to obtain reasonable, data enclosing bounds

These steps will be discussed in the subsequent sections in details.

4.3.1 Basis power spectrum

The basis power spectrum $S_{basis}(\omega_n)$ can be identified using different approaches. As the imprecise PSD function delivers an upper and lower bound regardless of any distribution of the data within those bounds, the mean spectrum or the midpoint spectrum are reasonable choices for

the basis power spectrum. The mean spectrum can be obtained by

$$S_{mean}(\omega_n) = \frac{1}{R} \sum_{i=1}^R S^{(i)}(\omega_n), \quad (4.10)$$

where the superscript indicates the i 'th PSD function in the ensemble and R is the cardinality of the ensemble, i.e. the total number of PSD functions. The midpoint spectrum can be obtained by computing the midpoint between maximum and minimum values of the ensemble, i.e. the vector consisting of all minimum values of the ensemble is defined as $S_{min}(\omega_n) = \min(S^{(i)}(\omega_n)) \forall i \in R$ and accordingly the vector of all maximum values is $S_{max}(\omega_n) = \max(S^{(i)}(\omega_n)) \forall i \in R$, such that

$$S_{midpoint}(\omega_n) = \frac{1}{2} (S_{max}(\omega_n) + S_{min}(\omega_n)). \quad (4.11)$$

If the PSD functions are relatively evenly distributed between the maximum and minimum values, the midpoint spectrum can be useful. If the data is unevenly distributed, the mean spectrum may be a better choice, as it will draw the basis power spectrum towards the direction of the majority of PSD functions.

4.3.2 Fitting an RBF network

To fit the RBF network to the basis power spectrum $S_{basis}(\omega_n)$, the hyperparameters N_B , the number of basis functions, as well as s_B , the basis function spreads, are required. For an exact interpolation of the basis power spectrum $S_{basis}(\omega_n)$, it is required to use as many basis functions (i.e., neurons in the RBF network) as frequency points in the ensemble. As such a representation will often yield in a highly spiky power spectrum and the subsequent optimisation of the bounds will yield in the minimum and maximum value of the ensemble at each frequency, it is advisable to choose a lower number of basis functions. This will result in a smoother approximation for $S_{basis}(\omega_n)$. However, the objective of this work is to find optimal bounds rather than an exact interpolation. Since an exact interpolation is not feasible due to the poor scaling of interval propagation schemes in terms of dimensionality, optimal bounds with a significantly reduced number of basis functions compared to frequency points in the PSD are sought.

The choice of the hyperparameter N_B and s_B is crucial. It must be kept in mind, that the choice of the hyperparameters will also affect the subsequent optimisation of bounds. This can result in the bounds of the imprecise PSD may being too wide or too narrow and therefore not correspond to the actual data set or the constraints of the optimisation are violated. An unfavourable choice of these hyperparameters can lead to unreliable results and will falsify the subsequent simulation analysis. Furthermore, if a low N_B is chosen, the RBF network operates as a smoother for its realisations.

There are several approaches in the literature to find a set of optimal hyperparameters, such as pruning methods, see e.g., [144, 145, 147] and references therein. Since the fitting of the RBF network is followed by the optimisation of the boundaries, the problem here is somewhat more

complex. Later in this work, it will be discussed that finding good parameters is not a trivial task considering the subsequent optimisation of the bounds. Finding appropriate parameters can be challenging, but defining these parameters is crucial for deriving optimal bounds. This section only presents the proposed idea of how to derive these optimal bounds for two examples with predefined hyperparameters. In Section 4.4.2 the influence of different hyperparameters on the resulting bounds is discussed and in Section 4.3.4 an optimisation of the hyperparameters is suggested.

4.3.3 Obtaining optimised bounds

The derivation of optimal bounds is done by optimising the weights calculated via the fitting of the RBF network. This requires the definition of an optimal weight $w^{up} \in \mathbb{R}^{N_w}$ and $w^{low} \in \mathbb{R}^{N_w}$ for the upper and lower bounds, respectively, as optimisation parameters that control the sensitivity of the respective basis functions and thus the distance between the basis power spectrum S_{basis} and the upper and lower bounds, respectively.

For the calculation of an upper and lower bound, the term from the RBF network (Eq. 4.9) must be adapted for the following optimisation problem. The upper bound thus results in

$$\overline{S_{opt}}(\omega_n; w^{up}) = \sum_{i=1}^m w_i^{up} \varphi_i + b_0 \quad (4.12)$$

and the lower bound is

$$\underline{S_{opt}}(\omega_n; w^{low}) = \sum_{i=1}^m w_i^{low} \varphi_i + b_0 \quad (4.13)$$

with ω_n and n as defined in Eq. 4.2. The basis functions φ_i and the bias b_0 including the spread s_B result from fitting the RBF network to the basis power spectrum S_{basis} , similarly for the weights w which are the initial values for $w^{up} = w$ and $w^{low} = w$. This leads to a total number of parameters to be optimised of $w^{up} + w^{low} = 2N_B$.

To ensure that representative and optimal bounds are derived for the data set, the norm of the difference between the upper and lower bound will be the objective function for the optimisation. This optimisation is subject to the condition, such that the resulting upper bound shall be larger than the maximum of the ensemble and the resulting lower bound shall be smaller than the minimum of the ensemble to ensure that all data points are included in the bounds. For physical reasons the lower bound must not be smaller than 0 by all means as negative values are not possible in terms of power spectral densities. The optimisation problem results as follows

$$\begin{aligned} \min \quad & \sum_{\omega_n} \left| \overline{S_{opt}}(\omega_n; w^{up}) - \underline{S_{opt}}(\omega_n; w^{low}) \right| \\ \text{s.t.} \quad & \overline{S_{opt}}(\omega_n; w^{up}) \geq S_{max}(\omega_n) \\ & \underline{S_{opt}}(\omega_n; w^{low}) \leq S_{min}(\omega_n) \\ & \underline{S_{opt}}(\omega_n; w^{low}) \geq 0 \end{aligned} \quad (4.14)$$

for $n = 1, \dots, N_\omega$. If the weights w^{up} and w^{low} are optimised, reasonable bounds can be provided.

4.3.4 Optimisation of the hyperparameter

In general, it may be a difficult task to find the optimal hyperparameters manually. Therefore, it seems natural to leave the choice of the hyperparameters N_B and s_B to an optimisation. Since the hyperparameters also influence the subsequent optimisation of the bounds, a nested optimisation must be carried out. This means that the hyperparameters are determined in an outer optimisation, while the bounds are defined in a nested inner optimisation. A study of different optimisation algorithms has proven that the best results are obtained with a Bayesian optimisation for the hyperparameters and a non-linear constrained optimisation for the bounds. However, different problems arise. Finding optimal hyperparameters is no trivial task as multiple local minima exist, which makes it even for advanced algorithms difficult to find the global optimum. Moreover, the number of basis functions is an integer value, which is a challenge in optimisation problems in general, see for instance [148]. In addition, a large number of basis functions N_B leads to better results, as already confirmed by the results in the previous section, which is why the optimisation for both parameters tends towards a higher number of basis functions.

Since the number of basis functions is also decisive for a simulation following the optimisation of the bounds, e.g. an interval propagation as part of a reliability analysis, it is desirable to obtain a lower number of these. Since it makes sense, especially with regard to interval propagation, for the analyst to have control over the number of basis functions and since optimising an integer value is difficult, it is suggested to predefine a feasible number of basis functions N_B and optimise only the parameter s_B . In this way, control over the trade-off, more basis functions for better results, less basis functions for better a interval propagation, is left to the analyst.

4.4 Academic examples

This section illustrates the derivation of the imprecise PSD with two academic examples. Although two specific examples are used in this case, it should be noted that in general any PSD function can be employed. Therefore, this choice of PSD functions does not affect the general nature of the approach. Note that in these examples, most physical units are omitted, as they have no effect for the purpose of illustrating the application of the proposed approach.

The first PSD function utilised is the Kanai-Tajimi PSD function of the form

$$S_1(\omega) = S_0 \frac{1 + 4\xi^2 \frac{\omega^2}{\omega_p^2}}{\left(1 - \frac{\omega^2}{\omega_p^2}\right)^2 + 4\xi^2 \frac{\omega^2}{\omega_p^2}} \quad (4.15)$$

is utilised in this section and throughout this work. In this equation, $S_0 = 0.25$ is a constant, $\omega_p = 3\pi$ describes the peak frequency and $\xi = 0.5$ indicates the sharpness of the peak [59, 86].

Furthermore, the upper cut-off frequency is defined to be $\omega_u = 50$ rad/s. For verification, a second PSD function is utilised, which is given in [84].

$$S_2(\omega) = \frac{1}{4}\sigma^2 b^3 \omega^2 e^{-b|\omega|} \quad (4.16)$$

In this PSD, the parameter $\sigma = 1$ is the standard deviation of the underlying stochastic process and $b = 1$ is a parameter proportional to the correlation distance of said stochastic process [84, 85]. The upper cut-off frequency is $\omega_u = 12.5$ rad/s.

For both PSD functions, three time signals were generated using SRM shown in Eq. 4.1, which were then transformed back into the frequency domain using the periodogram as in Eq. 4.3, to generate two data sets for the subsequent derivation of the imprecise PSD. Due to the influence of the random variables in SRM and the poor estimation quality of the periodogram, these data reflect a certain randomness and to a certain extent have the character of real data. Both data sets, or so-called ensembles, are depicted in Fig. 4.1. The ensembles utilised aim to illustrate the capabilities of the imprecise PSD in dealing with different forms of datasets, such as a more tight and a more variant dataset, i.e. with low and high spectral variance, respectively. The respective ensembles will be called ensemble A and ensemble B throughout this work.

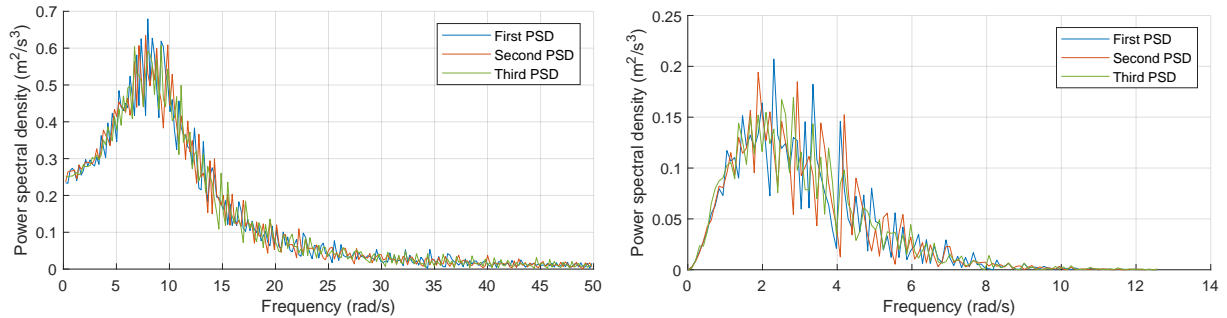


Figure 4.1: Ensemble A and ensemble B consisting of 3 PSD functions each, utilised to estimate the imprecise PSD function.

For the illustration of the estimation of the imprecise PSD, in this work the midpoint spectrum is utilised for establishing the basis power spectrum S_{basis} .

4.4.1 Estimation of an imprecise PSD function

Since this section aims to illustrate the approach in a comprehensible way by means of examples, the optimisation of the hyperparameters will be omitted. Instead, a predefined number of basis functions and spread for both examples are used.

For ensemble A the number of frequency points is $N_\omega = 238$. The number of basic functions has been chosen to be $N_B = 10$ with a spread of $s_B = 3.8$. Ensemble B consists of $N_\omega = 121$ frequency points. $N_B = 5$ and $s_B = 2$ are the predefined parameters here. For both ensembles, the weighted basis functions derived via the RBF network are shown in Fig. 4.2. In addition, the calculated basis spectra (target) and the basis spectra approximated via the basis functions

(output) are given.

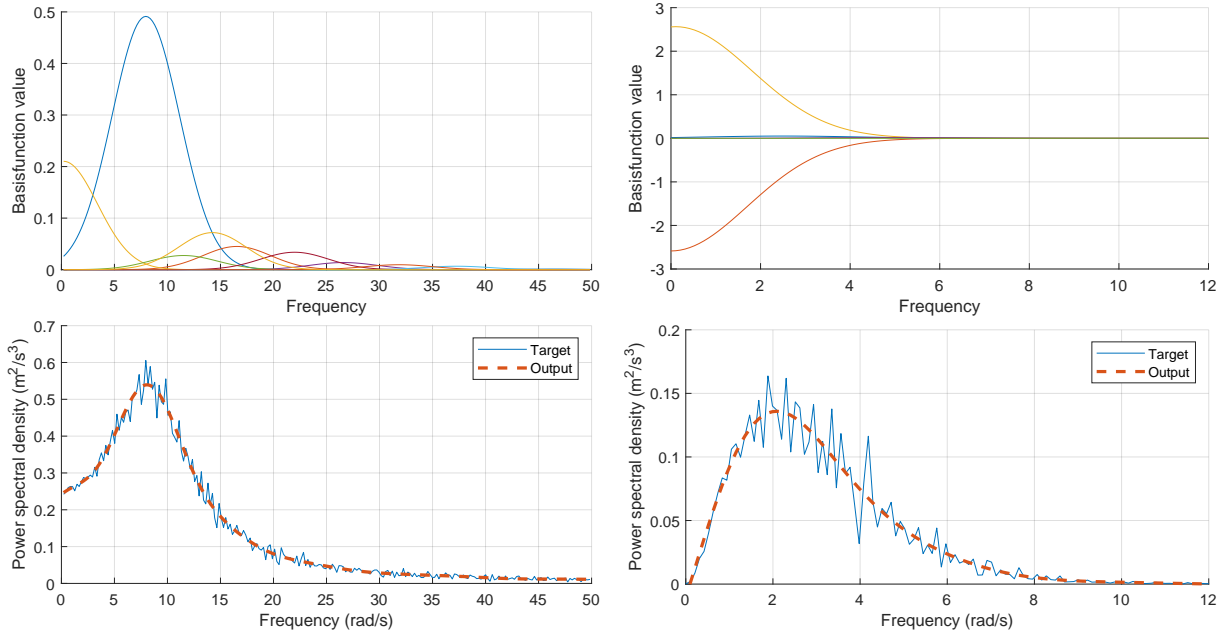


Figure 4.2: Weighted basis functions used to approximate the basis power spectrum.

The imprecise PSD function for the ensembles given in Fig. 4.1 are shown in Fig. 4.3. The final objective value for ensemble A is 1.825 and for ensemble B 0.885. These objective values will be of importance for section 4.4.2, where the influence of the hyperparameters on the resulting bounds is investigated.

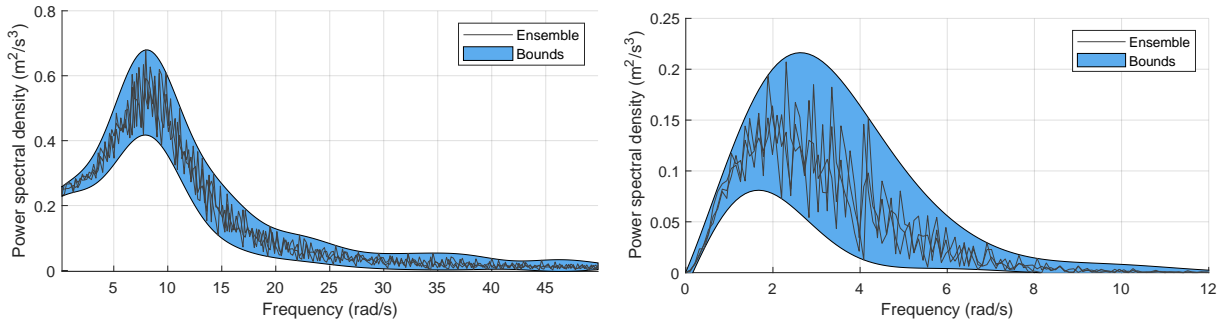


Figure 4.3: Bounds of the estimated imprecise PSD function.

In some cases, PSD values may lie (by a small margin) outside the optimised bounds or the bounds may take values smaller than 0 and thus violate the optimisation constraints. This can occur because the basis functions and their optimised weights are not able to capture all values, so the optimisation problem can be too inflexible. Values smaller than zero are an artefact of the optimisation. However, both violations of the optimisation constraints are justifiable, since the general form of the PSD and the underlying physics were nevertheless captured very well. Moreover, these violations often only occur at low PSD values, whose influence is of minor importance. However, this is more of an implementation problem than a theoretical issue.

4.4.2 Influence of the hyperparameter on the optimised bounds

The optimisation of the bounds has been carried out in a brute-force manner for both ensembles. This section aims to evaluate the influence of the hyperparameters and to show how complex finding optimal parameters can be. For each possible parameter combination of N_B and s_B , the optimal bounds were calculated. N_B was run over the values 3 to 50, while s_B was run from 0.01 to 10 with increment 0.01, yielding a total of 43,248 optimisations for each ensemble. The resulting objective values for ensemble A are depicted in Fig. 4.4, while those for ensemble B can be obtained in Fig. 4.5. In both figures, the colour scale is adjusted to reasonable objective values, i.e. the optimised bounds with such an objective value are considered acceptable.

The figures show that there are many local minima, which complicates finding suitable parameters. A higher number of basis functions often leads to better results, which seems to be reasonable because with a high number of basis functions the signal can be better captured. This is clearly reflected in the figures. However, since a lower number of basis functions is desirable in the propagation of intervals, this is in contradiction to each other. Therefore, the aim must be to find favourable parameters, under the condition that the number of basis functions does not become too high while still maintaining an acceptable objective value for the optimised bounds. Although this is of course case-dependent and influenced by the shape of the input data, it can be reasonably concluded that the optimal trade-off here is around 15 basis functions and a basis function spread of 3-4 for ensemble A, and around 6-8 neurons and neuron spread of 2-3 for ensemble B.

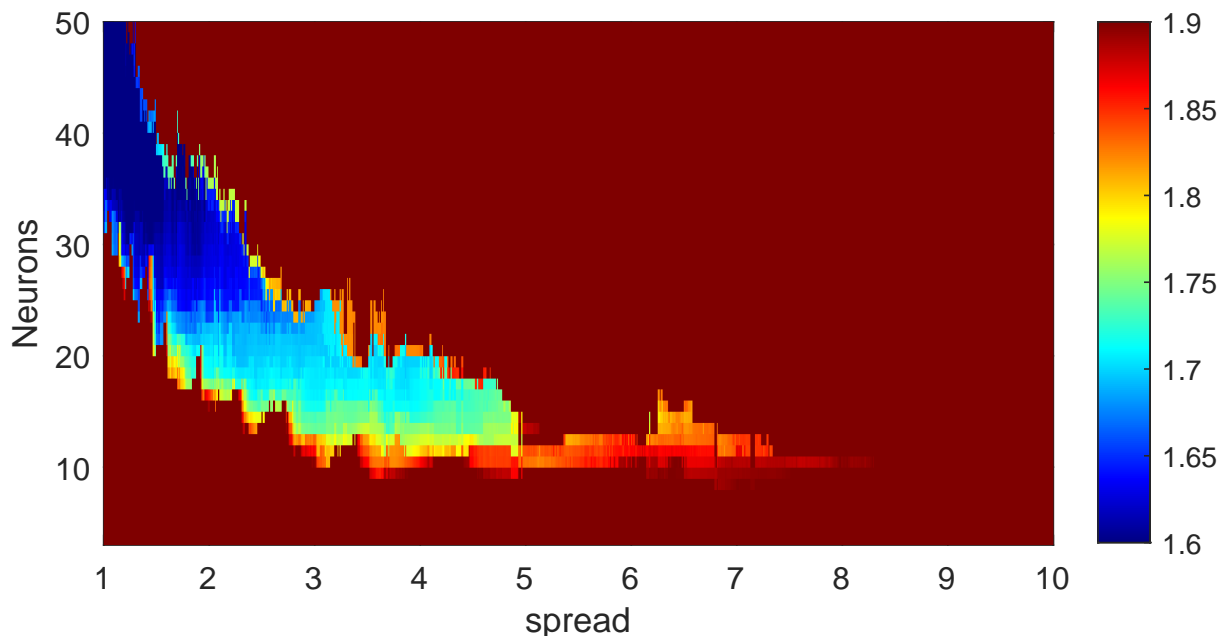


Figure 4.4: Objective values for ensemble A.

However, it can be concluded here that it is a highly complicated task and challenging for the analyst to find optimal parameters, which motivates to incorporate the optimisation for

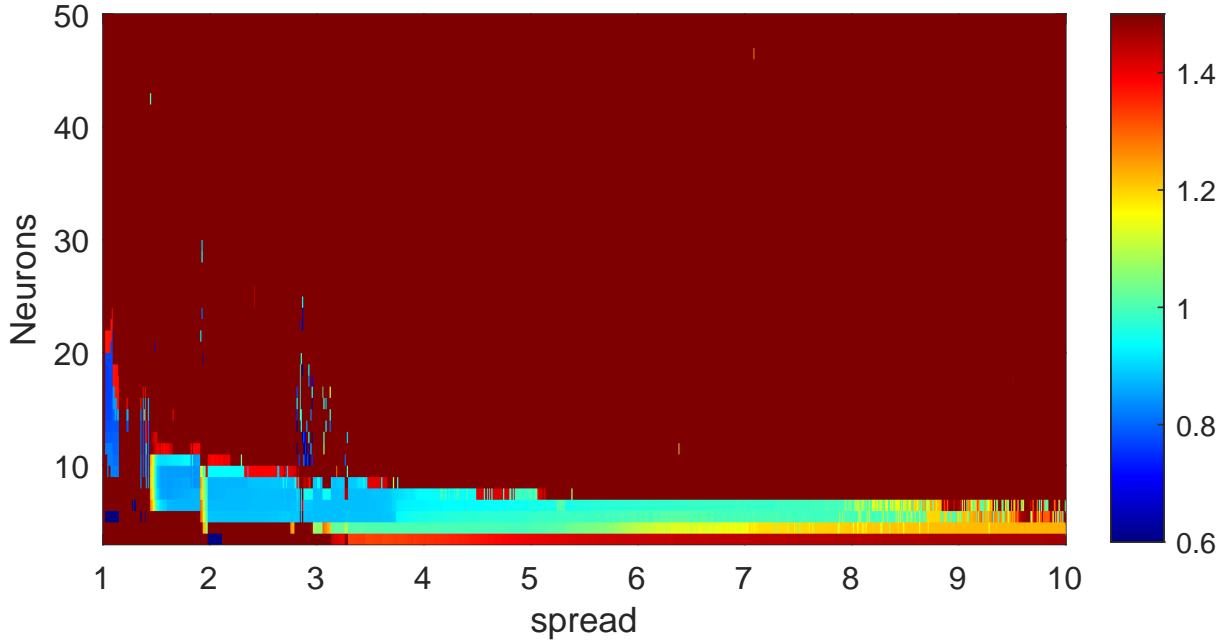


Figure 4.5: Objective values for ensemble B.

identifying the parameters such as described in Section 4.3.4.

4.5 Optimising the bounds of real data

In order to show the derivation of optimal bounds not only for academic examples, but also to demonstrate its applicability to a real case, the proposed method is applied to a real data set in this section. The data used here come from the PEER database [112, 149] and are the records of the El Centro earthquake on 18 May 1940. Under the premise of this work that the proposed method is in particular useful for limited data, only two time signals are used, recorded in north-south and east-west direction, which are shown in Fig. 4.6. The signals have a total length of $T = 53.46$ s and a time discretisation of $\Delta t = 0.02$ s. At this point, only the stationary PSD function is estimated from the given earthquake data as a simplification. Nevertheless, it should be noted that an earthquake always has a non-stationary character and an estimate of the evolutionary PSD function taking into account the time-frequency resolution provides a more realistic representation.

The two time signals are transformed into the frequency domain using the periodogram (Eq. 4.3), which leads to the PSD functions given in Fig. 4.7. Using this data set for optimisation poses some problems. The data set shows a high spectral variance, a problem that arises from the use of the periodogram, as already mentioned in Section 4.2. Due to the high variance, many PSD values, including those near the peak frequencies, are close to zero, which poses a challenge for the proposed method. This problem can be solved with a more suitable estimator. Instead of the periodogram, Welch's method (Eqs. 4.4 and 4.5) can be used, which usually leads to smoother results by averaging and windowing the input signals. This can be appreciated in Fig. 4.8.

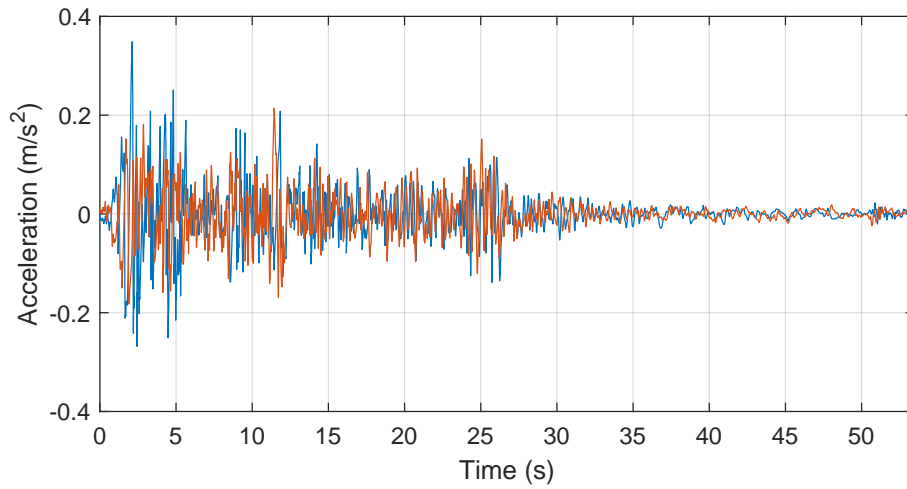


Figure 4.6: Time records of the El Centro Earthquake.

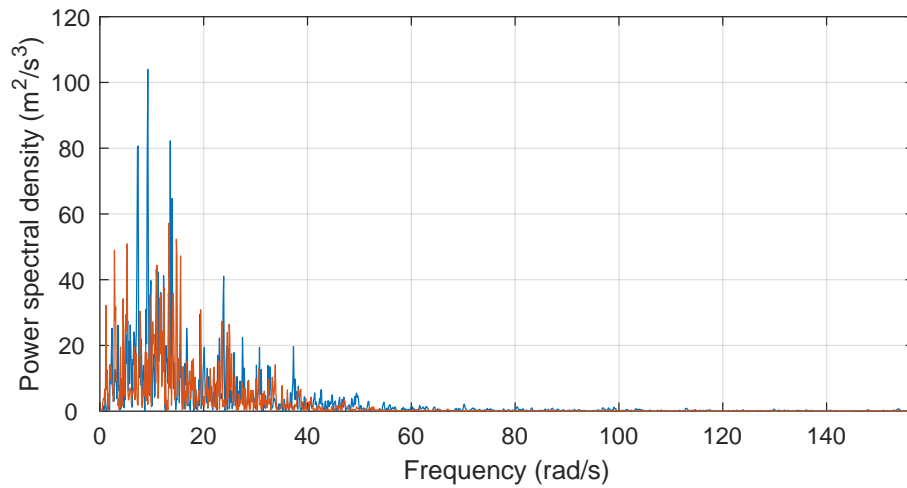


Figure 4.7: El Centro earthquake records transformed to frequency domain with the periodogram.

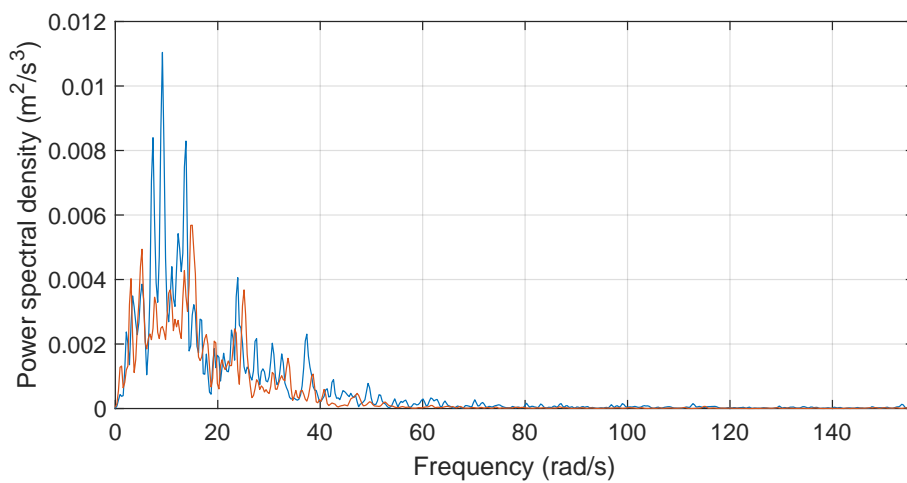


Figure 4.8: El Centro earthquake records transformed to frequency domain with Welch's method.

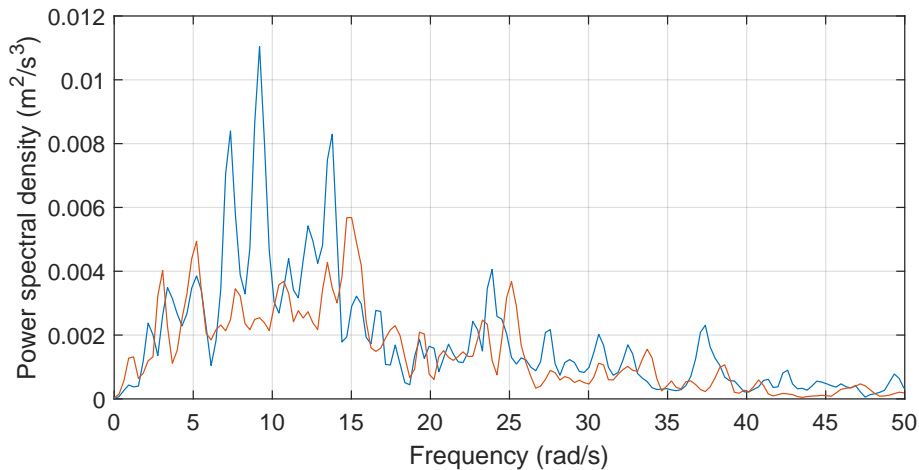


Figure 4.9: El Centro earthquake records transformed to frequency domain with Welch’s method and definition of suitable cut-off frequency.

Another problem is that although the spectral variance has now been reduced, many values are still close to zero. This is natural as it indicates that the spectral density for high frequencies in this data set are close to zero, a typical pattern for earthquake data. However, since the optimisation of the bounds is problematic at values close to zero, a suitable cut-off frequency ω_U must be chosen. This can reasonably be done since high frequencies with low spectral densities have a negligible small effect on the simulation results anyway. In [84], it is suggested that ω_U be chosen such that 99% of the total energy is still contained in the PSD function. Here, the cut-off frequency is set at 95% of the total energy, which is $\omega_U = 50$ rad/s, because, as it can be seen in Fig. 4.8, a very large amount of frequency components are close to zero.

After appropriate pre-processing of the data set, the bounds can be optimised according to the proposed approach. As described earlier, the analyst has control over the number of basis functions N_B , so the optimisation of the bounds was performed for $N_B \in \{8, 10, 15, 20\}$. The resulting bounds can be seen in Fig. 4.10. From the optimised bounds it can be seen that the more basis functions are used, the smaller the area between the upper and lower bounds becomes. Further, the bounds are more data-enclosing for a higher number of basis functions. This behaviour can easily be explained by the fact that, as mentioned before, a high number of basis functions is better able to capture the signal. Nevertheless, a high number of basis functions is not always useful in terms of interval propagation, so it is reasonable to obtain a higher area between upper and lower bound in exchange to a lower number of basis functions. For comparison, the corresponding optimised spreads s_B and objective values for the optimised bounds are given in Table 4.1.

To draw the reader’s attention to the importance of selecting suitable parameters and the pre-processing of the data, two counter-examples are given here. If the number of basis functions and their spread are incorrectly selected or the pre-processing of the data set was not done thoroughly enough, the optimised bounds can lead to a highly unrepresentative results. The optimisation

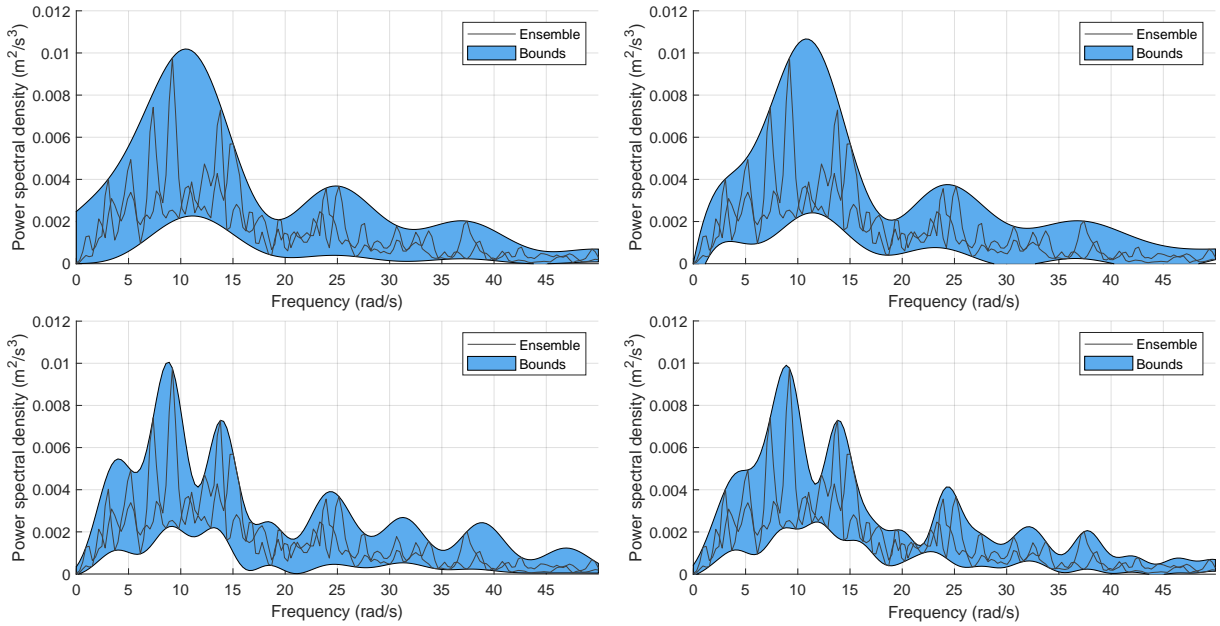


Figure 4.10: Optimised bounds of the real data set for $N_B \in \{8, 10, 15, 20\}$ basis functions.

Table 4.1: Optimised spread and resulting objective value depending on the number of basis functions.

N_B	s_B	Objective value
8	4.9141	0.0468
10	6.7423	0.0459
15	2.6417	0.0383
20	1.7656	0.0354

was carried out for the pre-processed data set but with a bad combination of parameters, i.e. $N_B = 15$ and $s_B = 0.5375$. The resulting bounds are given in Fig. 4.11 (left). For the second counter-examples the optimisation was carried out for the ensemble given in Fig. 4.7, where the signals were transformed to frequency domain using the periodogram 4.3. The counter-example is depicted in Fig. 4.11 (right). No further discussion is required to prove that such bounds do not reflect the data set. Therefore, these bounds are not acceptable and cannot be used for a subsequent simulation.

4.6 Conclusions

Accounting for uncertainties in data sets to obtain reliable simulation results is of paramount importance in engineering. Especially when only limited data are available, uncertainties can have a large impact on the results and can easily lead to wrong conclusions. This may result in disastrous consequences, e.g., when an actually catastrophic result is shifted into an acceptable range due to incorrect consideration of uncertainties. In such a case, it is important to correctly interpret the data and to quantify uncertainties rigorously. For the generation of appropriate load models it is important to account for those uncertainties. From a large amount of data, it

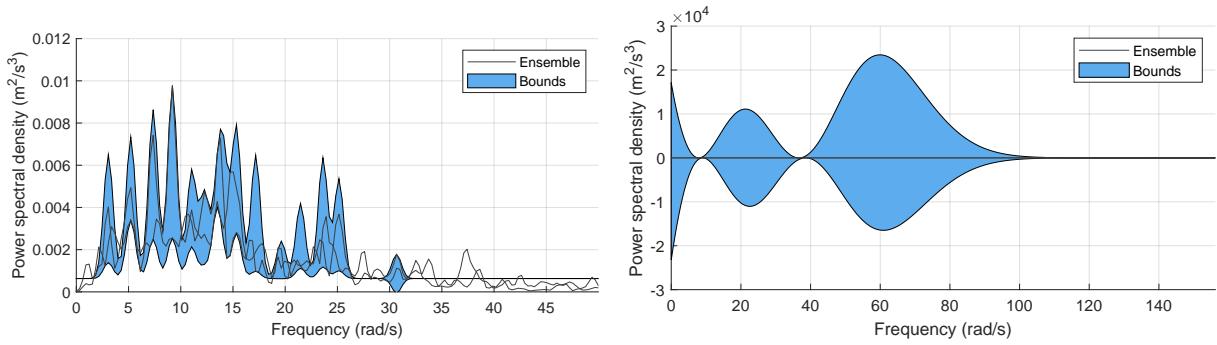


Figure 4.11: Counter-examples: a poor choice of hyperparameters (left) and a poorly pre-processed data set (right).

is often possible to derive a robust model that provides reliable simulation results. However, as this is often not possible from limited data, an imprecise model of a PSD function is proposed in this paper, which provides optimal bounds of the data set. Moreover, by using an RBF network, the physics of the underlying stochastic process is reflected and dependencies between frequencies are taken into account. The resulting basis functions of the RBF network are used to optimise the weights to obtain an upper and lower bound for the data set. One advantage of this approach is that no assumptions have to be made about the distribution of the data within the bounds, as this would be difficult in any case due to the limited data. Another advantage is that the choice of the number of basis functions is left to the analyst, which is particularly important for the propagation of intervals in the context of a reliability analysis. This also allows some flexibility in modelling the bounds. An important aspect is the pre-processing of the data, as the method may not yield acceptable results for high variant data or in case many spectral density values are close to zero. Adequate pre-processing of the data is therefore essential. The proposed approach was not only elaborated using academic examples, but its applicability to real data was also demonstrated. The imprecise PSD presented here is able to obtain optimal bounds and is thus suitable for quantifying uncertainties due to a limited amount data. This work only refers to the derivation of the optimised bounds for a data set consisting of only a few data records. Future works will address the robust propagation of the bounds in the context of reliability analyses.

5 | Development of a relaxed stationary power spectrum using imprecise probabilities with application to high-rise buildings

The following is an earlier development of the imprecise PSD described in Chapter 4 and addresses the same problem of capturing uncertainties in a limited data set. Also in this case, it is not possible to obtain reliable information from the data set because there is insufficient data for a probabilistic representation. An interval or fuzzy-based approach is seen as the remedy for this problem.

The basic idea is to generate a Fuzzy PSD function from an ensemble of PSD functions so that the spectral density at each individual frequency is represented by a Fuzzy number. First, a basis power spectrum is determined, which can be the mean power spectrum, for instance. Next, confidence intervals of each frequency are calculated at different confidence levels, which are then subjectively assigned to the membership function of the Fuzzy number. Since the basis power spectrum is identified to be the most reliable spectrum, the membership value 1 is assigned to it. The remaining confidence intervals are correspondingly assigned to lower membership values. This results in a PSD function where the spectral density value at each individual frequency is translated to a Fuzzy number.

The PSD model presented is validated by means of numerical examples. It is shown that the generation of such a model has the advantage that the epistemic uncertainties are quantified. This is especially important for the use in the context of reliability analyses, since only small uncertainties can lead to a wrong interpretation of the results. A load model that quantifies these uncertainties in terms of Fuzzy numbers has significant advantages compared to a discrete-valued PSD function. Because the system response can also be given as Fuzzy numbers, a range of possible system responses can be calculated instead of a discrete response. This approach allows critical system behaviour to be determined more reliably, taking these uncertainties into account.

Development of a Relaxed Stationary Power Spectrum using Imprecise Probabilities with Application to High-rise Buildings

Marco Behrendt^{a,b}, Liam Comerford^b, Michael Beer^{a,b,c}

^aInstitute for Risk and Reliability, Leibniz Universität Hannover, Germany

^bInstitute for Risk and Uncertainty, University of Liverpool, United Kingdom

^cDepartment of Structural Engineering & Shanghai Institute for Disaster Prevention and Relief, Tongji University, Shanghai, P.R. China

Published in *2019 IEEE Symposium Series on Computational Intelligence (SSCI)* on 20 February 2020

Abstract

Modern approaches to solve dynamic problems, where random vibrations are the governing excitations, are in most cases based on the concept of the power spectrum as the core model for the representation of excitation and response processes. This is partly due to the practical applicability of spectral models for frequency domain analysis. In addition, compatible time-domain samples can easily be generated. Such samples can be used for numerical performance evaluation of systems or structures represented by complex non-linear models. The development of spectral estimation methods that use ensemble statistics to generate a single or finite number of deterministic spectra results in spectral models that can be applied directly in structural analysis. However, the properties of the measured environmental process are still lost. In order to produce reliable and realistic power spectra for the application to systems, in most cases not enough real data sets are available. To capture the epistemic uncertainties of the model by taking into account inherent statistical differences that exist across real data sets, an approach for a stochastic representation of the loads can be used. In this work, the epistemic uncertainties in the spectral density of the process are captured by using an interval approach which, in combination with the stochastic nature of the process, leads to an imprecise probability model. From all the available power spectra of the ensemble, one power spectrum is identified on which the resulting relaxed power spectrum is based. To relax the power spectrum, interval parameters are implemented, thereby forming an enveloping boundary for all estimated power spectra. In order to capture the epistemic uncertainties and to present this information effectively, imprecise probabilities are used in this newly developed load representation. The relaxed power spectrum is validated by application to a single-degree-of-freedom system and a multiple-degree-of-freedom system by determining and analysing the response spectra of the systems.

Keywords: Stochastic process, Power spectrum estimation, Relaxed power spectrum, Imprecise probabilities, Fuzzy set theory, Uncertainty quantification, Random vibrations.

5.1 Introduction

Many systems and structures in engineering underlying random vibrations and thus lead to a problem of stochastic dynamics in which, for example, environmental influences have a significant impact. High-rise buildings excited by earthquake and wind loads, or offshore platforms in the ocean excited by waves are typical examples for these type of problems.

To determine the influence of environmental processes on the structures by performing simulations, these influences must be recorded and applied to a model with system excitation/response process [76].

For the analysis of a structure e.g. earthquake ground motions can be recorded and applied. Due to various reasons, uncertainties occur in these data. These can be: limited number of samples; damaged sensors or equipment failures, possibly due to the earthquake itself; sensor threshold limitations; measurement errors. Additionally, for several reasons the sensors may capture the data incorrectly due to sensor maintenance, bandwidth restrictions or data acquisition restrictions. Thus, the recorded data may be of poor quality. For this reason, the real data sets must be adequately represented, and uncertainties reduced as much as possible [106].

As the stochastic dynamics have been investigated very efficiently in the last decades, different models have been developed. One of them is the power spectrum density (PSD), which is widely used in the modelling of stochastic processes, especially in applications such as earthquake, wind and ocean engineering ([52, 54, 55]). In earthquake engineering, the use of PSD can be traced back to Housner [58] or Kanai [59], for example. At least in the linear case considered in this work, an elegant relationship between the power spectra of the input data and the output data can be derived for the system's behaviour [91].

The more real data records are available, the better are the results of the simulations, since they are more accurate for a large amount of data and represent environmental processes better than just a few data. Additionally, the underlying physics of the system should be well understood. Since in reality often insufficient real data records are available and the underlying physics is not well known, other approaches must be found to develop a load model that best represents the data and mitigates uncertainties [106].

This work is focused on the case when only a few real data records are available. Since it is not possible to calculate an average power spectrum which represents the whole ensemble adequately, other approaches must be pursued. Traditional statistical power spectrum estimation could result in a highly unrepresentative model of the process. Such a power spectrum might be too limiting and too restrictive regarding the numerical simulations, the response process would not correspond to the actual behaviour and the inherent uncertainties are not mitigated. In this case an interval approach is used to capture the epistemic uncertainty in the spectral density of the process, which leads to an imprecise probabilistic model. Therefore a basis power spectrum is defined which is then relaxed by implementing intervals such that the resulting model bounds form an envelope. For the calculation of the interval a method according to Beer [150] is used. The proposed method combines confidence intervals and α -level sets from fuzzy set theory [151]. This work is structured as following: An explanation in which form the environmental processes are available and how a power spectrum density is estimated from a stochastic process is presented in Section 5.2. In Section 5.3 it is described how the relaxed power spectrum density is developed using imprecise probabilities. Based on the derived relaxed power spectrum, two numerical examples, one of a single-degree-of-freedom system and one of a multiple-degree-of-freedom system are presented in Section 5.4. The conclusion is given in Section 5.5.

5.2 Stochastic Process Representation and Power Spectrum Estimation

This section describes the generation of a stochastic process and the transformation of such a process into a power spectral density.

Given a real-valued stationary stochastic process $X(t)$, $-\infty < t < \infty$. For such a process exists a corresponding orthogonal process $Z(\omega)$. Thus, $X(t)$ can than be written in the form

$$X(t) = \int_0^{\infty} e^{i\omega t} dZ(\omega). \quad (5.1)$$

This equation is called the spectral representation of the process $X(t)$, where the process $Z(\omega)$ has the following properties:

$$E [dZ(\omega)] = 0 \quad (5.2)$$

and

$$E \left[\left| dZ^2(\omega) \right| \right] = 4S_X(\omega)d\omega \quad (5.3)$$

with $E(\cdot)$ as expectation operator [11]. In this equation $S_X(\omega)$ describes the two-sided power spectrum of the stationary process $X(t)$. To generate a stationary stochastic process, a model proposed by Shinozuka and Deodatis ([85], [84]) is considered in this work:

$$X(t) = \sum_{n=0}^{N-1} \sqrt{4S_X(\omega_n)\Delta\omega} \cos(\omega_n t + \Phi_n), \quad (5.4)$$

where

$$\begin{aligned} \omega_n &= n\Delta\omega, \quad n = 0, 1, 2, \dots, N-1 \\ \Delta\omega &= \frac{\omega_u}{N} \end{aligned} \quad (5.5)$$

with $N \rightarrow \infty$ and Φ_n as uniformly distributed random phase angles in the range $0 \leq \Phi_n < 2\pi$. The power spectral density

$$S_X(\omega) = \frac{1}{4} \sigma^2 b^3 \omega^2 e^{-b|\omega|}, \quad -\infty < \omega < \infty \quad (5.6)$$

is utilised. In this equation σ is the standard deviation of the stochastic process and b is a parameter proportional to the correlation distance of the stochastic process. An example of a generated stochastic process is depicted in Fig. 5.1 (left). To transform a stochastic process from time domain to frequency domain the periodogram is utilised [12]. This is a frequently used estimator of the power spectrum and can be understood as the squared absolute value of the discrete Fourier transform of the time signal $x(t)$:

$$S_X(\omega_k) = \lim_{T \rightarrow \infty} \frac{2\Delta T}{T} \left| \sum_{t=0}^{T-1} x_t e^{-2\pi i k t / T} \right|^2 \quad (5.7)$$

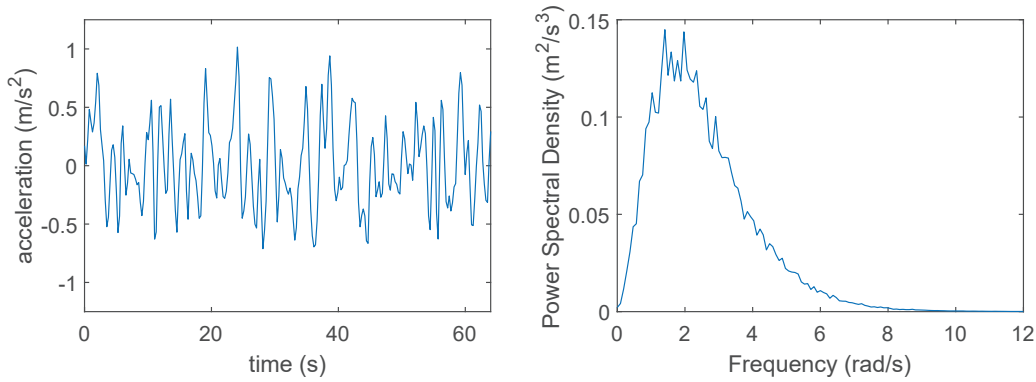


Figure 5.1: Example for a generated stochastic process (left) and the estimated power spectrum (right).

In this equation, T is the number of data points, t describes the data point index in the record, k is the integer frequency for $\omega_k = \frac{2\pi k}{T_0}$ and T_0 is the total length of the record. The stochastic process depicted in Fig. 5.1 (left) is transformed to the power spectrum, which is also depicted in Fig. 5.1 (right).

5.3 Estimation of a Relaxed Power Spectrum

This section discusses the estimation of a relaxed power spectrum from an ensemble of power spectra using a semi-heuristic concept similar to Beer [150] for individual random variables.

This method combines the calculation of confidence intervals with the fuzzy set theory introduced by Zadeh [151] and extensively studied in [152] and [153].

The starting point is the estimation of power spectra from the stochastic processes by Eq. 5.7. Once the power spectra are estimated from the time histories, the calculation of the relaxed power spectrum is based on three steps:

1. Identification of the basis power spectrum of the ensemble,
2. Calculation of the confidence intervals,
3. Assignment of the confidence intervals to the fuzzy membership values.

From the ensemble of power spectra the basis power spectrum is identified. There are different approaches to identify this power spectrum and the choice not limited. In this work the mean of the ensemble of the power spectra is assumed to be the basis power spectrum. Therefore, for each angular frequency the mean of all values for the corresponding frequencies is calculated. Based on additional knowledge about the stochastic process and the underlying physics, other identifications of the basis power spectrum can also be utilised.

Next, the confidence intervals are calculated based on the type of distribution function. The ensemble of power spectra is assumed to be normal distributed in this case. Therefore, additional to the mean m_X , the standard deviation σ of the ensemble is determined. With mean and

standard deviation the confidence interval is calculated by Eq. 5.8 for different α -levels:

$$C = m_X \pm t_{(1-\frac{\alpha}{2};n-1)} \frac{\sigma}{\sqrt{n}}, \quad (5.8)$$

where n is the number of power spectra in the ensemble and $t_{(1-\frac{\alpha}{2};n-1)}$ is the $(1 - \frac{\alpha}{2})$ -quantile of the t-distribution with $n - 1$ degrees of freedom.

As the last step, the confidence levels are assigned to the membership function $\mu(X)$ of the fuzzy set by subjective assignment. Since the mean m_X of the power spectra in this example is identified as the basis power spectrum, and thus the most plausible value, it is assigned to the membership value one. Since the largest confidence level describes the outer bounds of the calculated intervals, these values are almost not plausible. The membership value for the this confidence level is therefore set to zero and thus, the α -level 0.0 specifies bounds for the possible range of power spectra, which leads to an envelope of the ensemble of power spectra. For the remaining confidence levels the assignment should be combined with expert knowledge. If there is no additional information about the underlying process, the assignment can be done accordingly to distances between the individual confidence levels. Thus, the confidence level 0.9 would be assigned to the membership value 0.25, the confidence level 0.75 assigned to membership value 0.5 and confidence level 0.5 to membership value 0.75. Since the confidence level 0.99 and the mean of the ensemble are already assigned to the fuzzy membership values 0 and 1, respectively, all confidence levels are assigned to membership values. This model of a relaxed power spectrum based on confidence intervals and fuzzy numbers can be applied to systems. Since the excitation power spectrum is represented in a relaxed model, the response spectrum of the system will also be a relaxed model, which will result in bounds combined with fuzzy numbers of possible failure of the structure.

5.4 Numerical Example

In this section a numerical example for the calculation of a relaxed power spectrum will be given. Later it will be applied to a single-degree-of-freedom (SDOF) system and a multiple-degree-of-freedom (MDOF) system.

The stochastic processes and the estimated power spectra derived in Section 5.2 are utilised for the numerical example. For the generation of the stochastic processes and the estimation of the corresponding power spectra equations 5.4 – 5.7 are used. The time duration is $T = 64s$ with time step size $\Delta t = 0.25s$. For the frequency parameters an upper frequency $\omega_u = 12$ rad/s and a frequency step size of $\Delta\omega = 0.05$ rad/s are used.

In order to generate a variation of the different power spectra, the parameters σ and b in Eq. 5.6 are uniformly distributed random variables within the range $0.9 \leq \sigma, b \leq 1.2$. In this example five power spectra are used. From the ensemble of power spectra the mean is calculated, as described in Section 5.3 (Fig. 5.2).

The next step is the calculation of the confidence intervals. Here this is done for the confidence

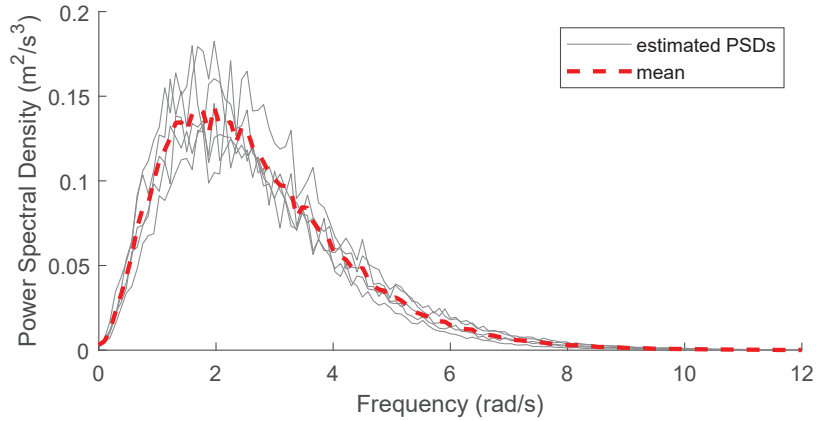


Figure 5.2: Ensemble of power spectra with mean power spectrum.

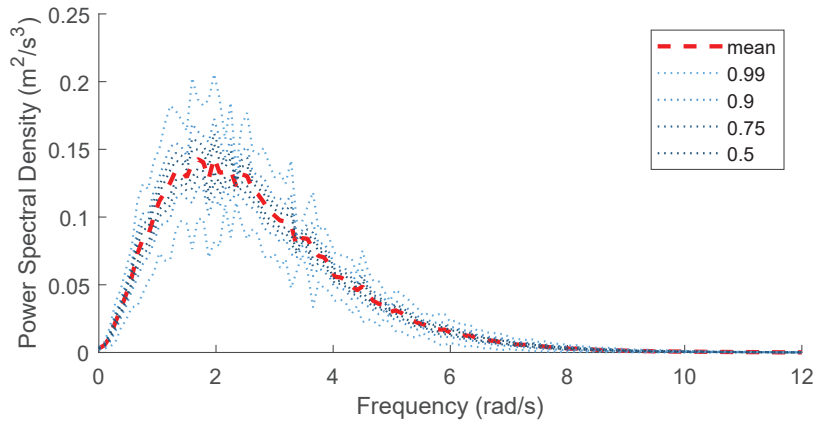


Figure 5.3: Mean power spectrum with confidence intervals.

levels 0.99, 0.9, 0.75 and 0.5 (Fig. 5.3). Based on the confidence level 0.99, which describes the largest confidence interval, the bounds of the possible range of the power spectra are determined, what is depicted in Fig. 5.4. It can easily be seen that these bounds form an envelope to the original ensemble of power spectra. This ensures that the relaxed model is defined for realistic values only, based on the ensemble of power spectra.

The subjective assignment of the confidence levels to the fuzzy numbers is shown, for instance, for three selected frequencies in Fig. 5.5. The plots of the power spectral density are related to the lines labelled with ‘A’, ‘B’ and ‘C’ in Fig. 5.4, which are describing intersections of the power spectral density at the specified frequencies. Corresponding to Fig. 5.5, in table 5.1 a summary of the values for the frequency $\omega = 1.97$ rad/s (‘A’) are given, as an example.

In the next sections the derived relaxed power spectrum will be applied to a SDOF system and a MDOF system, in order to determine and analyse the resulting response spectra of the systems.

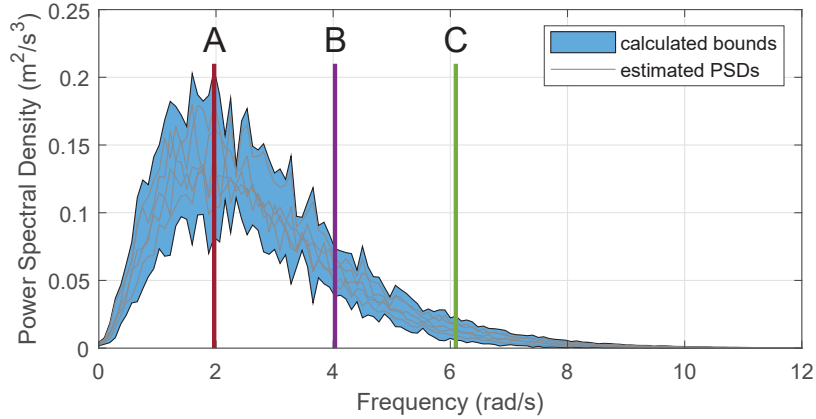


Figure 5.4: Power spectra bounds.

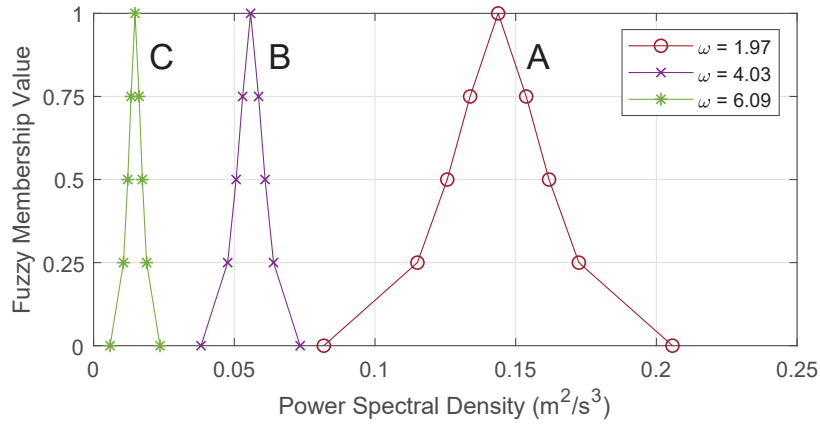


Figure 5.5: Membership function for selected frequencies.

5.4.1 Single-Degree-of-Freedom System

The derived relaxed power spectrum is applied to a SDOF system with the equation of motion

$$m\ddot{y} + c\dot{y} + ky = F(t), \quad (5.9)$$

where \ddot{y} , \dot{y} and y are the acceleration, velocity and displacement, respectively; m , c and k are the mass, damping coefficient and stiffness, respectively, and $F(t)$ is the external load. $F(t)$ is assumed to be a zero-mean stationary stochastic process with power spectral density $S_X(\omega)$ [154].

If the SDOF system is excited by the power spectral density $S_X(\omega)$, the power spectrum $S_Y(\omega)$ of the system response can be obtained by

$$S_Y(\omega) = |H(\omega)|^2 S_X(\omega). \quad (5.10)$$

This equation is called transfer function, where $H(\omega)$ is referred to as the frequency response

Table 5.1: Confidence levels and membership values for frequency $\omega = 1.97$ rad/s (labelled with ‘A’ in Fig. 5.5).

Confidence level	Spectral Power	α -level
—	0.1438	1
0.5	[0.1338, 0.1537]	0.75
0.75	[0.1257, 0.1619]	0.5
0.9	[0.1151, 0.1724]	0.25
0.99	[0.0818, 0.2057]	0.0

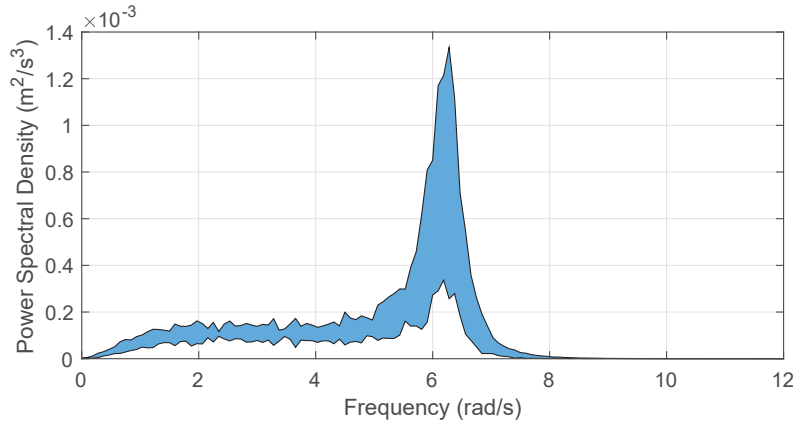


Figure 5.6: Bounds of the response spectra for the SDOF system.

function:

$$H(\omega) = \frac{1}{\omega_0^2 - \omega^2 + 2\zeta\omega_0\omega i} \quad (5.11)$$

In this equation ω_0 is the natural frequency of the system, ζ is the damping ratio and $i = \sqrt{-1}$ is the unit of the imaginary number [91].

If the relaxed power spectrum derived in Section 5.3 is applied to the SDOF system via the transfer function (eq. 5.10) with a natural frequency of $\omega_0 = 2\pi$ rad/s and a damping ratio $\zeta = 0.05$, the following relaxed response spectrum can be obtained, depicted in Fig. 5.6. The resulting relaxed power spectrum indicates the bounds for the possible range of response power spectra. The fuzzy membership values for the natural frequency of $\omega_0 = 2\pi$ rad/s are depicted in Fig. 5.7 and for the exact values shown in table 5.2.

5.4.2 Multiple-Degree-of-Freedom System

Similarly to the SDOF, the equation of motion for a MDOF system is

$$M\ddot{Y} + C\dot{Y} + KY = BX(t), \quad (5.12)$$

where \ddot{Y} , \dot{Y} and Y are the acceleration, the velocity and the displacement vector, respectively, M , C and K are the mass, the damping and the stiffness matrices, respectively, X is the

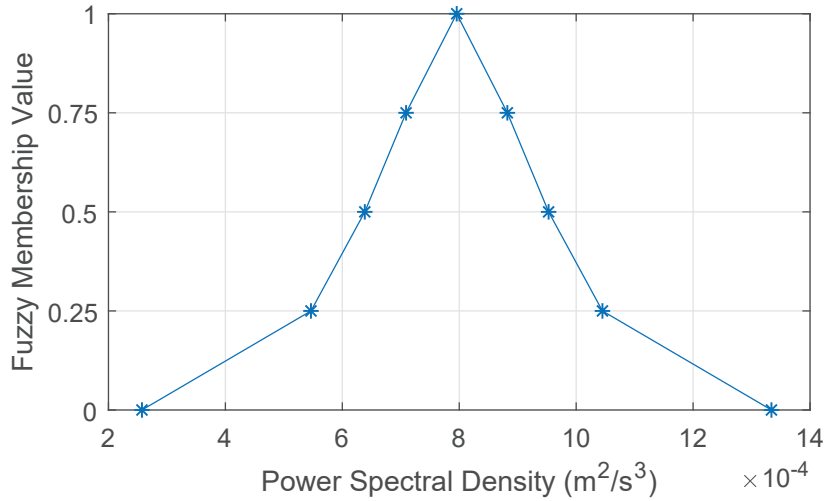


Figure 5.7: Fuzzy membership values for the natural frequency of $\omega_0 = 2\pi$ rad/s for the SDOF system.

Table 5.2: Confidence levels and membership values for the natural frequency $\omega_0 = 2\pi$ rad/s for the SDOF system.

Confidence level	Spectral Power (x10 ⁻⁴)	α -level
—	7.9566	1
0.5	[7.0905, 8.8227]	0.75
0.75	[6.3846, 9.5286]	0.5
0.9	[5.4638, 10.4494]	0.25
0.99	[2.5731, 13.3402]	0.0

stochastic excitation vector and B the input force influence matrix. For the MDOF system a nine-storey shear frame structure (Fig. 5.8) is investigated. The corresponding values for mass and stiffness for each of the storeys are given in table 5.3. This example, which includes both the system and the system properties, is based on Chen et al. [91].

Similarly to the SDOF system, the frequency response function of the MDOF system

$$H(\omega) = (K - \omega^2 M + i\omega C)^{-1} \quad (5.13)$$

can be obtained and the power spectrum of the response is

$$S_Y(\omega) = H(\omega) B S_X(\omega) B^T H^*(\omega), \quad (5.14)$$

where $H^*(\omega)$ is the complex-conjugate matrix of $H(\omega)$ and $S_X(\omega)$ is the power spectral density of $X(t)$. Since it is usually hard and time-consuming to obtain $H(\omega)$ directly from Eq. 5.13, alternatively the pseudo-excitation method (PEM) is adopted here. A brief summary is given here, for a detailed explanation see Lin et al. [155] and Li and Chen [52].

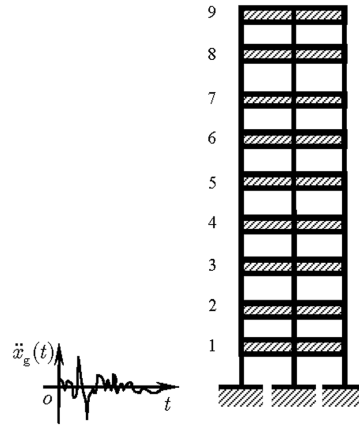


Figure 5.8: Nine-storey shear frame structure [91].

Table 5.3: Mass and stiffness of the shear frame structure [91].

Storey no.	Mass ($\times 10^5$)	Stiffness ($\times 10^8$)
1	3.5	1.47
2	3.3	1.63
3	3.0	1.62
4	3.0	1.60
5	3.0	1.60
6	3.0	1.92
7	3.0	1.85
8	2.7	0.96
9	2.7	0.89

Pseudo-excitation method

Consider a SDOF system with a harmonic excitation. The stable-state response of the system is

$$Y(\omega, t) = H(\omega)\sqrt{S_X(\omega)}e^{i\omega t} \quad (5.15)$$

and its complex-conjugate is

$$Y(\omega, t)^* = H^*(\omega)\sqrt{S_X(\omega)}e^{-i\omega t}. \quad (5.16)$$

By multiplying both sides of equations 5.15 and 5.16 the response spectrum $S_Y(\omega)$ can be derived:

$$S_Y(\omega) = Y(\omega, t)Y^*(\omega, t) = |H(\omega)|^2 S_X(\omega). \quad (5.17)$$

This indicates that the power spectrum of the response at certain angular frequency can be calculated by multiplying the deterministic response of the system to a harmonic excitation of the same angular frequency by its complex-conjugate part. Note that this excitation is made

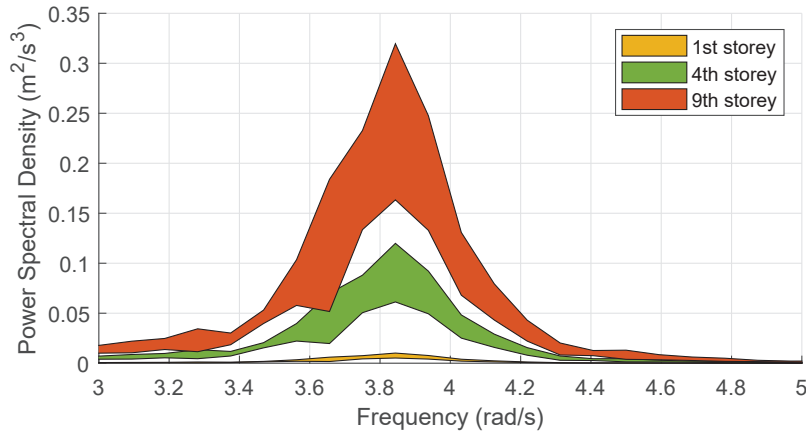


Figure 5.9: Bounds of the response spectra for the 1st, 4th and 9th storey of the MDOF system.

artificially, why this method is called *pseudo*-excitation method.

Numerical Procedure

For each angular frequency ω_j , $j = 1, \dots, N_\omega$, of the considered power spectrum a complex pseudo-excitation $\sqrt{S_X(\omega_j)}e^{i\omega_j t}$ or two real pseudo-excitations, one sine type $\sqrt{S_X(\omega_j)}\sin(\omega_j t)$ and one cosine type $\sqrt{S_X(\omega_j)}\cos(\omega_j t)$, $t \in [0, T]$, are generated. These pseudo-excitations are set as the outer excitation of the system and the corresponding stable-state response, $Y(\omega_j)$, is recorded. The power spectrum of the response $S_Y(\omega_j)$ can be obtained via Eq. 5.17 for each angular frequency ω_j .

Application to a MDOF System

The determination of the natural frequencies of the system yields

$$\omega_0 = [3.86, 10.84, 16.63, 23.67, 29.05, 33.28, 37.93, 42.87, 46.85].$$

Since the spectral power of the excitation power spectrum for $\omega \geq 10$ rad/s, which includes all other natural frequencies, are close to zero, the focus is on the first natural frequency. The application of the power spectrum derived in Section 5.3 yields to the response spectra depicted in Fig. 5.9. The response spectra for the first, fourth and ninth storey are shown. As it can be seen, the response spectra are similar, only the power of the response is higher and the bounds are widened, for higher storeys, since the system is a linear model. This behaviour can also be obtained in Fig. 5.10, where the fuzzy membership values of the corresponding storeys for the first natural frequency of $\omega_0 = 3.86$ rad/s are given. The shapes of the single fuzzy numbers are similar, they are only widened.

For instance, the fuzzy membership values of the ninth storey of the building for the response spectrum at the natural frequency $\omega_0 = 3.86$ rad/s are given in Table 5.4, corresponding to Fig. 5.10.

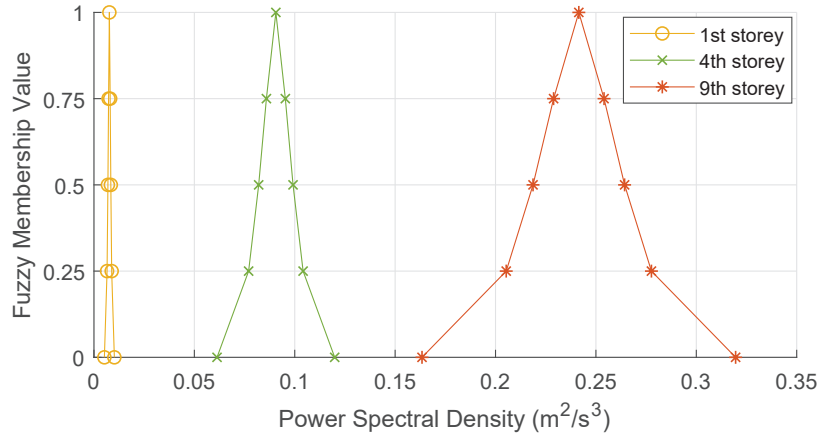


Figure 5.10: Membership values for the natural frequency $\omega_0 = 3.86$ rad/s for the 1st, 4th and 9th storey of the MDOF system.

Table 5.4: Confidence levels and membership values for the natural frequency $\omega_0 = 3.68$ rad/s for the 9th storey in Fig. 5.10.

Confidence level	Spectral Power	α -level
—	0.2415	1
0.50	[0.2289, 0.254]	0.75
0.75	[0.2187, 0.2643]	0.50
0.90	[0.2053, 0.2776]	0.25
0.99	[0.1634, 0.3196]	0.00

5.5 Conclusion

In this work the calculation of a relaxed power spectrum was presented. The new developed relaxed power spectrum offers several advantages compared to a traditional power spectrum described by discrete values for each frequency. The relaxed spectrum is characterized by the fact that it is more robust and able to capture the epistemic uncertainties. The model is especially useful when there are only a few real data records available. In this case, an approach with imprecise probabilities is useful in order to calculate a relaxed power spectrum because, in particular, due to the limitations mentioned in the Section 5.1, no reliable statistical information of the ensemble can be obtained.

Non-ergodic process realisations can be generated by the developed load model, which are directly applicable to Monte Carlo simulation analyses. Since the response process of the system, where the relaxed power spectrum is applied, will be presented in the same relaxed form of the load model, outer bounds and the corresponding fuzzy numbers can be calculated. The propagation of the fuzzy numbers through the system results in a fuzzy representation of each individual frequency of the response process, which allows due to the membership values of the fuzzy numbers a gradual failure range to be defined in the structural analysis. Thus, instead of

a crisp but uncertain response of the system a gradual failure range with mitigated uncertainties of the process can be determined.

In the examples considered in this work, the system behaviour is linear, what makes the propagation of the fuzzy numbers easy. For more complex systems with a non-linear behaviour, the propagation of the fuzzy numbers will be more difficult, and thus, there is a need of improvement for these cases.

Three crucial points must be taken into account when calculating the new developed relaxed power spectrum: First, the identification of the basis power spectrum: The shape of the resulting relaxed power spectrum depends strongly on the chosen or calculated, respectively, power spectrum. The mean of the ensemble of power spectra is used in this work as the basis power spectrum. However, other power spectra may be used, depending on the available data; Second, the calculation of the confidence interval: Depending on the available data, and the type of distribution function, different approaches may be used to calculate the confidence intervals, which will result in different relaxed power spectra. In addition, the confidence levels and their number is not limited to the way presented in this work. In other cases it might be useful to calculate more confidence intervals in order to get a more accurate presentation of the fuzzy numbers and thereby the response process; Third, the subjective assignment of the membership values: Since this affects the shape of the fuzzy sets for the individual frequencies, different relaxed power spectra can result. Therefore, it should be combined with expert knowledge to assign the values appropriately.

Practically in all cases, earthquakes, as a short-term process, have a relevant transient behaviour instead of a steady-state response on dynamic accelerations characterized by the response to a power spectrum. Since this work only considers the stationary case, it is necessary to extend the developed relaxed power spectra to the non-stationary case, where harmonic wavelets are utilised to estimate the frequency in the time-frequency domain.

In summary, the newly developed relaxed load model offers new possibilities in the analysis of structures. The uncertainties that arise due to various reasons are largely reduced. The load model provides a relaxed response process, resulting in spectral powers of frequencies with different membership values. Based on this, a failure range of the structures can be calculated.

Acknowledgement

This work was funded by the Deutsche Forschungsgemeinschaft (German Research Foundation) grants BE 2570/4-1 and CO 1849/1-1 as part of the project ‘Uncertainty modelling in power spectrum estimation of environmental processes with applications in high-rise building performance evaluation’.

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

A typical problem in engineering when using real data records is that they are never identical. This complicates the generation of a load model that describes the underlying stochastic process. The relaxed PSD (Chapter 2) and the imprecise PSD (Chapter 5 and 4) are examples of such load models.

In this part of the thesis, the data are classified according to their similarity to generate different load models from these data, which in turn contribute to the refinement of the simulation results. Uncertainties and especially dissimilarities are in the nature of such stochastic processes, so that no two identical earthquakes will ever occur, for instance. In particular, data sets are considered that have a high spectral variance in the frequency domain and are thus too dissimilar to generate a valid load model therefrom. Such a model would not represent the data set because the entire spectral range cannot be covered. Furthermore, selecting similar data reduces the variability in the simulation results, which is of high importance to obtain reliable results.

Therefore, a classification approach is presented that groups the data based on their spectral similarity. Although a variety of classification approaches already exist, they often require a set of parameters to obtain a valid classification. The approach presented here does not require any parameters other than a maximum number of groups to be determined. The similarity of the individual PSD functions is determined using Bhattacharyya distance. Based on this distance measure, the individual data sets are classified into groups using the k -means algorithm. From each of these groups, an individual load model can be generated, which reflects the similarity of the data and reduces the variability in the simulation results. Since each load model results in a new independent simulation, it is also proposed to determine the optimal number of groups using the Silhouette method. This ensures that as many simulations as necessary are carried out, but as few as possible.

The method is presented using two real data sets and two numerical examples in different scenarios. On the one hand a linear single degree of freedom system is used, on the other hand a non-linear model of a bridge pier is used to show its applicability to realistic cases. In each simulation, the maximum system displacement is determined. Based on this, a CDF is calculated to obtain an approximation of the failure probability. It is shown that the classification of the data leads to more accurate simulation results and that the previously determined optimal classification is valid. Using the probability of occurrence, a weighted CDF can also be formed that takes all load models into account. The classification approach is thus able to achieve a refinement of the simulation results. Moreover, the approach is useful to determine the similarity of the data, for the generation of the relaxed PSD or the imprecise PSD. These load models are also considered valid only if the data used have a certain similarity, which can be ensured with the proposed PSD classification.

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}

^aInstitute for Risk and Reliability, Leibniz Universität Hannover, Germany

^bInstitute for Risk and Uncertainty, University of Liverpool, United Kingdom

^cDepartment of Civil Engineering, Kanto Gakuin University, Japan

^dInternational Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University, Shanghai, China

Published in *Engineering Structures* on 1 October 2022

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 created 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 records 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 form 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.

6.1 Introduction

The simulation and subsequent reliability assessment of buildings and structures under specific loads have become increasingly important in engineering in the recent decades [53, 54, 111, 124]. 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 [52, 55, 126]. A general understanding of the dynamic behaviour of structures, especially under earthquake loads, is given in [125]. 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 [58, 61]. 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 [84], 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 [9, 22, 110]. Modal analysis and frequency decomposition methods by singular value decomposition of the PSD function are alternative approaches to MC simulations for characterising system responses, see for example [156–160]. 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 in the future. An overview about risk assessment of earthquakes is provided in [161]. Simulations are necessary to provide an understanding of the real case and to obtain an 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 [162], is often not possible due to the 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 special 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) [163–165], 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 [166–170]. In [171] a computationally efficient method for analysing the seismic fragility of structures is proposed, while in [172] fragility analyses are linked to artificial neural networks. Seismic fragility analysis is combined with Bayesian linear regression demand models in [173], yielding more accurate results compared to traditional methods. Other works deal with fragility analysis for specific structures, such as highway bridges [174], concrete dams [175] or railway bridges [176]. Both, the fragility analysis and IDA are concerned with the selection of seismic ground motions and with the definition of efficient and sufficient intensity measures [177, 178] of ground motions [179].

The definition of appropriate seismic intensities plays a key role in 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 [180–182]. Thanks to the ever-increasing databases of environmental processes (e.g. [112, 183]), a large amount of data is available from which corresponding load models can be generated. Although a pre-selection of the data can be conducted 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 [184] 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 [185]. In addition, uncertainties due to, for example, measurement errors, incorrect calibration, a damaged sensor or total failure of the sensor can complicate the selection and the subsequent analysis. Despite the fact that the data can be pre-selected according to the criteria mentioned above, 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 [186]. 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 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 variety 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 [187] that can be used for fast local solutions [188]. For example, in [189] a method is presented that takes the spectral shape into account. In [190], different frequency content-based parameters are used to classify the earthquakes using k-means and self-organising maps (SOM). The moment magnitude and the Joyner-Boore distance [191] are used in [192, 193] to classify earthquake ground motions with the k-means algorithm as well, while in [194] and [195] fuzzy-based approaches are employed. All these approaches require different parameters from the time and/or the 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 desired classification, which in turn affects 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 records 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. Moreover, these signals are used in the field of stochastic processes and dynamic systems, hence it is useful to classify ground motions based on their frequency characteristics. For the classification, the Bhattacharyya distance [196–198] is used to determine the similarity of the individual PSD functions. Utilising 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 statistic 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 [199, 200] is proposed to determine the optimal 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. For instance, such a classification is necessary to determine the similarity of data records to generate a probabilistic load model from multiple data, as proposed in [127].

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 [183]. This demonstrates that the proposed method is also feasible for practical applications. The developed load models from 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 a non-linear dynamic system.

This work is organised as follows: Section 6.2 summarises briefly the theoretical background used in this work. In Section 6.3 the approach for classifying PSD functions is explained. Section 6.4 provides the classification approach for two examples of real earthquake ground motions. In Section 6.5 the classified ensembles are applied to two numerical examples to show the strength of the approach. The work concludes with Section 6.6.

6.2 Stochastic processes and power spectrum estimation

Stochastic processes are determined by random influences, so that each value is described by a random variable. It is not possible to describe a stochastic process in a deterministic way. The frequency composition of a zero-mean stationary stochastic process $X(t)$ can be derived via the Fourier transform of its auto-correlation function $R_X(\tau) = E[X(t)X(t + \tau)]$

$$S_X(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R_X(\tau) e^{-i\omega\tau} d\tau \quad (6.1)$$

and the inverse Fourier transform

$$R_X(\tau) = \int_{-\infty}^{\infty} S_X(\omega) e^{i\omega\tau} d\omega, \quad (6.2)$$

where $S_X(\omega)$ describes the PSD function and τ denotes the time lag. Eq. 6.1 and Eq. 6.2 are the so-called Wiener-Khinchine theorem (see e.g. [11, 12, 52]).

The Spectral Representation Method (SRM) [84] can be employed to artificially generate stochastic processes $X(t)$. The prerequisite is a known expression of a PSD function S_X . SRM is described by the following expression

$$X(t) = \sum_{n=0}^{N-1} \sqrt{4S_X(\omega_n) \Delta\omega} \cos(\omega_n t + \varphi_n), \quad (6.3)$$

where the frequency points are defined as $\omega_n = n\Delta\omega$ with $n = 0, 1, 2, \dots, N-1$, N is the number of frequency points, $\Delta\omega$ is the frequency discretisation, φ_n are uniformly distributed random phase angles in the range $[0, 2\pi]$ and t is the time vector.

To estimate the PSD function from either real data records, e.g. an earthquake ground motion, or an artificially generated stochastic process, the periodogram [12, 52] can be utilised, which is based on the discrete Fourier transform and is as follows

$$\hat{S}_X(\omega_k) = \lim_{T \rightarrow \infty} \frac{\Delta t^2}{T} \left| \sum_{t=0}^{T-1} X(t) e^{-\frac{i2\pi}{T} kt} \right|^2. \quad (6.4)$$

In this expression, Δt describes the time discretisation, T is the total length of the record and k is the integer frequency for $\omega_k = 2\pi k/T$.

6.3 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.

6.3.1 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. 6.1. The time signals given here are derived from two different source PSD functions but the exact same random variables for generating the time signals (Eq. 6.3) are utilised. The PSD functions estimated from the time signals are given on the right side of Fig. 6.1. 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 conducted, and it may be useful to define two or more load models.

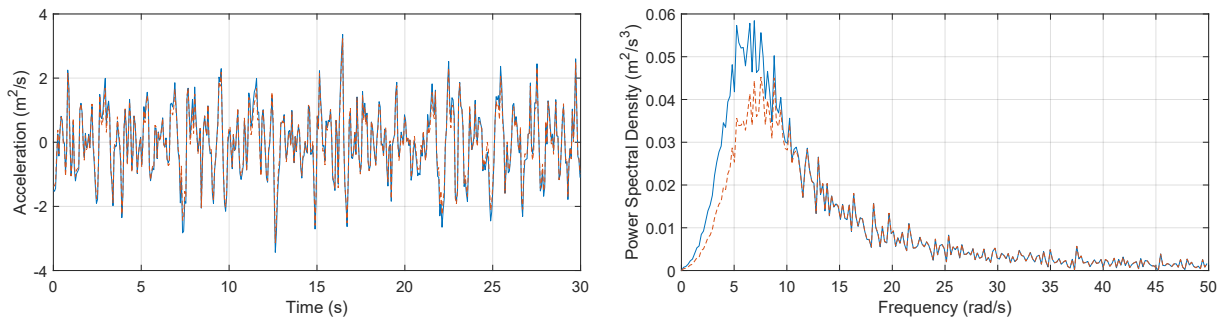


Figure 6.1: Example of two signals that show hardly any differences, but reveal the differences in the frequency domain.

6.3.2 Methodology

To identify different groups of PSD functions, the spectral similarity between two PSD functions S_1 and S_2 must be determined. In this work, it is proposed to use the Bhattacharyya distance [196–198]

$$D_B(S_1, S_2) = -\log \left(\sum_{\omega \in \Omega} \sqrt{S_1(\omega) S_2(\omega)} \right). \quad (6.5)$$

Due to its 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, for example, 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 used as a reference spectrum. Using the Bhattacharyya distance (Eq. 6.5), the similarity of each individual spectrum to the ensemble mean is determined. Therefore, Eq. (6.5) 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 mean lead to 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 6.3.3.

6.3.3 Optimal number of spectral groups

here are various methods to determine the optimal number of groups k_{opt} , such as the 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 on the methods used to measure the distances. In this work, it is suggested to use the silhouette method [199, 200] for the determination of k_{opt} . Other approaches can be found in [201] 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 $2, 3, \dots, k_{max}$ 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 records 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 s_C 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 $s(i)$

$$s_C = \frac{1}{n_i} \sum_{i=1}^{n_i} s(i) \quad (6.6)$$

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

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}. \quad (6.7)$$

According to [199], $a(i)$ 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 coefficient s_C is chosen as the optimal number of groups k_{opt} .

6.3.4 Example

In order to illustrate the proposed classification approach including the determination of the optimal number of groups by utilising the silhouette method, a short academic example is given in Fig. 6.2. In this example 3 different underlying analytical expressions of power spectra are utilised to generate 10 PSD functions each, with slightly 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 3 classified groups are completely disjoint, 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. (6.7) and are depicted in Fig. 6.3 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 obtained 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. (6.6) is used to compute the silhouette coefficient for each individual classification, which is the mean value of all silhouette values for the respective classification. This yields the silhouette coefficients shown in Fig. 6.4. The maximum of all silhouette coefficients reveals the optimal number of groups, accordingly $k_{opt} = 3$ in this example.

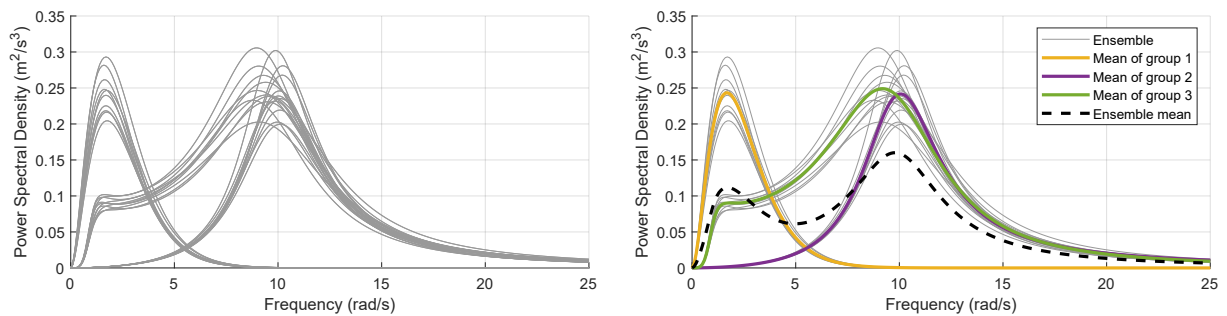


Figure 6.2: Unclassified ensemble (left) and the corresponding mean values of the classified groups and the entire ensemble (right).

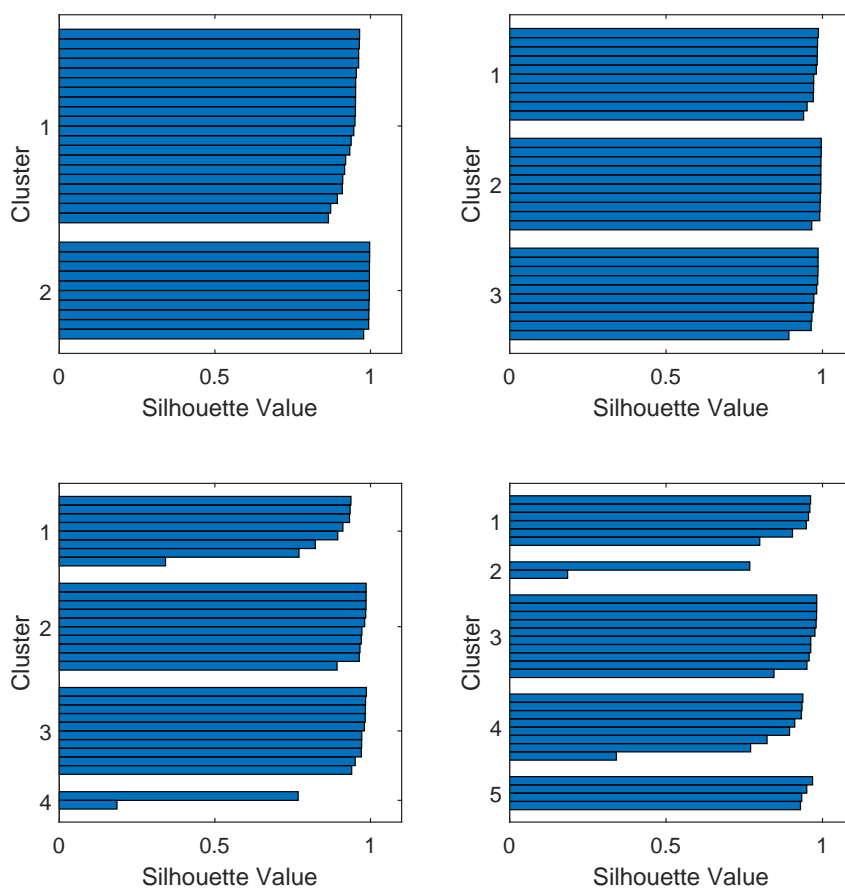


Figure 6.3: Silhouette values for the classification into 2, 3, 4 and 5 groups.

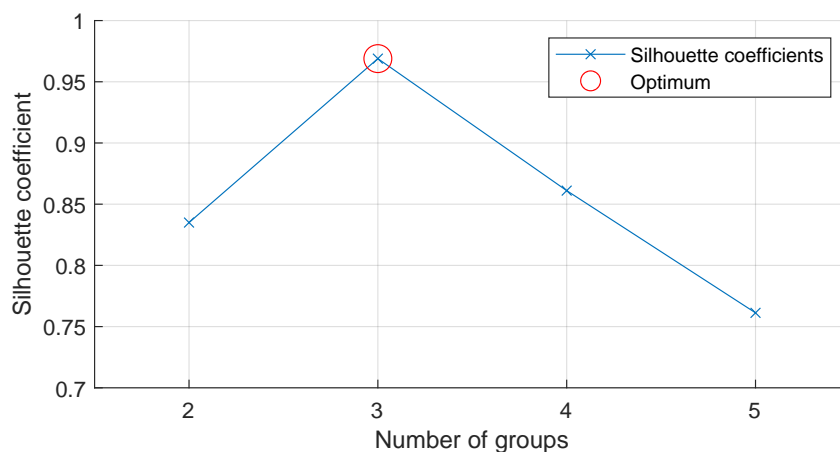


Figure 6.4: Silhouette coefficients for 2, 3, 4 and 5 groups. The optimal number of groups is $k_{opt} = 3$.

6.3.5 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:

- (i) Set up a load model based on the mean spectra of the classified groups. Then utilise Eq. 6.3, 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).

6.4 Classification of real data records

The real data records utilised in this work are provided by the K-NET and KiK-net database [183] 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 motions are illustrated.

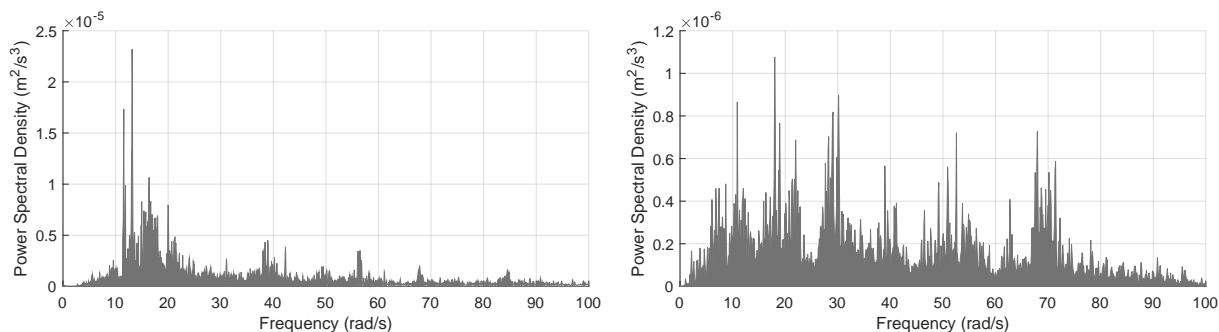


Figure 6.5: Source-specific (left) and site-specific (right) ensemble of real data records.

6.4.1 Source-specific data ensemble

For the classification of data records and the application to structures, real data of a specific earthquake event was utilised, see Fig. 6.5 (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 pre-selected according to their PGA, as it is meaningless to compare and classify data with completely different amplitudes. Therefore, only ground motions in the range $0.02 \text{ m/s}^2 \leq PGA \leq 0.06 \text{ m/s}^2$ 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. (6.4) and then classified using the Bhattacharyya distance (Eq. 6.5) and k-means algorithm. This was done for $k = 2$ and $k = 3$ groups. Fig. 6.6 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.

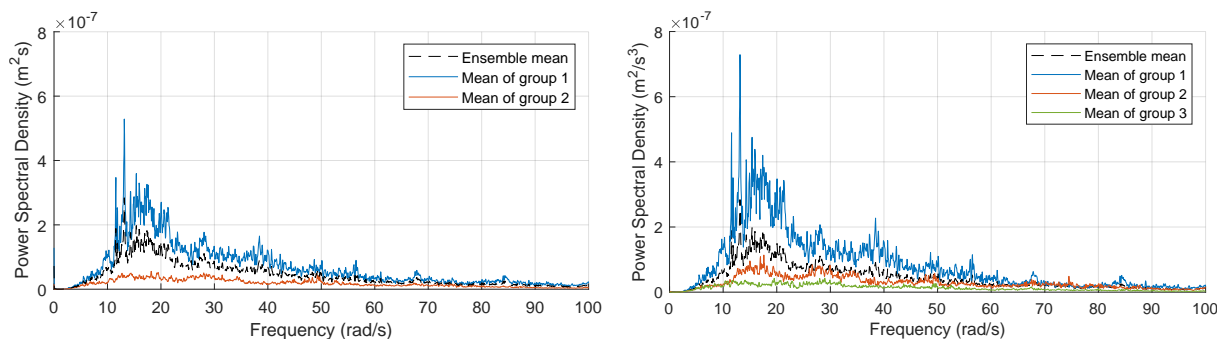


Figure 6.6: Mean PSD functions of the classified groups estimated from source-specific seismic ground motions with total length $T = 120$ s.

According to the silhouette method described in Section 6.3.3, the optimal number of groups is $k_{opt} = 2$, as shown in Fig. 6.7 and 6.8. For illustration purposes, however, classification and simulation are 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 6.1. In the time domain, minimum, maximum and mean value of the PGA of the classified groups were determined and are shown in Table 6.2. 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.

6.4.2 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. 6.5 (right). Data from July 2010 to July 2021 were used for the classification. The utilised earthquake ground motions have a

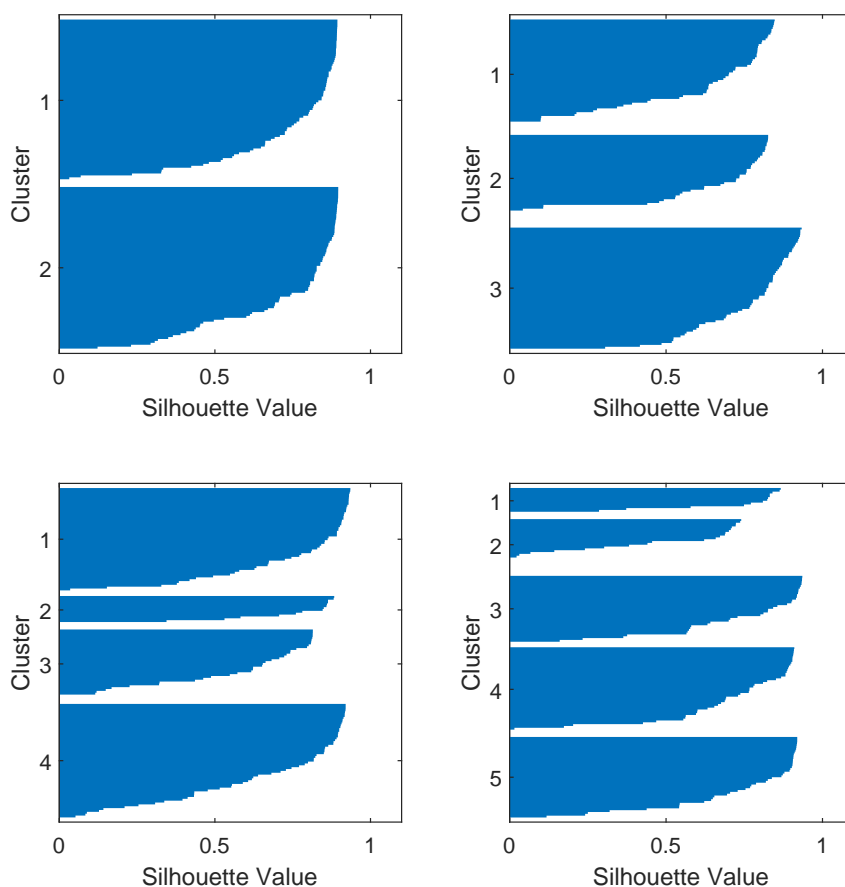


Figure 6.7: Silhouette values for the ensemble of source-specific data for the classification into 2, 3, 4 and 5 groups.

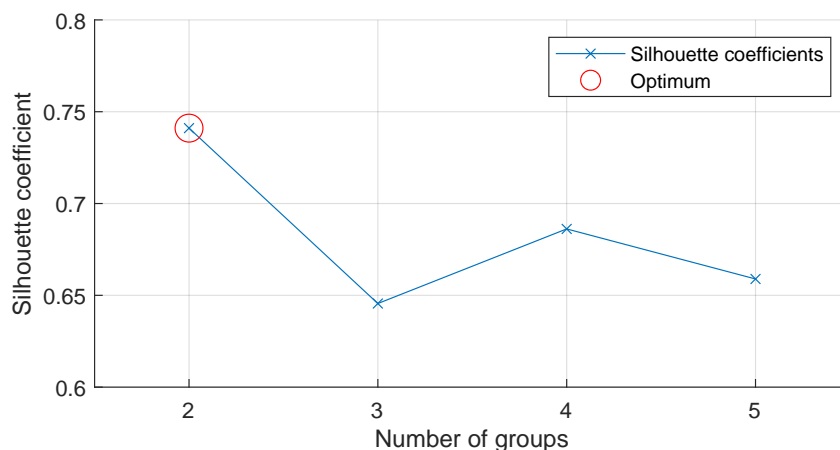


Figure 6.8: Silhouette coefficients for the source-specific ensemble for 2, 3, 4 and 5 groups. The optimal number of groups is $k_{opt} = 2$.

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 pre-selected according to their 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

Table 6.1: 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 6.2: 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

of 64 individual records. After transforming the data into the frequency domain according to Eq. (6.4) and classification for $k = 2$ and $k = 3$ groups, the corresponding mean PSD functions are obtained in Fig. 6.9. 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.

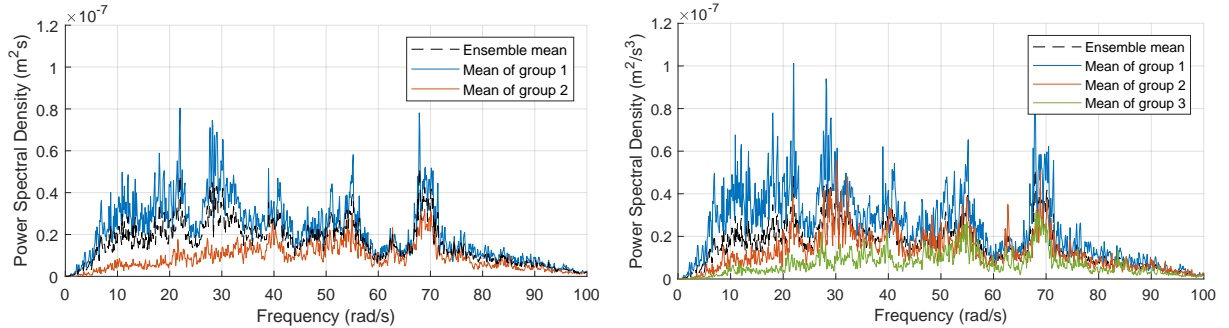


Figure 6.9: Mean PSD functions of the classified groups estimated from site-specific seismic ground motions with total length $T = 60$ s.

Determining the optimal number of groups using the silhouette methods yields $k_{opt} = 3$, as shown in Fig. 6.10 and 6.11.

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 6.3. The minimum, maximum and mean values of the PGA in the time domain of the groups were also determined and are shown in Table 6.4. As with the source-specific classification, a clear trend can also be seen in these values.

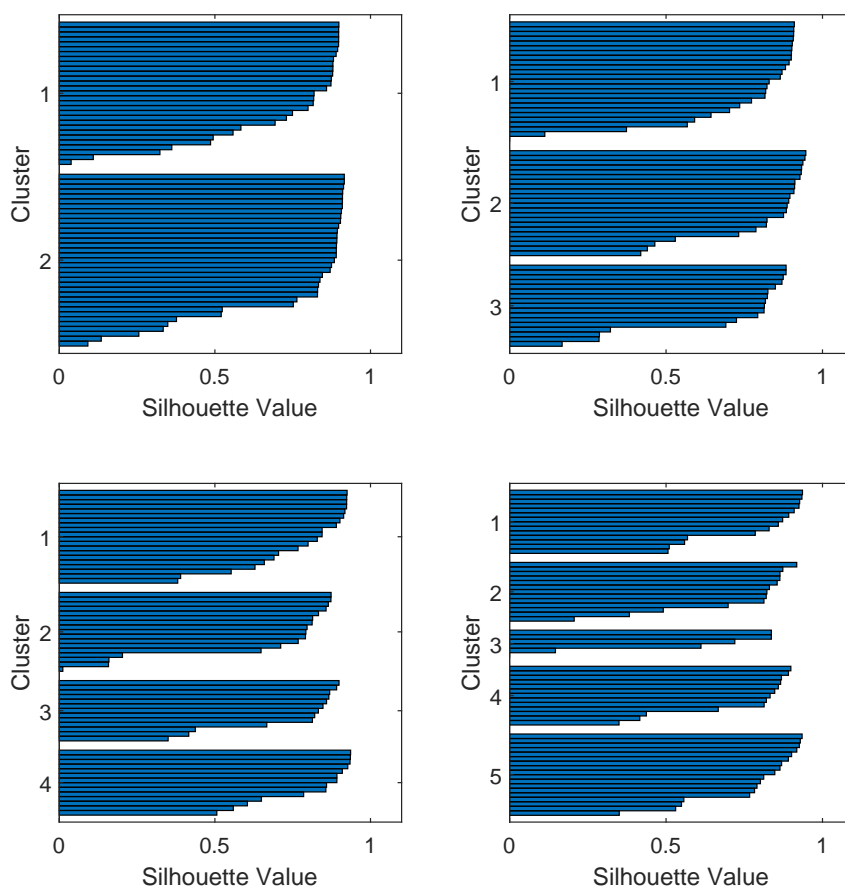


Figure 6.10: Silhouette values for the ensemble of site-specific data for the classification into 2, 3, 4 and 5 groups.

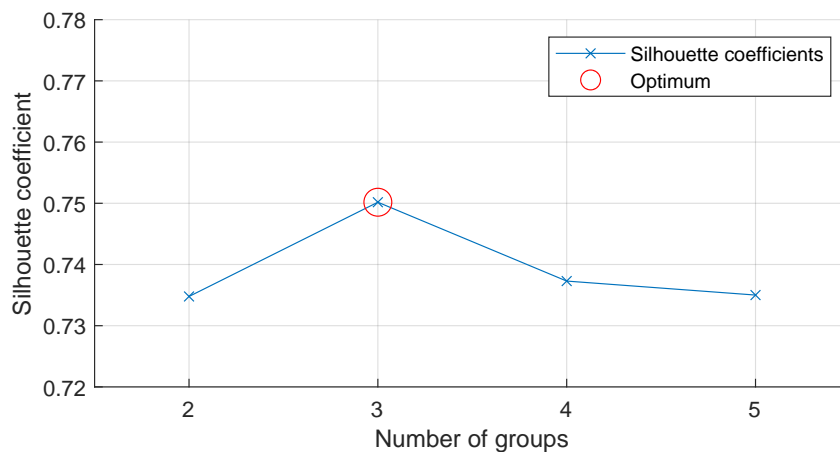


Figure 6.11: Silhouette coefficients for the site-specific ensemble for 2, 3, 4 and 5 groups. The optimal number of groups is $k_{opt} = 3$.

6.5 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

Table 6.3: Classification values for site-specific data in frequency domain

Classification	Group	Total power	Peak frequency value
$k = 2$	1	$[1.24 \cdot 10^{-5}, 3.56 \cdot 10^{-5}]$	$[1.34 \cdot 10^{-7}, 1.08 \cdot 10^{-6}]$
	2	$[4.35 \cdot 10^{-6}, 1.39 \cdot 10^{-5}]$	$[3.77 \cdot 10^{-8}, 4.08 \cdot 10^{-7}]$
$k = 3$	1	$[1.73 \cdot 10^{-5}, 3.56 \cdot 10^{-5}]$	$[2.57 \cdot 10^{-7}, 1.08 \cdot 10^{-6}]$
	2	$[8.85 \cdot 10^{-6}, 1.73 \cdot 10^{-5}]$	$[1.06 \cdot 10^{-7}, 8.96 \cdot 10^{-7}]$
	3	$[4.35 \cdot 10^{-6}, 9.33 \cdot 10^{-6}]$	$[3.77 \cdot 10^{-8}, 2.46 \cdot 10^{-7}]$

Table 6.4: 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

approach. The first example aims to demonstrate the effectiveness of the proposed classification approach and verify the identified optimal classification, using a linear mass-spring-damper system considering different scenarios in the relationship between the natural frequency of the system 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. 6.6 and Fig. 6.9) are used in the following for the numerical examples. For these, SRM (Eq. 6.3) is utilised to generate adequate time signals as excitation for the systems. 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 structures. For each sample, the maximum displacement of the systems in the time domain are determined, from which a cumulative distribution function (CDF) is calculated that can be used to estimate the probability of failure for specific displacements.

Stationary stochastic processes are generated for the mass-spring-damper system discussed in 6.5.1, while non-stationary stochastic processes emulated by an envelope function are generated for the bridge pier model in 6.5.2, since the response properties of non-linear dynamic systems are strongly affected by the non-stationarity of input ground motions. The envelope function is given by

$$g(t) = k \left(e^{-at} - e^{-bt} \right), \quad (6.8)$$

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. 6.12.

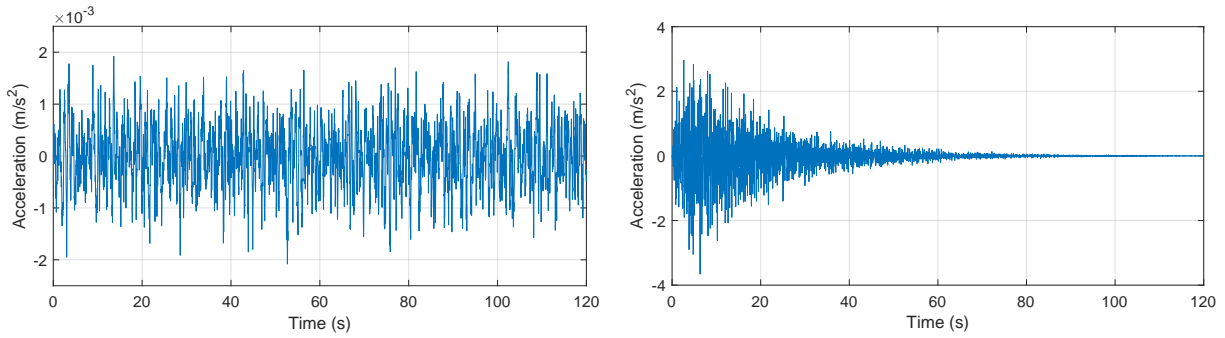


Figure 6.12: Stationary (left) and non-stationary ground motion acceleration (right), generated by SRM (Eq. 6.3) and the envelope function (Eq. 6.8).

6.5.1 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 following equation of motion

$$m\ddot{x} + c\dot{x} + kx = F(t), \quad (6.9)$$

with m as mass, c as damping coefficient and k as stiffness. 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 6.4. An explicit Runge-Kutta scheme [117] is used to solve Eq. (6.9).

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 6.5.

Table 6.5: Parameters of the SDOF system for different scenarios.

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

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. 6.13. 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 the 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 (6.13, left). 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 (6.13, right).

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 natural frequency of the system. 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 groups, which supports that $k_{opt} = 2$ is reasonable and correct.

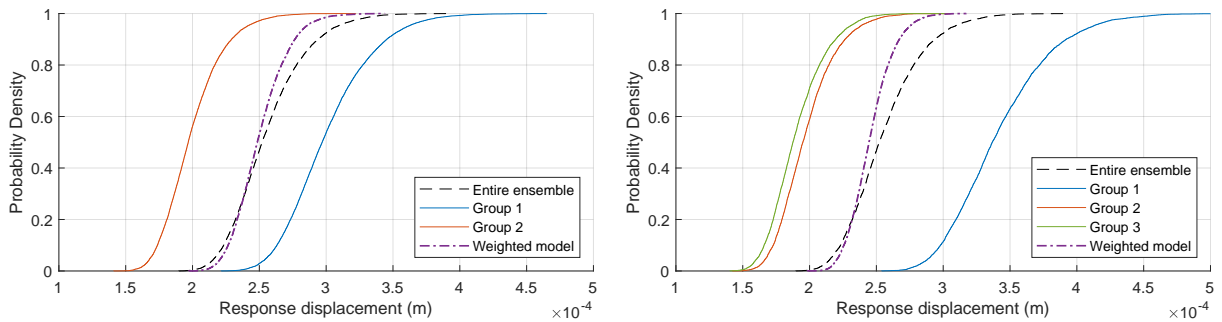


Figure 6.13: CDFs of the maximum response displacement of the linear mass-spring-damper system for scenario A for the classification into 2 groups (left) and into 3 groups (right) of the source-specific data ensemble.

In Fig. 6.14 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 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. 6.14, 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.

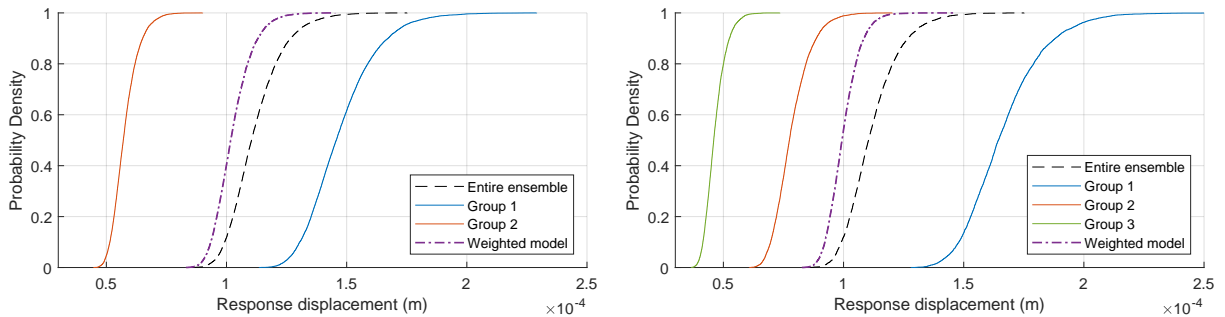


Figure 6.14: CDFs of the maximum response displacement of the linear mass-spring-damper system for scenario B for the classification into 2 groups (left) and into 3 groups (right) of the source-specific data ensemble.

Although the silhouette coefficients are often very close, see for example Fig. 6.8, the application of the classified models, however, shows that it is indeed effective. This is particularly evident in the results in Fig. 6.13. The classification results in the optimal number of groups $k_{opt} = 2$, which yields 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.

Results of site-specific data

The results of the site-specific data and the corresponding classifications show a similar behaviour as the results of the source-specific classification. In Fig. 6.15 the CDFs for the classification into 2 groups (left) and into 3 groups (right) are shown for scenario C. 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 sufficient for an accurate estimation of system responses, and thus the classification into 3 groups is reasonable.

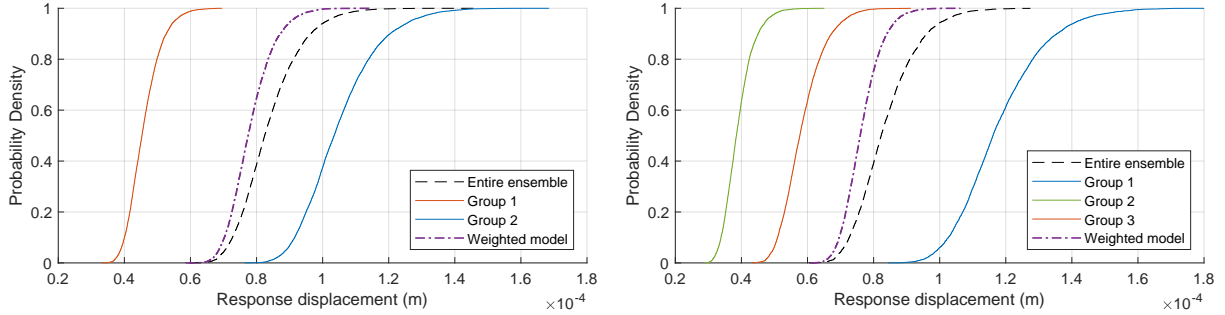


Figure 6.15: CDFs of the maximum response displacement of the linear mass-spring-damper system for scenario C for the classification into 2 groups (left) and into 3 groups (right) of the site-specific ensemble.

The results of scenario D, using the site-specific data and its classification, are shown in Fig. 6.16. 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. 6.16 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 the 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.

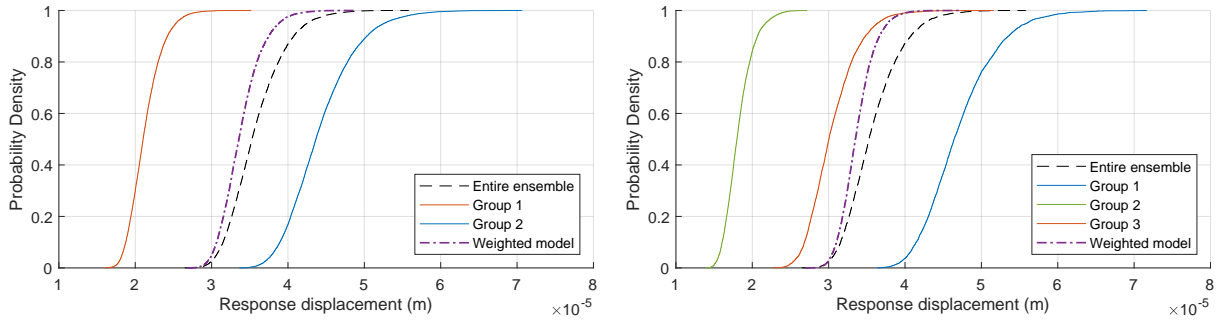


Figure 6.16: CDFs of the maximum response displacement of the linear mass-spring-damper system for scenario D for the classification into 2 groups (left) and into 3 groups (right) of the site-specific ensemble.

6.5.2 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 [202] and the manual on bearings for highway bridges [203]. The

bridge pier is modelled as a 2-DOF lumped mass system and consists of a superstructure and a reinforced concrete (RC) pier, which is modelled as a non-linear horizontal spring, see Fig. 6.17. 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 [204]. 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 6.6.

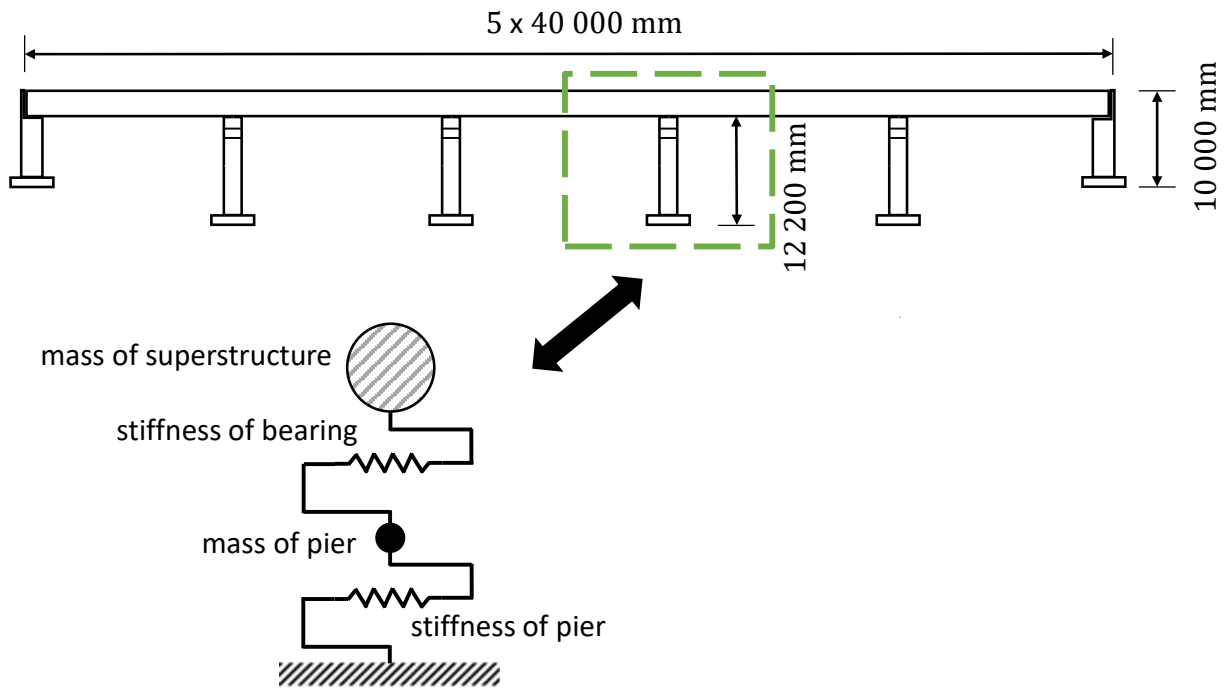


Figure 6.17: 2-DOF lumped mass model for the bridge pier.

Table 6.6: Model parameters of the bridge pier.

	Model parameter	Nominal value
Superstructure	Mass M_S (ton)	604
	Yield strength (kN)	1118
Rubber bearing	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)	110,100

An example of the non-linear force-displacement behaviour of the rubber bearing of the bridge

pier model is depicted in Fig. 6.18.

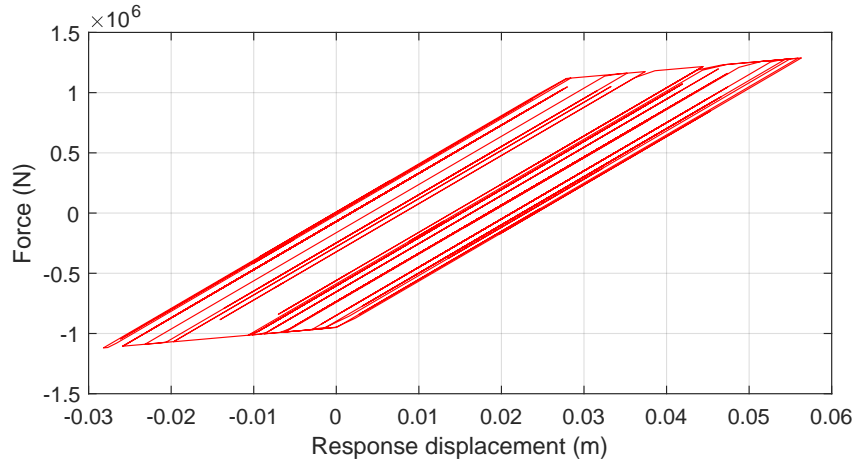


Figure 6.18: Force-displacement behaviour of the rubber bearings.

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 displacement at the RC pier and at the rubber bearings, is determined in the time domain, from which the CDF is calculated. In this case, only non-stationary earthquake ground motions were utilised 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 of groups for the classification, since response properties of the non-linear dynamic system 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 of the optimal classifications are presented.

Results of the non-linear bridge pier model

The results of the source-specific data are given in Fig. 6.19. The CDFs of the maximum displacements at the RC pier of the bridge (left) for the optimal number of groups $k_{opt} = 2$ and of those at the rubber bearings (right) 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.

In Fig. 6.20 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

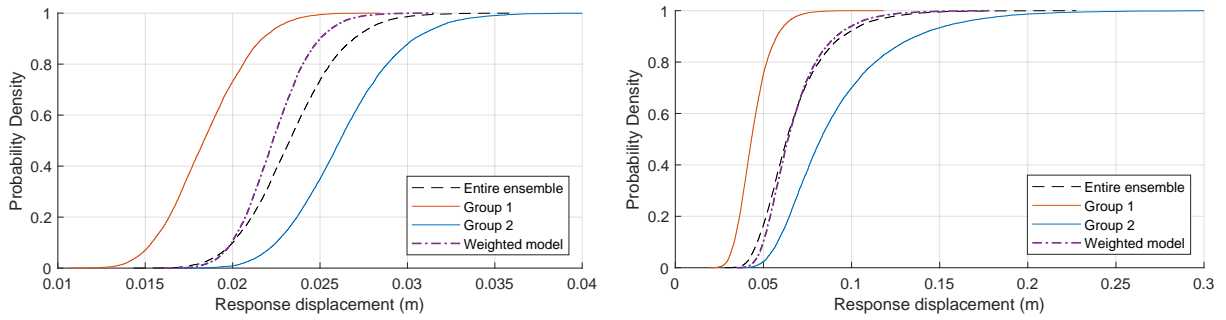


Figure 6.19: CDFs of the maximum response displacement of the seismic-isolated bridge pier model for the source-specific ensemble for the classification into $k_{opt} = 2$ groups. Results for the RC pier are shown on the left, results for the rubber bearings are shown on the right.

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 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.

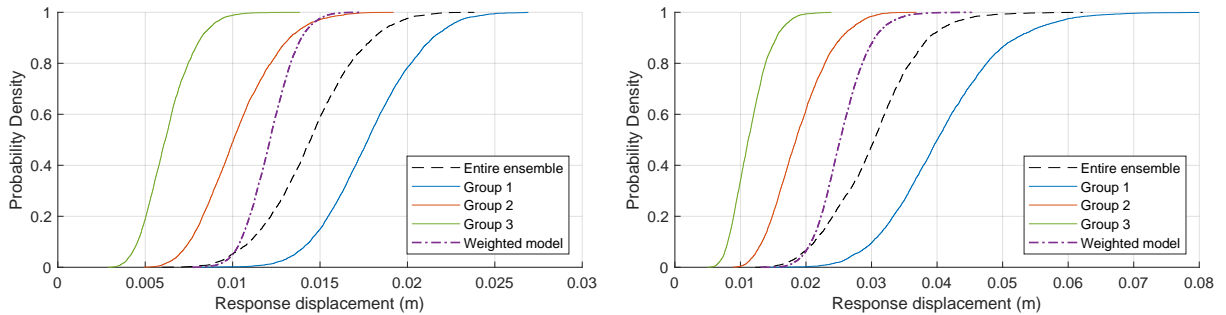


Figure 6.20: CDFs of the maximum response displacement of the seismic-isolated bridge pier model for the site-specific ensemble for the classification into $k_{opt} = 3$ groups. Results for the RC pier are shown on the left, results for the rubber bearings are shown on the right.

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.

6.6 Conclusions

A new technique has been proposed for developing load models from ensembles of PSD functions 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 results in higher system displacements that would otherwise remain undetected and a possible system failure would thus not be identified. 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 of groups for the 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 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.

7 | Projecting interval uncertainty through the discrete Fourier transform: An application to time signals with poor precision

In contrast to the previous chapters, where only discrete-valued signals were analysed, this chapter deals with the quantification of uncertainties of single interval-valued signals. The uncertainties in such a time signal are modelled as intervals. In order to calculate the Fourier amplitude and subsequently estimate the PSD function from such an interval signal, an algorithm is needed that can propagate intervals through the DFT. An algorithm that is capable of performing this task is presented in this part of the work. The algorithm allows to calculate the exact bounds of the Fourier amplitude and the PSD function. While plain interval arithmetic is sufficient to determine bounds, a more advanced algorithm is provided that also takes into account the repeated variables problem that arises in this computation. When a variable occurs multiple times in a computation, plain interval arithmetic can overestimate the interval uncertainty and thus artificially inflate it, so that the bounds can be unreasonably large. The algorithm presented here takes repeated variables into account and thus provides exact bounds. The proposed interval DFT algorithm is explained in detail in this part of the thesis, with particular emphasis on the individual steps of the algorithm. In addition, the entire algorithm is verified in a comparison with plain interval arithmetic and it is shown that the exact bounds are in fact calculated. Three different algorithms are considered, one that employs plain interval arithmetic, a brute-force algorithm and the selective algorithm. The latter one is based on the brute-force algorithm but computes the bounds significantly faster due to the calculation of the convex hull of all endpoints for each calculation step. A comparison of the computational costs is presented, as well as pseudocodes for all algorithms.

The algorithm is demonstrated using an example of an offshore wind turbine subject to wave loads. An artificially generated signal describing the sea waves is translated into an interval signal assuming uncertainty. This interval signal is then transformed into the frequency domain by means of the proposed interval DFT algorithm in order to obtain the exact bounds of the PSD function. By means of an analytical description of the input-output relationship via the frequency response function, the bounded PSD function is propagated through the system such that bounds can be obtained for the system response as well. This allows an immediate determination of whether the system response is in a critical range. The interval DFT algorithm can therefore be used directly for reliability analysis, taking interval uncertainties in time signals into account.

Projecting interval uncertainty through the discrete Fourier transform: An application to time signals with poor precision

Marco Behrendt^{a,b}, Marco de Angelis^b, Liam Comerford^b, Yuanjin Zhang^c, Michael Beer^{a,b,d}

^aInstitute for Risk and Reliability, Leibniz Universität Hannover, Germany

^bInstitute for Risk and Uncertainty, University of Liverpool, United Kingdom

^cSchool of Safety Science and Emergency Management, Wuhan University of Technology, China

^dInternational Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University, Shanghai, China

Published in *Mechanical Systems and Signal Processing* on 1 June 2022

Abstract

The discrete Fourier transform (DFT) is often used to decompose a signal into a finite number of harmonic components. The efficient and rigorous propagation of the error present in a signal through the transform can be computationally challenging. Real data is always subject to imprecision because of measurement uncertainty. For example, such uncertainty may come from sensors whose precision is affected by degradation, or simply from digitisation. On many occasions, only error bounds on the signal may be known, thus it may be necessary to automatically propagate the error bounds without making additional artificial assumptions. This paper presents a method that can automatically propagate interval uncertainty through the DFT while yielding the exact bounds on the Fourier amplitude and on an estimation of the Power Spectral Density (PSD) function. The method allows technical analysts to project interval uncertainty—present in the time signals—to the Fourier amplitude and PSD function without making assumptions about the dependence and the distribution of the error over the time steps. Thus, it is possible to calculate and analyse system responses in the frequency domain without conducting extensive Monte Carlo Monte Carlo simulations nor running expensive optimisations in the time domain. The applicability of this method in practice is demonstrated by a technical application. It is also shown that conventional Monte Carlo methods severely underestimate the uncertainty.

Keywords: Discrete Fourier transform, Complex intervals, Dependency tracking, Interval arithmetic, Power spectral density estimation, Uncertainty quantification.

7.1 Introduction

The discrete Fourier transform (DFT) is ubiquitous in signal processing and in engineering computing in general. The DFT allows signals to be decomposed into single harmonics and so facilitate their data compression and analysis. Its versatility resides in its mathematical property of being linear and fully invertible. The DFT is used in a range of different applications in science and engineering, such as in spectral analysis, random vibrations, differential equations, data compression, signal processing, image processing or probabilistic programming [205, 206]. There are various algorithms available for transforming a signal with the DFT, of which the best-known is probably the fast Fourier transform (FFT), presented by Cooley & Tukey [207]. Due to the increasing computational power, simulations and equivalent calculations can be carried out ever faster, which has also opened up the possibility of conducting the FFT analysis on-the-fly

on embedded systems. An overview of a variety of algorithms used can be found in abundance in the literature [208, 209].

The uncertainty in signals can arise, for example, from damaged sensors, equipment failures, measurement errors, operational range, inaccurately calibrated sensors, and incorrectly recorded extreme values. Moreover, data can be imprecise because of measurement uncertainty, e.g., due to degradation or digitisation. There are numerous cases and reasons why the signal can be imprecise, and yet still informative in sensing critical circumstances. In order to obtain accurate simulation results, e.g., in the context of reliability analysis, it is imperative to use suitable simulation methods and models. Since the simulation is only a representation of the real case, real data should be used by all means. Uncertain data increasingly complicates the task of obtaining meaningful results. Some approaches on how to deal with uncertain data can be found in [5, 17, 210].

Handling interval uncertainty is a broad problem at the intersection between engineering, mathematics and statistics [210–212]. For the reasons mentioned above, the signal can be of poor quality and the interpretation of the simulation results might lead to unrealistic or even dangerous outcomes, for example if an actually disastrous assumption is shifted into an acceptable range by incorrect quantification. Such a scenario is particularly conceivable in risk analysis, as even smallest deviations in the input data can lead to a system failure that must be detected. Especially in the dynamic behaviour of systems, data with poor precision can lead to enormous problems, e.g., when the natural frequencies of the system under consideration are excited. The behaviour of the system can change significantly in magnitude if imprecise rather than precise data is used. Therefore, every attempt must be made to account for uncertainties.

Especially in the field of random vibrations [54, 111] and stochastic dynamics [52], where environmental processes, such as wind and earthquake loads or sea waves, are the governing excitations, a precise analysis of structures and buildings is indispensable. The DFT is particularly useful here to determine the dominant frequencies of excitation for vibration analysis. A versatile method to represent stochastic processes is the Power Spectral Density (PSD) function, which represents the transformation of signals in the frequency domain and is also calculated by means of the DFT [12]. For instance, it is used in earthquake engineering to display the amplitude of acceleration with respect to frequencies given a signal. Therefore, the amplitude of the DFT is of special interest in these fields, as it has an important physical meaning.

Some approaches have already been taken to estimate a reliable amplitude of the DFT from uncertain data. In the field of stochastic dynamics, missing data were reconstructed under the assumption that they are normally distributed. This probability density function of those data was then propagated through the DFT [77, 105]. In other fields, the FFT and convolution were studied for signals with interval and fuzzy uncertainty [213].

In this work, signals with poor precision are represented by intervals whose width can either be constant or variable along the signal. The objective is to find the ranges of the absolute value of the DFT of such an interval signal, i.e. upper and lower bounds of the Fourier amplitude, and not

merely the ranges for real and imaginary parts of the transformed signal. Interval arithmetic is sufficient to propagate interval signals through the DFT, but since the calculation of the amplitude contains repeated variables, this approach usually does not yield exact bounds [41, 42, 44]. An interval algorithm has been proposed in [214] that fully addresses the dependency problem caused by the repeated variables, leading to the exact bounds on the amplitude of the DFT. In this paper, the proposed algorithm is presented in details and applied to an engineering example involving signals with poor precision.

This work is organised as follows: An overview about the theoretical background is given in Section 7.2. In Section 7.3, the novel algorithm for propagating interval uncertainty through the DFT is described. Section 7.4 presents an application for the utilisation of the proposed algorithm. The final conclusions are given in Section 7.5.

7.2 Preliminaries

In this section, some essential background is given on the rules of interval arithmetic used in this work, on the representation of interval signals and on the interval extensions of the transform. In addition, the repeated variables problem is introduced and the theory of PSD function estimation is described briefly.

7.2.1 Interval arithmetic

An interval \underline{x} is a compact subset of \mathbb{R} and is defined as $\underline{x} = [x, \bar{x}] = \{x \leq x \leq \bar{x}\}$. A complex interval $\underline{z} \subseteq \mathbb{C}$, consisting of intervals for the real component $\underline{z}_{re} \subset \mathbb{R}$ and the imaginary component $\underline{z}_{im} \subset \mathbb{R}$, is defined as

$$\underline{z} = \underline{z}_{re} + i \underline{z}_{im} = \{z_{re} + i z_{im} \mid z_{re} \in \underline{z}_{re} \wedge z_{im} \in \underline{z}_{im}\}. \quad (7.1)$$

Using the rectangular interpretation of a complex interval, there are four endpoints: $z_1 = z_{re} + i z_{im}$, $z_2 = z_{re} + i \bar{z}_{im}$, $z_3 = \bar{z}_{re} + i z_{im}$, $z_4 = \bar{z}_{re} + i \bar{z}_{im}$, each corresponding to a vertex of the rectangle. The addition between two complex intervals \underline{z} and \underline{w} , can be defined in terms of such endpoints as:

$$\underline{z} + \underline{w} = [\min\{z_i + w_j\}, \max\{z_i + w_j\}], \text{ for } i, j = 1, \dots, 4. \quad (7.2)$$

For a singleton $a \in \mathbb{R}$, the singleton-interval multiplication is:

$$a\underline{z} = [\min\{az_i\}, \max\{az_j\}], \text{ for } i, j = 1, \dots, 4. \quad (7.3)$$

The absolute value of a complex number is obtained by squaring real and imaginary components separately, followed by the square root of their sum, $|\underline{z}| = \sqrt{\underline{z}_{re}^2 + \underline{z}_{im}^2}$. The square and square

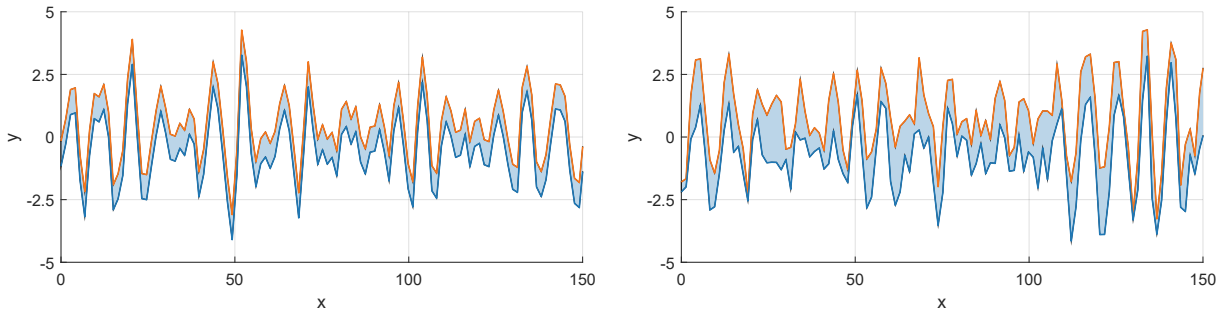


Figure 7.1: Two signals with poor precision: constant interval uncertainty $\xi = 0.5$ (left), variable interval uncertainty between $\xi_{min} = 0.1$ and $\xi_{max} = 1.5$ (right).

root of a real interval \bar{x} are

$$\bar{x}^2 = \begin{cases} [\underline{x}^2, \bar{x}^2], & \bar{x} > 0 \\ [\bar{x}^2, \underline{x}^2], & \bar{x} < 0 \\ [0, \max(\underline{x}^2, \bar{x}^2)], & \bar{x} \ni 0, \end{cases} \quad (7.4)$$

and

$$\sqrt{\bar{x}} = [\sqrt{\underline{x}}, \sqrt{\bar{x}}], \quad x > 0. \quad (7.5)$$

7.2.2 Signal with interval uncertainty

Sensor data can be subject to experimental uncertainty for a variety of reasons, given that their acquisition can be faulty, intermittent due to maintenance, and of poor precision [210]. The reasons are complex and range from simple measurement errors to total failure of the sensors. Often the sensors are not accurate enough, are not calibrated correctly or are subject to certain technical limitations. For example, extreme values may not be recorded due to sensor threshold limitations. In some cases, the data can be disturbed by external influences, specifically if sensors are used for long-term recordings. It is also possible that the sensor is damaged by the event it is supposed to record, for example an earthquake, and makes incorrect recordings or stops recording completely. The placement of the sensors also has a significant impact on the quality of the signal. In addition, sensors may be temporarily unavailable due to maintenance. If the time interval of unavailability is sufficiently short, intervals could be used to bridge this gap.

In this work, the assumption is made that sensors work accurately only within certain tolerances. These tolerances are represented by intervals, i.e. each value of the data record is described by an upper and lower bound and every arbitrary value within these bounds is possible. The interval uncertainty can be specified as constant or variable along the signal. Two examples of constant and variable interval uncertainty are illustrated in Fig. 7.1. Intervals are the most conservative model of error, because no assumption is made about how the error is distributed nor how it depends on anything else. Statistics for experimental data carrying interval uncertainty are being extensively studied in the literature and just to mention a few, include [210–212, 215].

An interval signal can be defined in central notation as

$$\bar{x} = x + \xi\Delta, \quad (7.6)$$

where $x \in \mathbb{R}^N$ is the vector of midpoints, $\Delta = [-1, 1]$ is the unitary interval and ξ is the precision of the signal. When the precision is variable across the signal, $\xi \in \mathbb{R}^N$ is a vector of differing scalars, otherwise it is a vector of identical scalars. This representation will be used to generate simulated interval signals from precise intervals given measurement error.

7.2.3 Interval extensions

The DFT converts a signal $x = x_0, x_1, \dots, x_{N-1}$ to its Fourier sequence $z = z_0, z_1, \dots, z_{N-1}$, for $k = 0, \dots, N - 1$.

The *interval extension* of the DFT is obtained by replacing the real signal with their interval values. The resulting interval Fourier transform at frequency number k is

$$\bar{z}_k = \sum_{n=0}^{N-1} \bar{x}_n \cdot e^{-\frac{i2\pi}{N}kn} = \sum_{n=0}^{N-1} \bar{x}_n \cdot \left[\cos\left(\frac{2\pi}{N}kn\right) - i \cdot \sin\left(\frac{2\pi}{N}kn\right) \right]. \quad (7.7)$$

The interval DFT amplitude is

$$\bar{A}_k = |\bar{z}_k| = \sqrt{\left[\sum_{n=0}^{N-1} \bar{x}_n \cdot \cos\left(\frac{2\pi}{N}kn\right) \right]^2 + \left[\sum_{n=0}^{N-1} \bar{x}_n \cdot \sin\left(\frac{2\pi}{N}kn\right) \right]^2}. \quad (7.8)$$

An alternative interval extension of the DFT can be obtained representing the interval signal as $\bar{x}_n = x_n + \xi\Delta$, i.e. separating interval from non-interval components, which results in

$$\bar{z}_k = \sum_{n=0}^{N-1} x_n \cdot e^{-\frac{i2\pi}{N}kn} + \sum_{n=0}^{N-1} \xi\Delta \cdot e^{-\frac{i2\pi}{N}kn}, \quad (7.9)$$

where $\Delta = [-1, 1]$ is the unitary interval and $\xi \in \mathbb{R}$ is the precision of the interval signal, expressed in the same units as the signal. This extension can also be used to derive algorithms for the interval propagation based on the zonotope representation as shown in [216, 217], and to make comparisons with other methods, like Monte Carlo, for computing the bounds. Nonetheless, this representation will not be used for the derivation of the main algorithm.

Remark: An *interval extension* is obtained replacing the original expression's variables directly with interval variables. The fundamental theorem of interval analysis states that every interval extension—obtained combining the four rules of arithmetic—is *inclusion monotonic* [42, 43]. This translates into the well-known conservatism of interval analysis for computing the bounds. In other words, the bounds obtained by interval analysis on interval extensions are always inclusive thus rigorous.

Remark: A *united extension* is defined as the union of all the images of a function evaluated on all the subsets of a given interval space. The image of a united extension is often not a box, but an arbitrarily-shaped set, which is referred to as the *united set*. In interval analysis, conservatism is often due to approximating the united set with its enclosing box. This kind of conservatism of interval computations arises in functions with repeated variables, or in back-calculation problems, and is also known as the wrapping effect.

7.2.4 Repeated variables problem

For readers not familiar with the *repeated variables problem* of interval computing, a brief recap is given as follows: Interval computations produce bounds that are *best possible*, i.e. without inflated uncertainty, when the interval variables appear only once in their mathematical expression. For example, the evaluation of a second-degree polynomial $a\bar{x}^2 + b\bar{x} + c$ with interval arithmetic can result in inflated bounds due to the interval \bar{x} repeating twice in the expression. For more details on this problem the reader can be referred to [218].

7.2.5 PSD function 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 [11]. An estimation of the stationary PSD function of a stochastic process can be obtained by the periodogram [12, 52], which can be defined by the squared amplitude of the DFT. The periodogram for a non-interval discrete signal x_n is

$$\hat{S}_X(\omega_k) = \frac{\Delta t^2}{T} \left| \sum_{n=0}^{N-1} x_n \cdot e^{-i\frac{2\pi}{N}kn} \right|^2, \quad (7.10)$$

where Δt is the time step size, T is the total length of the record, n describes the data point index in the record and k is the frequency number of $\omega_k = \frac{2\pi k}{T}$. The interval extension of Eq. (7.10) in trigonometric form for an interval signal \bar{x}_n is

$$\overline{\hat{S}_X}(\omega_k) = \frac{\Delta t^2}{T} \left(\left[\sum_{n=0}^{N-1} \bar{x}_n \cdot \cos\left(\frac{2\pi}{N}kn\right) \right]^2 + \left[\sum_{n=0}^{N-1} \bar{x}_n \cdot \sin\left(\frac{2\pi}{N}kn\right) \right]^2 \right). \quad (7.11)$$

Because Eq. (7.11) can be expressed in terms of the square of the Fourier amplitude as follows: $\overline{\hat{S}_X}(\omega_k) = \frac{\Delta t^2}{T} \overline{A}_k^2$, and because the Fourier amplitude \overline{A}_k is a non-negative real interval, the interval PSD function $\overline{\hat{S}_X}(\omega_k)$ can be computed without inflation, provided that its frequency components are not combined together in further computations.

7.3 The interval discrete Fourier transform algorithm

The objective of this work is to calculate the *exact* bounds on the interval extension of the Fourier amplitude in Eq. (7.8), and subsequently on the interval extension of an estimation of the PSD function in Eq. 7.11. Under no dependency assumption (noninteractivity) interval arithmetic suffices to obtain the exact bounds on the Fourier sequence. To support this statement, it will suffice to note that Eq. (7.7) has no repeated variables. Conversely, repeated interval variables occur in the calculation of the interval amplitude spectrum of Eq. (7.8), thus additional computational investment is needed. In order to obtain the exact bounds, the finite set of complex pairs determining the boundary of the *united set* of the DFT is needed. Because the DFT is a linear map, its image under the interval constraints, is a compact set called united set. In this section, an algorithm that computes such united set, leading to the exact bounds on the amplitude of the DFT is presented. This algorithm is referred to as the *selective* algorithm.

Because the DFT is a linear map, the united set can be obtained by endpoints analysis on Eq. (7.7). Tracking all the endpoints in Eq. (7.7) allows the representation of the united extension in terms of its finite boundary, but this task has exponential complexity $O(2^N)$, with N being the cardinality of the set of intervals, which coincides with the length of the signal in this study. So, a better algorithm is needed to track down the endpoints whose image through the DFT lies on the boundary of the united set. Fig. 7.2 shows the united set (octagon), the interval extension (box), the image through the DFT of the endpoints (blue dots). Among the latter, those on the boundary of the united set are depicted with orange diamond-shaped markers. The interval extension is obtained applying the rules of interval arithmetic presented in Section 7.2. The selective algorithm is tasked with the following three main steps:

1. Select the set of complex pairs on the boundary (vertices) of the united set, and track down the corresponding endpoints.
2. Select the two vertices that are farthest and nearest to the origin of the complex plane. These are named *anchoring points*. If the united set contains the origin of the complex plane, then return the origin as the nearest point.
3. Compute the absolute value of these two precise complex numbers to obtain the resulting interval amplitude. The absolute value of the farthest and nearest point is the upper and lower bound, respectively.

These three steps are repeated for each frequency number k in Eq. 7.8, to construct the interval Fourier spectrum.

7.3.1 Obtaining the boundary of the united set

In this section, the procedure used by the *selective* algorithm to track down the endpoints whose image is on the boundary of the united set is presented. To understand the procedure, the

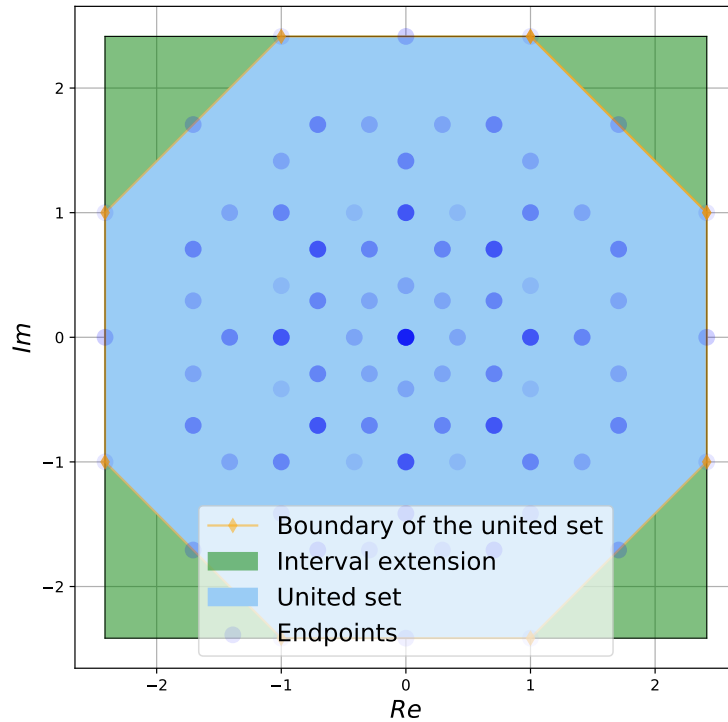


Figure 7.2: Boundary points (orange diamond markers) of the united set (light blue), interval extension box (green), and endpoints (blue dots) mapped through the Fourier transform for a short signal with $N = 8$.

Fourier sum of Eq. (7.7) can be conceptualised as a sequence of elementary additions. Each addend of this series, $\bar{x}_n \cdot e^{-\frac{i2\pi}{N}kn}$, is an interval in the complex plane, whose real and imaginary components are perfectly dependent. In order to see that real and imaginary components are perfectly dependent, the $(n + 1)$ _{th} addend for a given frequency number k in trigonometric form can be written as:

$$\bar{z}_{k,n} = \bar{x}_n \cos\left(\frac{2\pi}{N}kn\right) - i \bar{x}_n \sin\left(\frac{2\pi}{N}kn\right).$$

From the trigonometric form, it appears evident that real and imaginary components are the same interval \bar{x}_n differing by just a multiplicative factor.

In the complex plane, the interval addend $\bar{z}_{k,n}$ is better represented by an oriented segment rather than a box, as shown in Fig. 7.3. The Fourier series can be imagined as the sum of such dependent interval objects. What happens when two such objects are added together? Because a segment is a convex bounded set, the sum between two such sets is another convex set, whose vertices are obtained by summing the addends' vertices. Will all the vertices end up on the boundary of the resulting convex set? The answer is no, but why? Although there is an intuition as to which pairs of vertices will reach the boundary before the addition is done, the algorithm adds together all the vertices, and a posteriori discards the pairs of vertices that end up in the interior of the united set. An important theorem in computational geometry, that has led to the Minkowski addition, guarantees that the points in the interior are necessarily

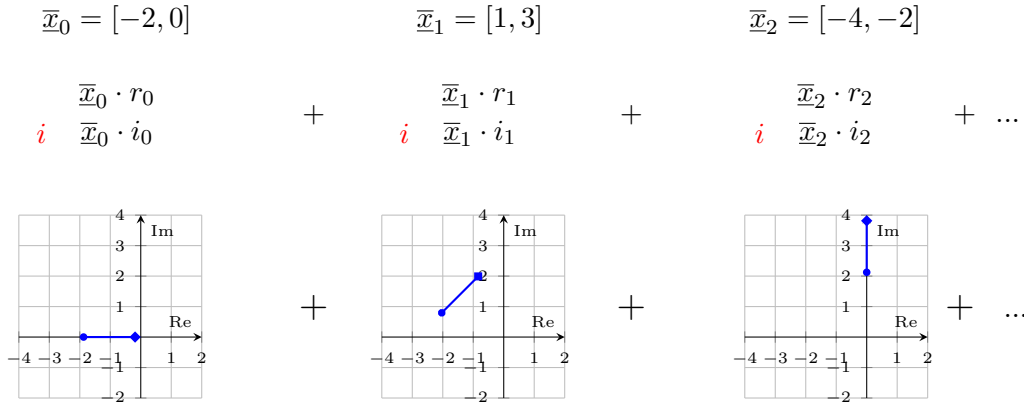


Figure 7.3: First three addends of the interval Fourier series seen as the sum of perfectly dependent intervals, where $r_n = \cos\left(\frac{2\pi}{N}kn\right)$ and $i_n = \sin\left(\frac{2\pi}{N}kn\right)$ for $N = 8$, and $k = 5$.

mapped in the interior of the resulting polygon. In practice, the interior points are discarded by constructing the convex hull of the set of vertices mapped through each addition, followed by the indexing of only the points on the boundary, which will be subsequently kept in memory. The polygon resulting from adding two dependent intervals—shown in Fig. 7.4—is then added to the third dependent interval as shown in Fig. 7.5. This time, the addition generates two polygons, whose convex hull is the resulting united set. Once again, all the mapped vertices that end up in the interior are removed from memory, whilst the process continues until each addend of the Fourier sum has been processed.

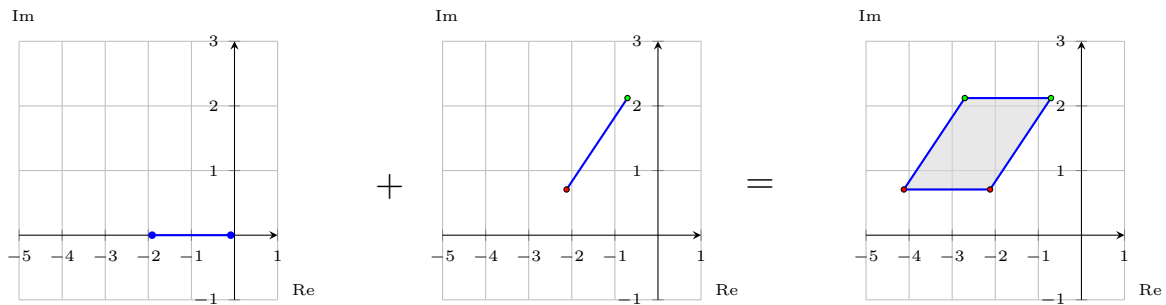


Figure 7.4: The first addition of the Fourier series results in a polygon with four vertices.

The key steps performed by the *selective* algorithm are also presented in the Pseudocode 1 and are described in detail in Appendix A.

Verification of the selective algorithm

The interval extension of the DFT can be used as verification tool to ensure the correctness of the algorithm on arbitrary signals. This is possible thanks to the fact that the bounding box tightly circumscribe (see Fig. 7.2) the united set because there are no repeating variables in Eq. 7.7. In other words, for the verification of the algorithm, it is sufficient to show that the boundary of the united set is always resting (for some points) on the boundary of the interval extension

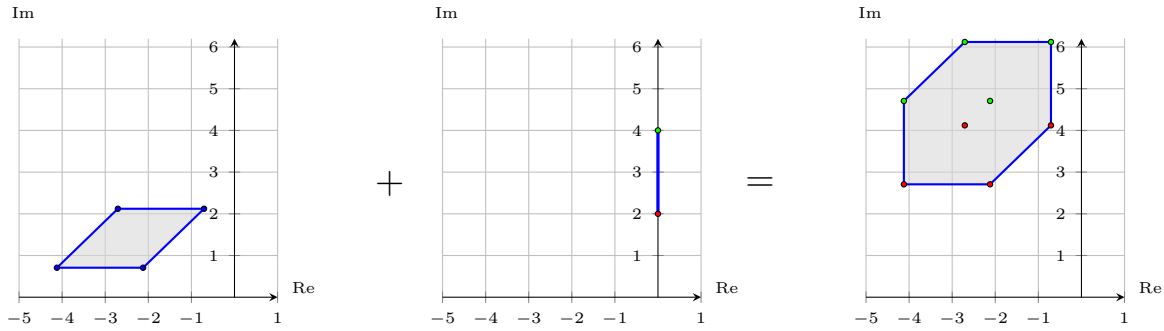


Figure 7.5: The second addition is the Minkowski sum of a polygon (resulting from the first addition) and the third dependent interval in the Fourier series. The two vertices mapped in the interior of the resulting polygon will be discarded.

Pseudocode 1 Selective algorithm

```

1: function EXACTAMPLITUDEBOUNDS( $\underline{x}$ ,  $k$ )
2:    $N \leftarrow l(\underline{x})$  ▷ length of interval signal
3:    $v \leftarrow [e^{-i2\pi 0k/N} \underline{x}[0], e^{-i2\pi 0k/N} \bar{x}[0]]$  ▷ compute first two vertices
4:    $c_{hull} \leftarrow v$  ▷ initialise convex hull
5:   for  $n \leftarrow 1$  to  $N - 1$  do
6:      $v \leftarrow [e^{-i2\pi nk/N} \underline{x}[n], e^{-i2\pi nk/N} \bar{x}[n]]$  ▷ compute subsequent vertices
7:      $ep \leftarrow$  add the two vertices  $v$  separately to the list  $c_{hull}$  to get a list of new endpoints
8:      $c_{hull} \leftarrow$  get convex hull of  $ep$  in  $\mathbb{R}^2$  with real and imaginary components as coordinates
9:   end for
10:  if origin is inside  $c_{hull}$  then
11:     $\bar{A}_k \leftarrow$  Interval(0,  $\max(|c_{hull}|)$ )
12:  else
13:     $\bar{A}_k \leftarrow$  Interval( $\min(|c_{hull}|)$ ,  $\max(|c_{hull}|)$ )
14:  end if
15:  return  $\bar{A}_k$  ▷ bounds on the Fourier amplitude for frequency  $k$ 
16: end function

```

box. This verification procedure can be done automatically and systematically within the code that implements the algorithm. This result can be appreciated in Fig. 7.6 where the boundary of the united set, the interval extension box and the endpoints are all depicted in the same plot for four different frequency numbers k . The interval extension algorithm, which is presented in the Pseudocode 2 in Appendix A, can also be used to obtain an outer approximation of the united set, when speed is a priority. This verification algorithm is much faster than the selective algorithm (Pseudocode 1), because no convex hull evaluation ought to take place. More details about the cost of the selective and interval algorithms are provided in Section 7.3.4.

7.3.2 Determining the two anchoring points

The two anchoring points are defined as the farthest and the nearest vertices of the united set from the origin of the complex plane. The absolute value of these two anchoring points yield the

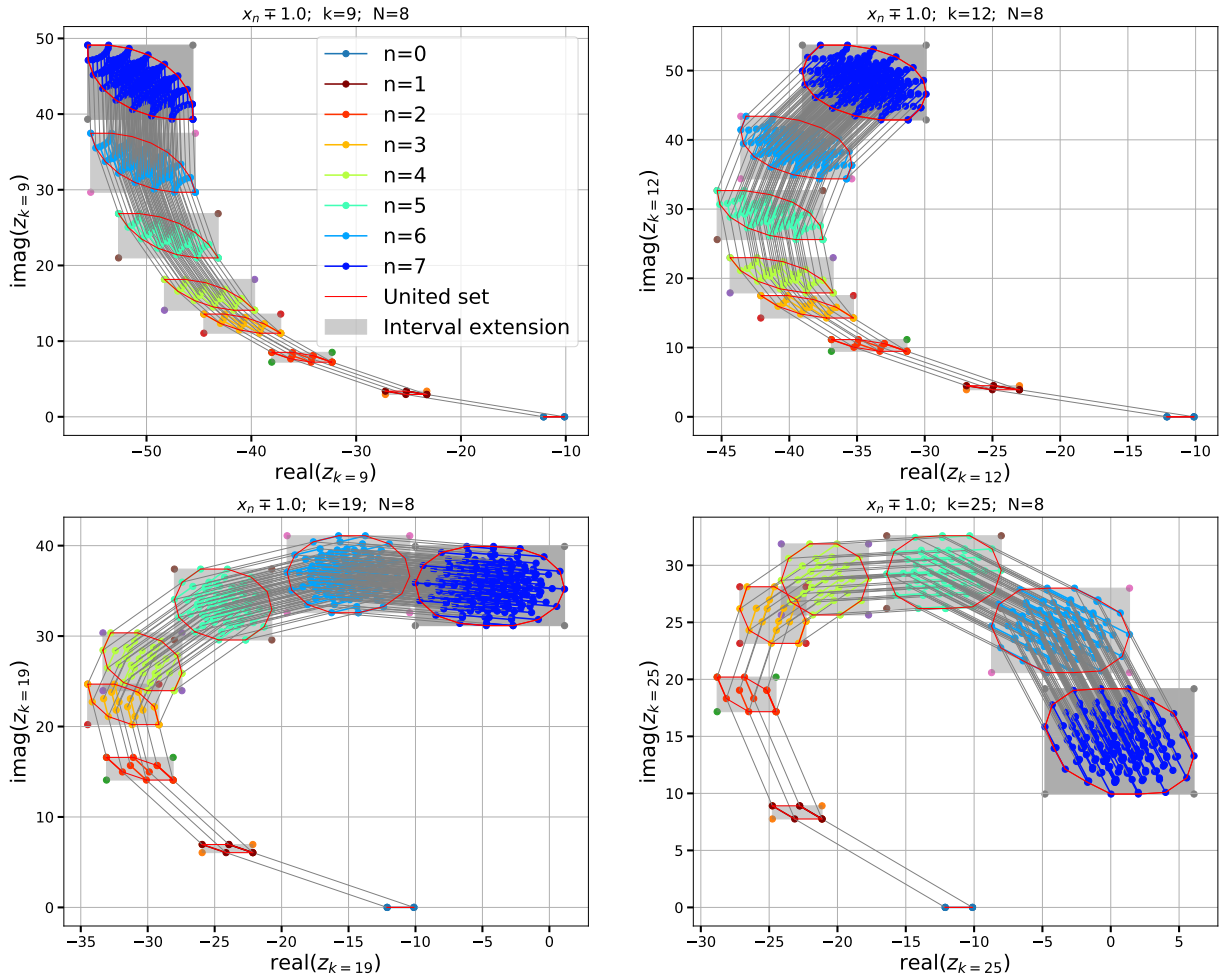


Figure 7.6: Boundary of the united set in red, image of the endpoints, and interval extension box, at each term of the Fourier series, up to $n = 7$, for frequencies $k \in \{9, 12, 19, 25\}$.

minimum and maximum amplitude at any given frequency. This holds under the interpretation that the absolute value, thus the amplitude of Eq. (7.8), can be seen as the Euclidean distance from the origin of an equivalent point in \mathbb{R}^2 whose coordinates are the real and imaginary components. In Appendix B, a simple proof is provided to show that the maximal distance is attained at one of the vertices of the united set.

If the united set at a given frequency number k is denoted by Z_k , and its border by ∂Z_k , the interval amplitude of the Fourier transform can be computed taking the minimum and maximum over all the vertices of the united set,

$$\overline{A}_k = \max_{\min} \{|z_k| : z_k \in \partial Z_k\}. \quad (7.12)$$

This is exemplified in Fig. 7.7 for four different frequencies. When the united set contains the origin, the lower bound amplitude is zero, see Fig. 7.8. A function that checks if the origin of the complex plane is contained in the united set is needed to perform this task.

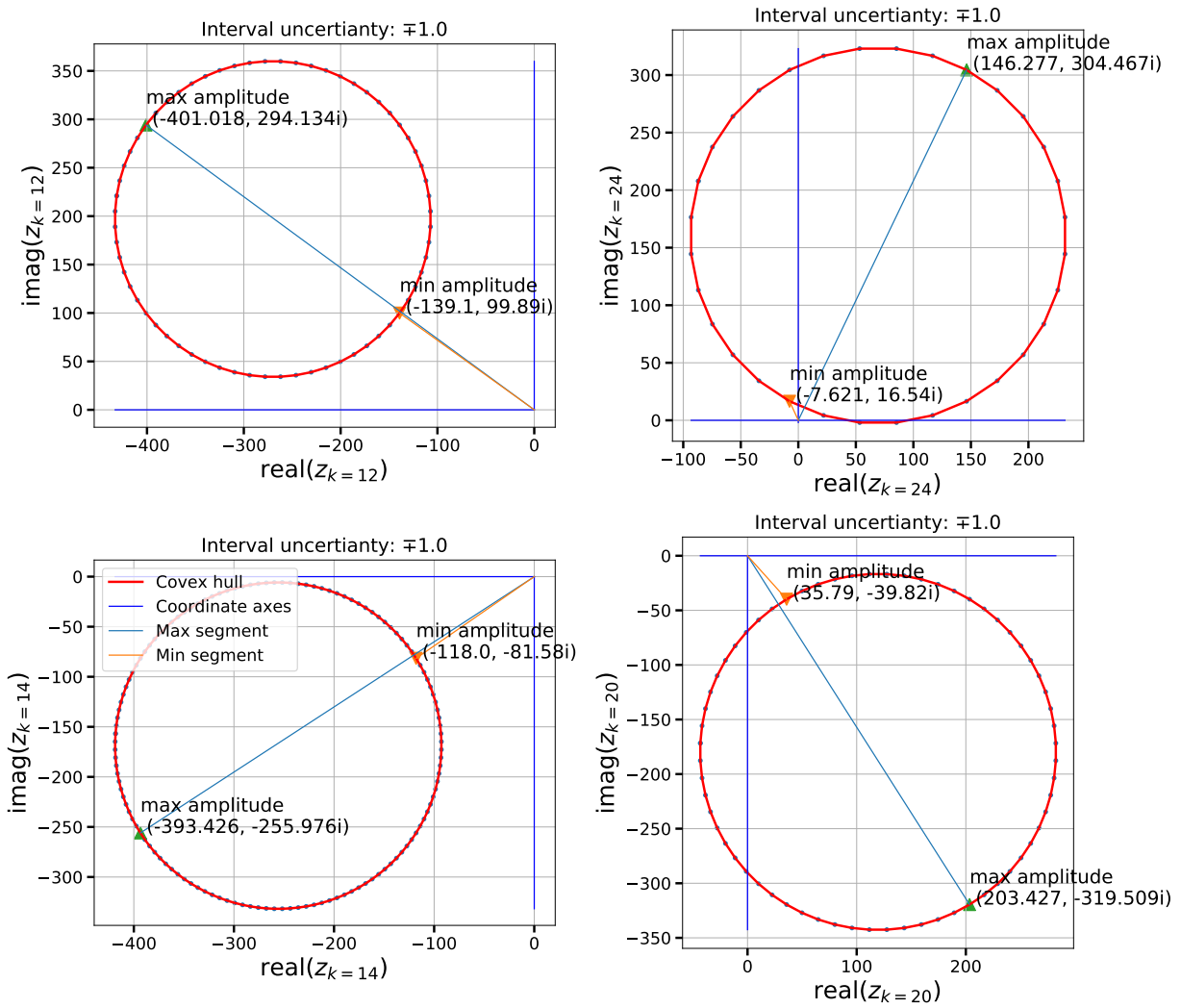


Figure 7.7: Anchoring points on the united set boundary at frequencies $k \in \{12, 14, 20, 24\}$, for a signal with ± 0.1 [units] imprecision. The anchoring points corresponding to maximum and minimum amplitude are depicted with a triangle marker pointing *up* for the maximum and *down* for the minimum.

7.3.3 Obtaining the interval amplitude spectrum

The amplitude spectrum of an arbitrary interval signal is constructed by collecting minimum and maximum amplitude for all frequencies $k \in \{0, \dots, N-1\}$, as shown in the previous section. A few examples of this interval spectrum are shown in Fig. 7.9 for the same simulated signal with different values of interval uncertainty. In Fig. 7.9 the interval amplitude spectrum obtained with the algorithm (united extension), displayed in dark orange, is superimposed on the interval amplitude spectrum obtained with interval arithmetic (interval extension), displayed in light blue. It can be seen from Fig. 7.7 that the number of vertices composing the boundary of the united set differs between frequency numbers. For example, frequency number $k = 24$ has less vertices than frequency number $k = 14$. While the algorithm should produce exactly 2 more vertices at each iteration, for certain frequencies the Fourier coefficient is such that some

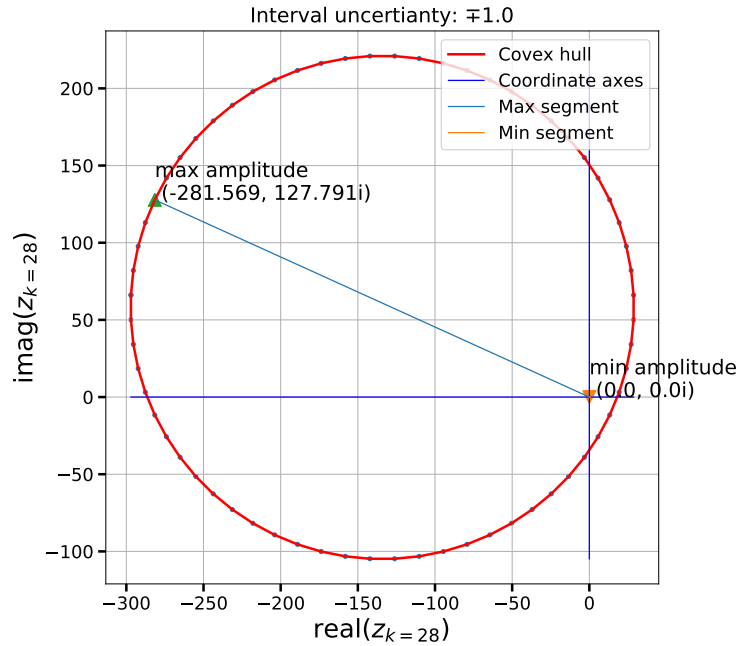


Figure 7.8: An example of a united set (at frequency $k = 28$) that contains the origin of the complex plane. In this case the minimum amplitude is zero. The anchoring points corresponding to maximum and minimum amplitude are depicted with a triangle marker pointing *up* for the maximum and *down* for the minimum.

vertices get projected onto the same vertex. For example, frequency numbers that are a power of two display projections that are always parallel to the coordinate axes, making the resulting united set identified by four vertices only. This behaviour can be appreciated in Fig. 7.9, where the united extension of the spectrum “peaks” at frequencies $k = \{32, 64, 96\}$. Therefore, these peaks are not due to a numerical artefact but are inherent to the process of propagating intervals through a discrete transform. For frequencies that are divisible by two, but have another cofactor like $k = 14 = 2 \cdot 7$ or $k = 24 = 2^3 \cdot 3$, some projections parallel to the coordinate axes make some of the vertices map onto the same one vertex, effectively making those aligned vertices collapse into one. When $N = 2^m$ with $m \in \mathbb{N}$, the Fourier coefficients $\frac{2\pi}{N}kn$ for frequency numbers k that are a power of two: $k = 2^p$, $p \in \mathbb{N}$, $p < m$, correspond to a rotation on the Euler circle with the following angle: $2\pi \frac{n2^p}{2^m} = 2\pi \frac{n}{2^{m-p}}$, which for $p = m - 1$ (or $k = N/2$) is a multiple of π . For such frequency numbers the united set is a rectangle, thus united extension interval extension coincide. This can be appreciated in Fig. 7.9, where for frequency number $k = 64$ the interval extension and the united extension meet in one point in the upper bound.

7.3.4 Computational cost

The selective algorithm (Pseudocode 1) tracks down the endpoints corresponding to the vertices of the united set for each frequency number. This process entails the addition of at most $4k$ complex numbers (twice $2k$ additions). Set aside the convex hull step, a total of at most $4N$ additions are performed at the end of the sum. At each k , however, the selective algorithm computes the convex hull of the set of vertices obtained by the polygon-segment addition de-

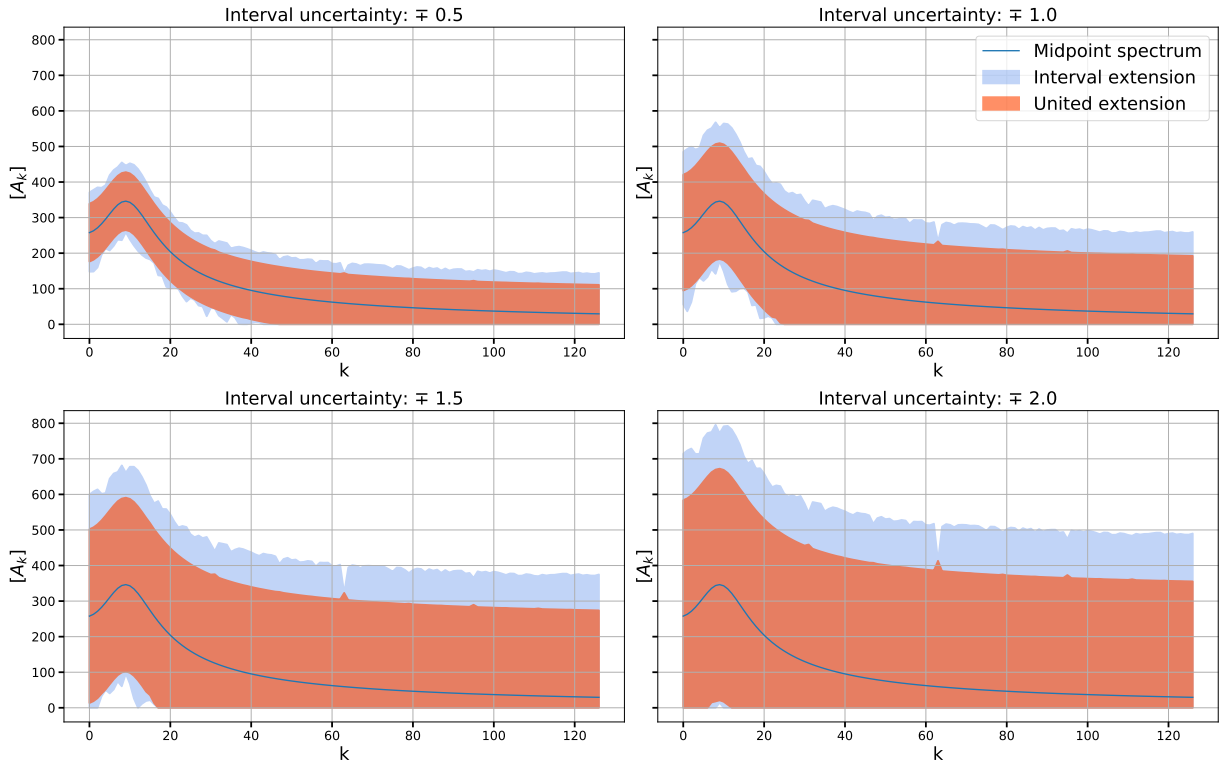


Figure 7.9: Interval Fourier amplitude spectrum obtained with the proposed algorithm (united extension) in dark orange versus interval spectrum obtained with interval arithmetic (interval extension), displayed in light blue, for signals with constant imprecision $\xi = \{0.5, 1.0, 1.5, 2.0\}$.

picted in Fig. 7.5, to eliminate the vertices that lie in the interior of the resulting polygon. Because of the convex hull, each addition step will cost at least $4k \log 4k$. In total, because $\sum_k^N k = \frac{N(N+1)}{2} \leq N^2$, $N = 1, 2, \dots$, the cost of obtaining the boundary points is at most $(4N \log 4N)^2$ for each frequency number k . For a signal of length N —with $N = 2^m$ a power of two—the DFT outputs exactly $N/2$ frequencies, because the matrix of Fourier coefficients is symmetric and orthogonal. Hence, overall the cost of running the selective algorithm is at most $\frac{N}{2} (4N \log 4N)^2 = 8N^3 \log^2 4N$. Using the big O notation, the cost of the selective algorithm is $O(N^3 \log^2 N)$, which means that the interval DFT can be performed in polynomial time.

In contrast, an exhaustive search of all the endpoints of the interval signal (described in Pseudocode 3) would require 2^k complex additions, so a total of $N2^{2N}$ floating-point operations. In big O notation such a brute-force algorithm costs $O(2^N)$, which makes it unusable in practice. The evaluation of the interval extension (described in Pseudocode 2), which yields the bounding box that circumscribes the united set, here used for verification, need not compute a convex hull at each addition, so it is much cheaper. The cost of obtaining the interval extension is thus equivalent to the standard DFT algorithm with the only exception that each operation is done within interval analysis. When interval multiplications (or divisions) are involved the cost of the intervalised algorithm is at most four times (eight times with complex intervals) the cost of the standard DFT algorithm. When there are only interval additions (or subtractions) and

non-interval multiplications, the cost of an intervalised algorithm is at most two times (four times with complex intervals) the cost of the standard DFT algorithm. In summary, the cost of the intervalised DFT is: $\frac{1}{2} N 4N = 2N^2$, which in big O notation is $O(N^2)$. Given that the DFT algorithm runs in $O(N^2)$, there is no appreciable additional cost in running its intervalised version.

An overview of the computational costs for the individual algorithms is provided in Table 7.1.

Table 7.1: Summary of computational costs

Algorithm	Computational costs
Interval extension	$O(N^2)$
Selective	$O(N^3 \log^2 N)$
Brute-force tracker	$O(2^N)$

7.4 Technical application

In order to demonstrate the propagation of intervals through the DFT, the pile foundation of an offshore wind turbine excited by sea waves is considered. This model was chosen because it is easy to understand, especially to demonstrate the proposed method, but still has the characteristics of a real structure. An exemplary illustration of the offshore wind turbine model is shown in Fig. 7.10. The total height of the structure is denoted by h_{pile} . Since the structure is hollow from the inside, the cross-section of the structure is specified. The outer radius R and the inner radius r are important parameters to describe the wind turbine. The structure is subjected to the continuous changes of the water height caused by the sea waves. When the water is at rest, the water height is h_{water} , which is indicated by the dashed line. The change in water height due to the sea waves is expressed by $h_{wave}(t)$, which is a time-dependent variable. The total water height including the sea waves is thus $H(t) = h_{water} + h_{wave}(t)$.

This structure will be idealised as a Single-Degree-of-Freedom (SDOF) mass-spring-damper system and described by means of the following equation of motion

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = F(t), \quad (7.13)$$

where m describes the mass of the system, c is the damping coefficient and k denotes the spring constant. x , \dot{x} and \ddot{x} describe the displacement, velocity and acceleration of the system, respectively. The external excitation of the system is $F(t)$. The following structural parameters are assumed for the wind turbine. The mass m is defined as

$$m = \rho_{steel} h_{pile}(R^2 - r^2)\pi + \text{CONST} = 1.71 \cdot 10^6 + \text{CONST} \quad [\text{kg}], \quad (7.14)$$

with $\rho_{steel} = 7800 \text{ kg/m}^3$ as the density of steel, $h_{pile} = 60 \text{ m}$ as the total height of the offshore

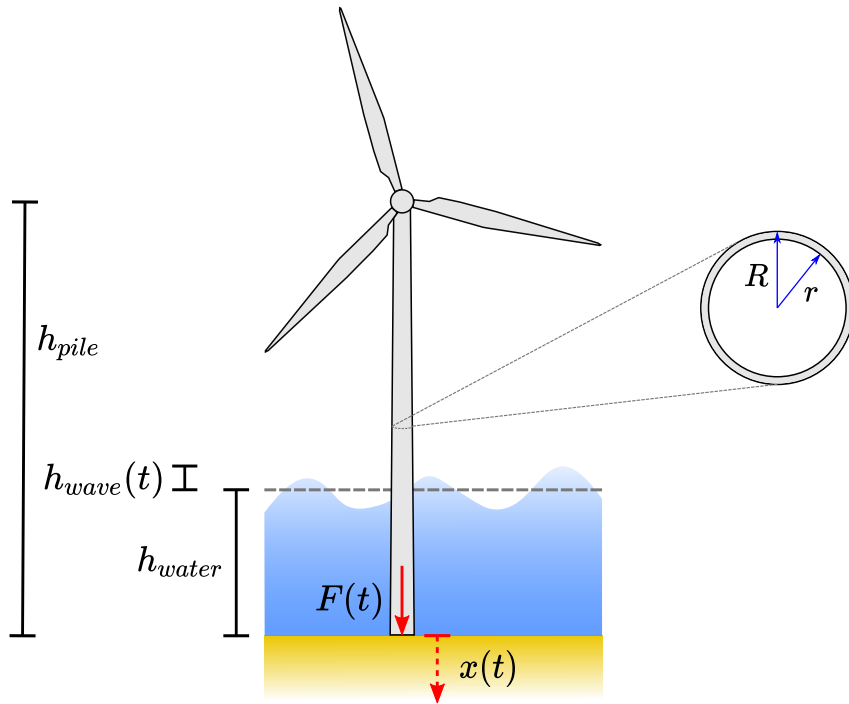


Figure 7.10: Exemplary illustration of the offshore wind turbine model used in this work.

pile, $R = 3$ m is the outer radius of the pile and $r = 2.8$ m is the inner radius of the hollow pile. The constant CONST in Eq. (7.14) denotes the weight of any other components of the offshore wind turbine, such as the turbine itself, the blades and other technical devices and equipment. Furthermore, the stiffness is $k = 10^6$ N/m and the damping coefficient is $c = 10^5$ Ns/m.

7.4.1 Modelling of the excitation by sea waves

The dynamic behaviour of the system is caused by the sea waves. The waves continuously change the water depth and thus the water pressure and the resulting buoyant force on the pile foundation. To model the dynamic behaviour of sea waves in the frequency domain, a PSD function derived within the Joint North Sea Wave Observation Project (JONSWAP) [89] is used, which describes an extension of the Pierson-Moskowitz PSD function [88] and reads as follows

$$S(\omega) = \frac{\alpha g^2}{\omega^5} \exp\left(-\frac{5}{4} \left(\frac{\omega_p}{\omega}\right)^2\right) \gamma^r \quad (7.15)$$

with

$$r = \exp\left(\frac{-(\omega - \omega_p)^2}{2\sigma^2\omega_p^2}\right). \quad (7.16)$$

In these equations α describes a spectral energy parameter, g is the gravity acceleration, ω_p describes the peak frequency, γ^r is the peak enhancement factor and σ the spectral width parameter. An example for the JONSWAP PSD function with $\alpha = 0.0081$, $w_p = 0.7$, $\gamma = 3.3$ and

$$\sigma = \begin{cases} 0.7 & \omega \leq \omega_p \\ 0.9 & \omega > \omega_p \end{cases}$$

is given in Fig. 7.11 (left).

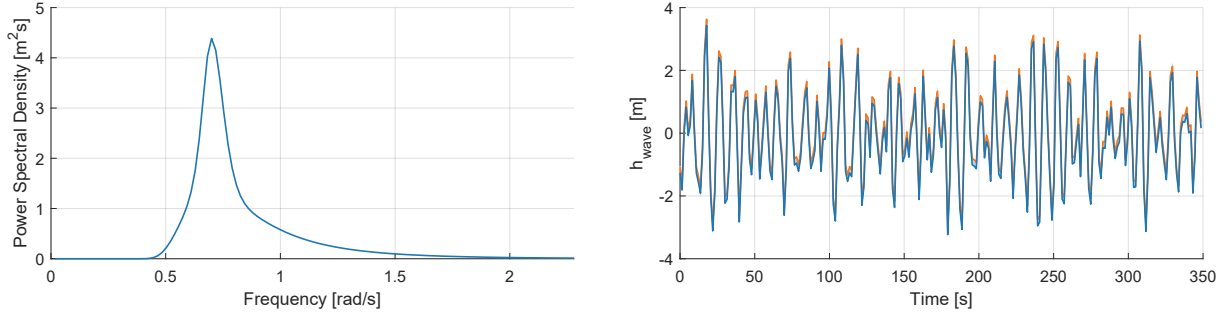


Figure 7.11: Example for the JONSWAP PSD function (left) and a generated signal of wave heights with interval uncertainty $\xi = 0.1$ m.

The JONSWAP PSD function (Eq. 7.15) is utilised to generate an artificial signal for the height of the sea waves, which will be the signal under consideration. The spectral representation method (SRM), given in Eq. (7.17), allows to generate stochastic processes from a given PSD function S_X [84]. It provides a suitable method for generating compatible time signals derived from and carrying the characteristics of an underlying PSD function. The method reads as follows

$$X_t = \sum_{n=0}^{N-1} \sqrt{4S_X(\omega_n)\Delta\omega} \cos(\omega_n t + \varphi_n), \quad (7.17)$$

with

$$\omega_n = n\Delta\omega, \quad n = 0, 1, 2, \dots, N-1 \quad (7.18)$$

where $N \rightarrow \infty$, t as time vector and φ_n as uniformly distributed random phase angles in the range $[0, 2\pi]$.

For generating the signal X_t with total length of $T = 350$ s, the time step size $\Delta t = 2\pi/(2\omega_u)$ and frequency step size $\Delta\omega = (2\pi)/T$ are determined according to [84]. The upper cut-off frequency is defined as $\omega_u = 2.2975$ rad/s. Thus, $\Delta t = 1.367$ s and $\Delta\omega = 0.018$ rad/s. To emulate the poor precision of the resulting signal X_t , it is intervalised using Eq. (7.6) with interval uncertainty $\xi = 0.1$ m. The entire interval signal \overline{X}_t is then represented by an interval at each point in time, see Fig. 7.11 (right). This step is performed here to simulate a real signal with poor precision.

7.4.2 Computing the interval PSD function and the interval system response

The intervalised signal \overline{X}_t of the wave height is transformed to the frequency domain as explained in Section 7.3 by means of the interval extension and the united extension using Eq. (7.11). Pushing the interval signal \overline{X}_t through Eq. (7.11) using the algorithm described in Section 7.3 yields the exact bounds on \hat{S}_X ,

$$\hat{S}_X(\omega_k) = \frac{\Delta t^2}{T} A_k^2 \quad (7.19)$$

because $A_k \geq 0$ is non-negative by definition, see Eq. (7.8), and shows no dependence with T and Δt .

The bounds on the interval PSD function $\overline{\hat{S}_X}$ are shown on the left side of Fig. 7.12. For comparison, the midpoint spectrum is also given, which is calculated via the ordinary DFT from the signal without interval uncertainty. This procedure allows the uncertainties of the interval signal \overline{X}_t to be quantified in the frequency domain and to obtain an upper and lower bound for the respective frequency components.

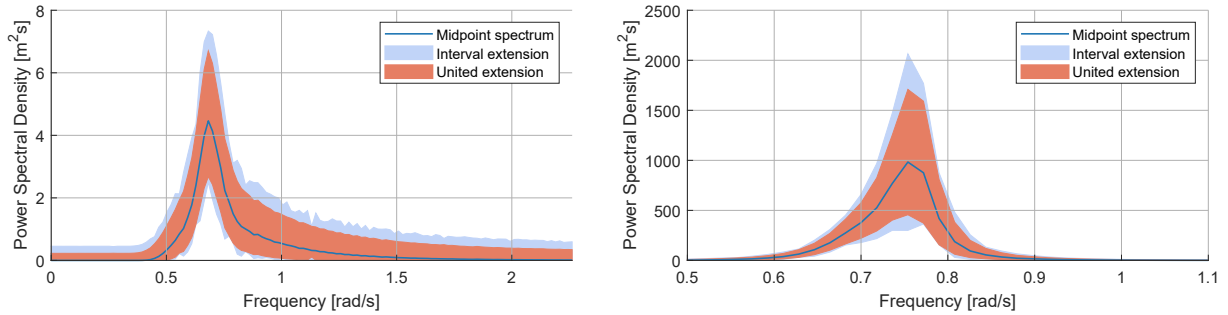


Figure 7.12: Bounds obtained by the interval extension and the united extension of the signal with interval uncertainty $\xi = 0.1$ m (left) and bounds on the resulting PSD function of the response (right) computed by Eq. (7.20). The bounds derived by interval extension are displayed in light blue, the exact bounds derived by the united extension in dark orange.

After the excitation of the wind turbine by the sea waves, a displacement of the pile foundation can be detected. The PSD function of the system response indicates the frequency components and their amplitude corresponding to the system displacement $x(t)$ of the pile foundation in the frequency domain. The system response is significantly dependent on the excitation and the system parameters and can be utilised to detect critical system behaviour or system failure. Instead of a numerical solution obtained by extensive MC simulation in time domain, the response of the pile foundation in frequency domain can be directly attained by the the frequency response function (e.g., [52, 124])

$$S_Y(\omega) = S_X(\omega)|H(\omega)|^2 \quad (7.20)$$

with $H(\omega)$ as transfer function

$$H(\omega) = \frac{1}{\omega_0^2 - \omega + 2\zeta\omega_0\omega i}. \quad (7.21)$$

In these equations $\omega_0 = \sqrt{k/m}$ describes the natural frequency of the system, $\zeta = c/(2\omega_0 m)$ is the damping ratio and i is the imaginary unit, while the PSD function of the excitation is denoted by S_X .

The natural frequency of the system thus results in $\omega_0 = 0.7657$ rad/s and the damping ratio is

$\zeta = 0.0383$. The constant CONST in Eq. (7.14) will be neglected for illustration purposes. The upper and lower bounds of the interval PSD function of the response $\widehat{\underline{S}}_Y$ can be determined separately by interval arithmetic using the singleton-interval multiplication (Eq. 7.3) for the frequency response function (Eq. 7.20) on the upper and lower bounds of the interval PSD function $\widehat{\underline{S}}_X$. The uncertainty in the original signal of wave heights is thus propagated to the PSD function of the response $\widehat{\underline{S}}_Y$ and determined by their upper and lower bounds, see Fig. 7.12 (right). The determination of a bounded system response in the frequency domain thus leads to a significant improvement in the evaluation of system responses in general, taking into account the uncertainties in the input signal. For each frequency component, an upper and lower bound is now identified in the response, which can be employed to determine whether possible responses are within an acceptable window or result in dangerous system behaviour. This procedure thus makes it possible to propagate the uncertainty of the input signal to the system response and consequently to detect critical or dangerous system behaviour taking these uncertainties into account. In particular, the range of natural frequencies of a system can thus be assessed more accurately in this way.

7.4.3 Comparison with Monte Carlo

To determine the uncertainty of an interval signal, random signals within the bounds of the interval signal are usually sampled using Monte Carlo (MC). In this example, a signal with constant interval uncertainty $\xi = 0.5$ m is used. Within the interval bounds of the signal \overline{X}_t , any arbitrary signal is possible due to the definition of intervals. For illustration, a set of 20 sampled signals within the bounds is shown in Fig. 7.13. Each of these realisations is transformed separately using Eq. (7.10) to obtain the PSD function. The maximum and minimum at each frequency can then be determined from all individual transforms to obtain the envelope. These are the extrema of the PSD function determined via MC. A comparison of these extrema and the bounds determined by the interval and united extension is depicted in Fig. 7.13. To obtain the MC extrema a total of 10^6 sample signals were generated and transformed.

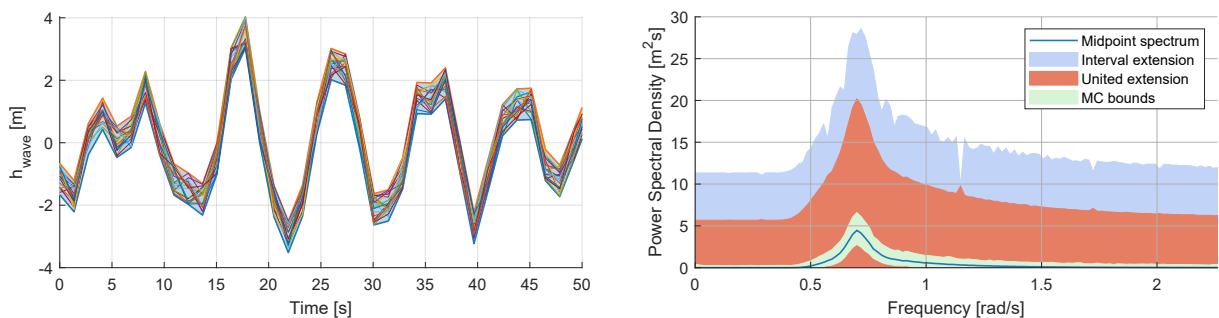


Figure 7.13: Signal with interval uncertainty $\xi = 0.5$ m and randomly sampled signals within the bounds (left), resulting PSD function bounds obtained by the interval extension in light blue, the united extension in dark orange and MC with 10^6 samples in light green (right).

It can be clearly seen that the minimum/maximum determined by MC are clearly suboptimal

compared to the bounds determined by the united extension and interval extension. Since the united extension produces exact bounds, it can be concluded that the determination of the bounds with MC severely underestimates the uncertainty of the signal, which leads to major problems and potential hazards in reliability assessment.

The MC method is tasked to identify the local minimum and maximum while complying with the interval constraints. The number of MC samples determines how close these two extrema are to the exact bounds. In principle, the more MC samples are utilised the closer the minimum and maximum are to the exact bounds. However, the number of MC samples that provides enough coverage can be prohibitively high, as shown in Fig. 7.13, where 10^6 MC samples were evaluated. In Fig. 7.14, the PSD functions were estimated for 3, 10, 100 and 500 sampled signals to show the progression of coverage as the number of MC samples increases. In Fig. 7.15 the envelopes of 10^1 , 10^3 and 10^6 MC samples are compared against each other. This problem of coverage for the MC method has been presented for example in [219, 220]. A numerical optimization study has also been conducted using *fmincon* in Matlab. The minimum/maximum found by the optimizer are significantly better than those obtained with Monte Carlo, whilst also requiring a lot less DFT evaluations. Nevertheless, the minimum and maximum outputted by the optimiser are only known to be local optima, so they are not guaranteed to approximate the exact bounds to the desired accuracy in general.

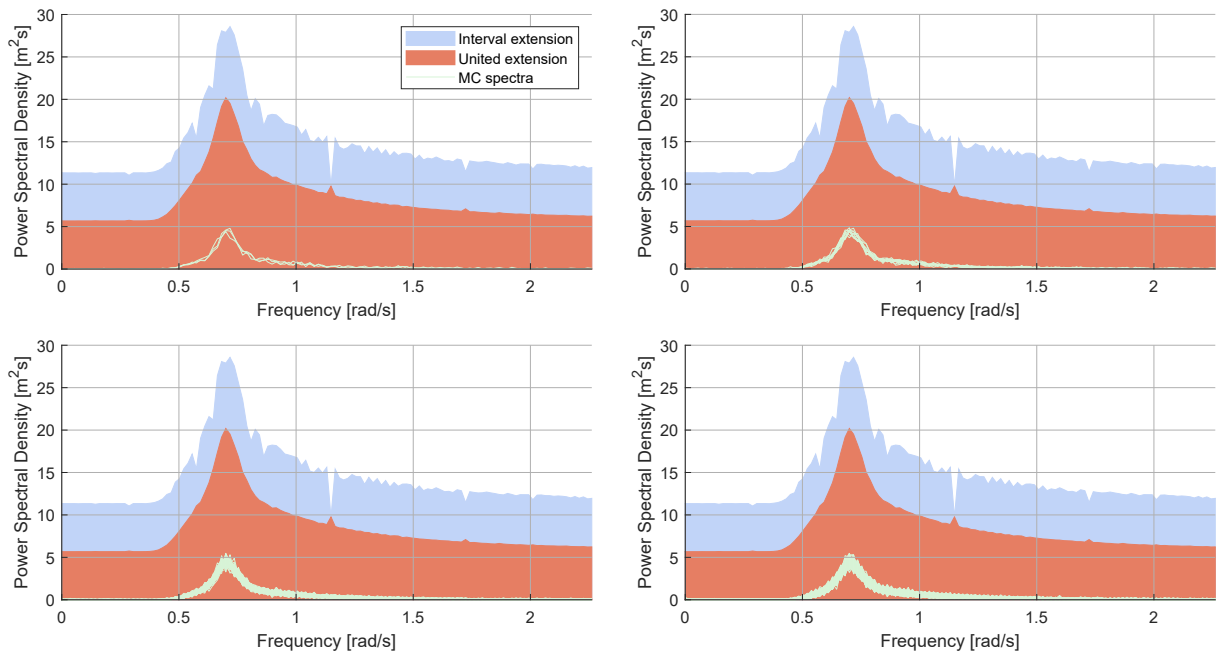


Figure 7.14: Estimated PSD functions for 3, 10, 100 and 500 sampled signals, against interval PSD function, for a signal with interval uncertainty $\xi = 0.5$ m.

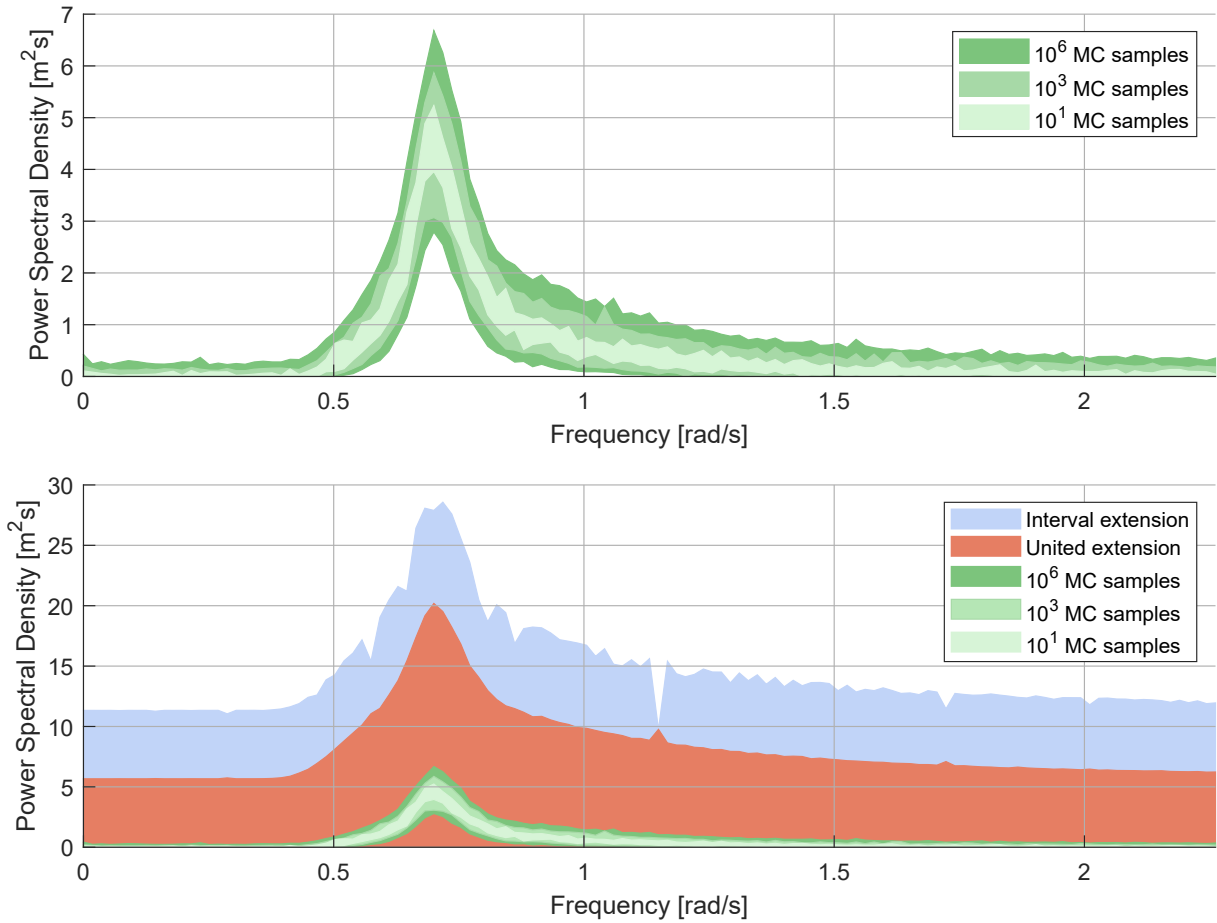


Figure 7.15: Envelopes of estimated PSD functions for 10^1 , 10^3 and 10^6 sampled signals for a signal with interval uncertainty $\xi = 0.5$ m (top) compared against the bounds obtained by the interval extension in light blue and the united extension in dark orange (bottom).

7.4.4 Sensitivity analysis

To investigate the influence of the interval uncertainty ξ of the signal on the interval width of the PSD functions, a sensitivity analysis is carried out in this section. The interval uncertainty of the signal used in the previous sections was successively increased by 0.1 up to 5.0 and the Fourier transform with interval extension and united extension were calculated. For comparison, the interval width of the spectra at the peak frequency of $\omega_0 = 0.7657$ rad/s, which is the natural frequency of the system, is used. The results for both proposed methods are shown in Fig. 7.16 (left). It should be noted that for an interval uncertainty of $\xi \geq 0.4$, the lower bound is zero for the interval extension. This is shown in Table 7.2, where some selected values of the interval bounds are presented and is also visible in Fig. 7.16 (right). The same applies for an interval uncertainty of $\xi \geq 0.5$ for the united extension. Accordingly, the interval width grows only by the increase of the upper bounds. From Fig. 7.16 (right) a linear trend between widths can be appreciated up to $\xi \geq 0.5$. For values of imprecision greater than $\xi \geq 0.5$, for which the lower bound is always zero, a non-linear trend between widths can be appreciated instead. Such

uncertainty growth can have significant impact on the analysis, since there could be a potentially quadratic increase of the energy content in the PSD function as the uncertainty of the signal grows.

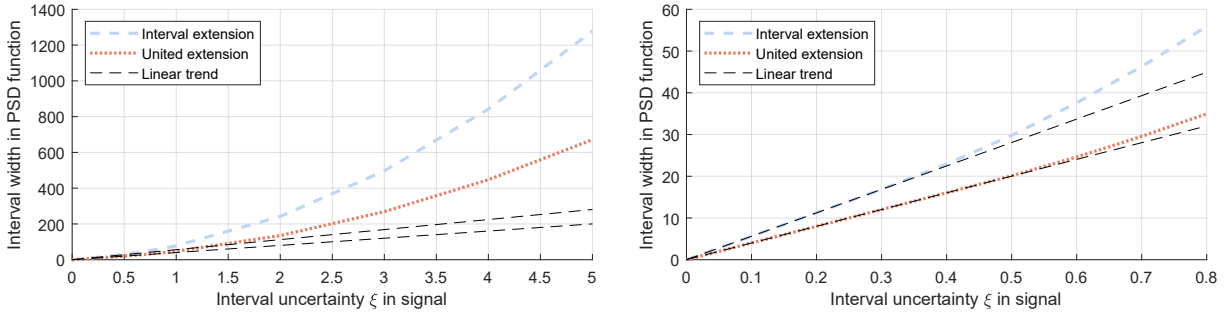


Figure 7.16: Influence of the interval uncertainty ξ of the signal on the interval width of the PSD function at the peak frequency $\omega_0 = 0.7657$ rad/s (left) and linear trend for smaller interval uncertainties in the signal (right).

Table 7.2: Interval width and bounds at the peak frequency $\omega_0 = 0.7657$ rad/s for the interval extension and the united extension of the PSD function \widehat{S}_X .

ξ	Interval extension		United extension	
	width (\widehat{S}_X)	\widehat{S}_X	width (\widehat{S}_X)	\widehat{S}_X
0	0	[4.4, 4.4]	0	[4.4, 4.4]
0.1	5.7	[2.0, 7.7]	4.1	[2.6, 6.7]
0.2	11.3	[0.6, 11.9]	8.1	[1.3, 9.4]
0.4	22.9	[0, 22.9]	16	[0, 16.0]
0.8	55.9	[0, 55.9]	35	[0, 35.0]
1	77.8	[0, 77.8]	47.1	[0, 47.1]
2	241.9	[0, 241.9]	135.2	[0, 135.2]
3	496.7	[0, 496.7]	268.6	[0, 268.6]
4	842.1	[0, 842.1]	447.3	[0, 447.3]
5	1278	[0, 1278]	671.3	[0, 671.3]

7.5 Conclusions

Especially in engineering, the quantification of uncertainties is a matter of paramount importance. In particular, it can be argued that real recorded data are hardly ever exact, being subject to uncertainties due to a wide variety of factors. The algorithm presented in this paper allows interval signals to be projected through the DFT to obtain an upper and lower bound on

the Fourier amplitude and an estimation of the PSD function, which has a significant impact on the quantification of uncertainty present in time signals. The propagation of signals can be computed using interval arithmetic, which generally provides outer bounds, as opposed to exact. An algorithm that characterises the geometry of the united set of the DFT is presented. The algorithm fully addresses the repeated variables problem arising in the calculation of the amplitude and therefore yields the exact bounds. Both the interval and the united extension were compared against the MC method. It was shown that the extrema derived via MC severely underestimate the uncertainty. The proposed algorithm is thus more adequately able to account for the uncertainties of signals with poor precision in the Fourier domain. One advantage of this algorithm is that no assumptions have to be made about the error distribution. This leads to a significant improvement in the assessment of engineering problems regarding risk, reliability and uncertainty. The strengths and advantages of the algorithm were illustrated by means of an example involving dynamic structural analysis, in which the bounds of the PSD estimation of an interval signal were calculated. These bounds can be propagated through the system to quantify the response behaviour and the impact of the signal's inherent uncertainty on the quantity of interest.

7.6 Replicability

The software for computing the interval DFT can be accessed in a single instance via GitHub at: <https://github.com/interval-fourier-transform/DFT> (last accessed February 2022); whilst the code running the presented examples is available at: <https://github.com/interval-fourier-transform/application-to-poor-precision-time-signals> (last accessed February 2022). The code, examples and numerical results presented in this paper are therefore fully replicable.

Acknowledgement

This research is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) with grant no. BE 2570/4-1 and CO 1849/1-1 and by the Engineering & Physical Sciences Research Council (EPSRC) with grant no. EP/R006768/1. DFG and EPSRC are greatly acknowledged for their funding and support.

Appendix A. Pseudocodes

The provided Pseudocodes 1 and 2 describe the computation of the interval Fourier amplitude of an interval signal \underline{x} for a specific frequency k with the selective algorithm and with the interval extension using interval arithmetic. Pseudocode 3 describes the tracking of all endpoints for each iteration step. This code is mainly used for illustration purposes and as the foundation of the selective algorithm.

Pseudocode 1: This algorithm, also called the selective algorithm, calculates the Fourier amplitude of an interval signal \underline{x} for any frequency k by determining the convex hull of the endpoints. First, the length N of the interval signal is determined. The first two vertices v are calculated by computing the first coefficient for $n = 0$ of the DFT for the first segment of the signal for the upper and lower interval bounds. These are stored in c_{hull} and describe the first two vertices of the endpoint analysis.

Within the for loop, the next vertices v are iteratively computed by calculating the next coefficients for $n = 1, \dots, N - 1$. These two vertices v are added separately to the list c_{hull} to get a new list of endpoints, which are stored in the variable ep . The convex hull is calculated from ep , where the real and imaginary components represent the coordinates in \mathbb{R}^2 . The vertices forming the convex hull are stored in c_{hull} and all other endpoints are discarded. This procedure is repeated for each n until the last iteration $n = N - 1$ is reached, thus c_{hull} constitutes the convex hull after each iteration.

Since the minimum and maximum distance from the convex hull to the origin of the coordinate system determine the interval bounds, these are obtained from the absolute values of c_{hull} . If the origin is inside the convex hull, the lower bound is 0, otherwise it is defined by $\min(|c_{hull}|)$. The upper bound is always defined by $\max(|c_{hull}|)$.

The algorithm returns the upper and lower bound on the Fourier amplitude \overline{A}_k for the respective frequency k .

Pseudocode 2: In this algorithm interval arithmetic is utilised to achieve the bounds on the Fourier amplitude. This algorithm is mainly used as a verification algorithm for the algorithm presented in Pseudocode 1.

First, the interval extension of the DFT (Eq. 7.7) of the interval signal \underline{x} for the frequency k is determined and stored in \underline{z}_k . From the complex interval \underline{z}_k the absolute values of the vertices of the bounding box are calculated. Those are defined by the 4 possible combinations of upper and lower bound for real and imaginary parts of \underline{z}_k : $d_{ll} = |\text{Re}(z_k) + i \text{Im}(z_k)|$, $d_{lh} = |\text{Re}(z_k) + i \text{Im}(\bar{z}_k)|$, $d_{hl} = |\text{Re}(\bar{z}_k) + i \text{Im}(z_k)|$ and $d_{hh} = |\text{Re}(\bar{z}_k) + i \text{Im}(\bar{z}_k)|$.

Since the minimum and maximum distances to the origin of the coordinate system define the interval bounds, these will be determined: The upper bound is the maximum value of all distances to the origin described by $\overline{A}_k = \max(d_{ll}, d_{lh}, d_{hl}, d_{hh})$. The lower bound is determined by considering different cases: If the origin of the coordinate system is inside \underline{z}_k , 0 is determined as the lower bound. If $\text{Re}(z_k) \leq 0 \leq \text{Re}(\bar{z}_k)$, then the lower bound will be defined as $\underline{A}_k = \min(|\text{Im}(z_k)|, |\text{Im}(\bar{z}_k)|)$. If $\text{Im}(z_k) \leq 0 \leq \text{Im}(\bar{z}_k)$, the lower bound will be $\underline{A}_k = \min(|\text{Re}(z_k)|, |\text{Re}(\bar{z}_k)|)$, and otherwise the lower bound is determined to be $\underline{A}_k = \min(d_{ll}, d_{lh}, d_{hl}, d_{hh})$.

The algorithm returns an outer approximation of the interval bounds \overline{A}_k of the Fourier amplitude for a given frequency k .

Pseudocode 3: This algorithm describes the tracking of all endpoints using a brute-force method. The idea behind the code is the same as in Pseudocode (1). The difference is that

the convex hull is not determined in each iteration and therefore no endpoints are discarded. Instead, all endpoints ep are stored in a binary *tree* for all N and are passed on to the next iteration. During this procedure the number of endpoints becomes significantly larger than in the selective algorithm, as the number of endpoints increases exponentially in base two and with exponent given by the iteration number. The total number of iterations is determined by length N of the interval signal. The resulting bounds for the Fourier amplitude are the same as in Pseudocode (1), but for practical simulations this code is not feasible as it is too inefficient to be used for long signals. Therefore, this code is herein mainly used for illustration purposes and as the foundation of the selective algorithm.

Pseudocode 2 Verification algorithm with rigorous interval arithmetic

```

1: function INTERVALAMPLITUDEBOUNDS( $\underline{x}$ ,  $k$ )
2:    $\underline{z}_k \leftarrow \text{DFT}(\underline{x}, k)$        $\triangleright$  interval extension of the discrete Fourier transform (Eq. 7.7) for frequency  $k$ 
3:    $d_{ll} \leftarrow |\text{Re}(z_k) + i \text{Im}(z_k)|$ 
4:    $d_{lh} \leftarrow |\text{Re}(z_k) + i \text{Im}(\bar{z}_k)|$ 
5:    $d_{hl} \leftarrow |\text{Re}(\bar{z}_k) + i \text{Im}(z_k)|$ 
6:    $d_{hh} \leftarrow |\text{Re}(\bar{z}_k) + i \text{Im}(\bar{z}_k)|$        $\triangleright d_{ll}, d_{lh}, d_{hl}$  and  $d_{hh}$  define distances of the vertices of the bounding
      box to the origin
7:    $\bar{A}_k \leftarrow \max(d_{ll}, d_{lh}, d_{hl}, d_{hh})$        $\triangleright$  get maximum distance to origin
8:   if origin is inside  $\underline{z}_k$  then       $\triangleright$  get minimum distance to origin depending on different cases
9:      $\underline{A}_k \leftarrow 0$ 
10:  else if  $\text{Re}(z_k) \leq 0 \leq \text{Re}(\bar{z}_k)$  then
11:     $\underline{A}_k \leftarrow \min(|\text{Im}(z_k)|, |\text{Im}(\bar{z}_k)|)$ 
12:  else if  $\text{Im}(z_k) \leq 0 \leq \text{Im}(\bar{z}_k)$  then
13:     $\underline{A}_k \leftarrow \min(|\text{Re}(z_k)|, |\text{Re}(\bar{z}_k)|)$ 
14:  else
15:     $\underline{A}_k \leftarrow \min(d_{ll}, d_{lh}, d_{hl}, d_{hh})$ 
16:  end if
17:   $\bar{A}_k \leftarrow \text{Interval}(\underline{A}_k, \bar{A}_k)$ 
18:  return  $\bar{A}_k$        $\triangleright$  bounds on the Fourier amplitude for frequency  $k$ 
19: end function

```

Appendix B. Anchoring points

The maximum amplitude of the DFT is the maximum distance from the boundary of the united set to the origin of the complex plane. This holds under the interpretation that a complex number is a point in \mathbb{R}^2 , whose coordinates are the real and imaginary components, and its complex absolute value is the Euclidean distance of that point from the origin. Because the boundary of the united set consists of a finite number of vertices, in this Appendix, it will be sufficient to show that the farthest point to the origin of the plane is attained at the vertices and not on any of the connecting edges.

Proposition 1. *Let $A, B \in \mathbb{R}^2$ two points on the real plane, O the origin of the plane, and \overrightarrow{AB} the segment connecting the two points. For every point $P \in \overrightarrow{AB}$, it holds that*

Pseudocode 3 Tracking of all the endpoints

```

1: function ENDPNTSTREETRACKER( $\bar{x}$ ,  $k$ )
2:    $N \leftarrow l(\bar{x})$  ▷ length of interval signal
3:    $v \leftarrow [e^{-i2\pi 0k/N} \underline{x}[0], e^{-i2\pi 0k/N} \bar{x}[0]]$  ▷ compute first two vertices
4:    $ep \leftarrow v$  ▷ initialise endpoints
5:    $tree[0] \leftarrow ep$ 
6:   for  $n \leftarrow 1$  to  $N - 1$  do
7:      $v \leftarrow [e^{-i2\pi nk/N} \underline{x}[n], e^{-i2\pi nk/N} \bar{x}[n]]$  ▷ compute subsequent vertices
8:      $ep \leftarrow$  add the two vertices  $v$  separately to the list  $ep$  to get a list of new endpoints
9:      $tree[n] \leftarrow ep$  ▷ store  $n$ th level of endpoints  $ep$  in  $tree$ 
10:  end for
11:  return  $tree$  ▷ endpoints of the Fourier sum for each iteration
12: end function

```

$$\|\vec{OP}\| \leq \max(\|\vec{OA}\|, \|\vec{OB}\|). \quad (7.22)$$

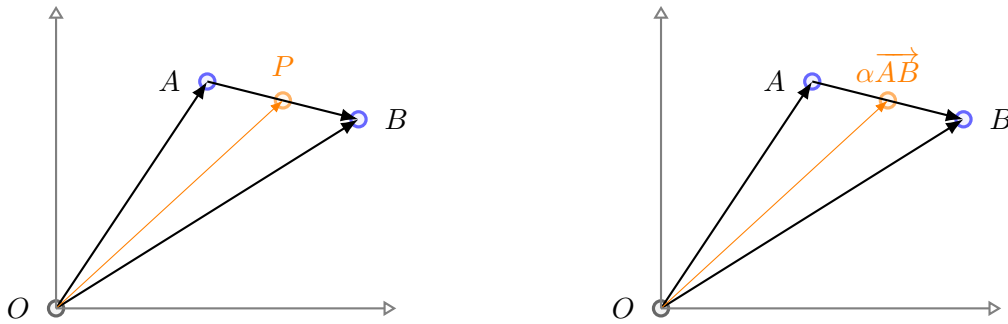


Figure 7.17: Two vertices on the boundary of the united set. On the right, the point P in the interior is obtained multiplying the vector \vec{AB} times a scalar $\alpha \in [0, 1]$.

Proof. It is well-known in Linear Algebra that the difference between two points yields the oriented segment between the two points:

$$\vec{OB} - \vec{OA} = \vec{AB}. \quad (7.23)$$

Any point $P = \vec{OP}$ in the interior of the segment \vec{AB} , see Figure 7.17, can be expressed as:

$$\vec{OP} = \vec{OA} + \alpha \vec{AB}, \quad \text{where } \alpha \in [0, 1]. \quad (7.24)$$

Therefore, any point in the interior of the segment is

$$P = \vec{OP} = \vec{OA} + \alpha (\vec{OB} - \vec{OA}), \quad (7.25)$$

Assuming that $\|\vec{OB}\| \geq \|\vec{OA}\|$, so that $\|\vec{OB}\|$ is maximum. Then, primarily because $\alpha \in [0, 1]$, it holds

$$(1 - \alpha) \|\vec{OA}\| + \alpha \|\vec{OB}\| \leq \|\vec{OB}\|. \quad (7.26)$$

The left-hand side of (7.26) satisfies the Triangle Inequality:

$$\left\| (1 - \alpha)\vec{OA} + \alpha \vec{OB} \right\| \leq (1 - \alpha) \left\| \vec{OA} \right\| + \alpha \left\| \vec{OB} \right\|. \quad (7.27)$$

Therefore, it holds

$$\left\| (1 - \alpha)\vec{OA} + \alpha \vec{OB} \right\| \leq \left\| \vec{OB} \right\|. \quad (7.28)$$

If the left-hand side of (7.28) is rearranged as follows, the original statement is proven:

$$\left\| \vec{OA} + \alpha (\vec{OB} - \vec{OA}) \right\| = \left\| \vec{OP} \right\| \leq \left\| \vec{OB} \right\|. \quad \blacksquare \quad (7.29)$$

□

8 | Assessing the severity of missing data problems with the interval discrete Fourier transform algorithm

The contribution in this chapter is a further investigation of the interval DFT algorithm (Chapter 7). While it was previously assumed that signals can be uncertain due to various reasons and influences and were therefore represented as interval signals, this part considers the problem of missing data in more detail. Often a failed sensor or corrupted data is the reason for missing data in a data record. This is a common problem when dealing with real data sets, as it may have a significant impact on the simulation results. Especially in spectral analysis, where a signal is transformed from time domain to frequency domain to determine its spectral properties, a signal with missing data can yield misleading results. There are several methods available to reconstruct the gaps in the signal, e.g. least squares or autoregressive methods, but none of these methods can guarantee an exact reconstruction. Furthermore, an incorrect reconstruction can lead to undesired side effects and the signal transformed into the frequency domain, i.e., the PSD function, may not reflect the actual characteristics of the process.

In this part of the thesis, the interval DFT algorithm described in the previous chapter is examined for its capabilities on signals with missing data. The missing data are reconstructed using only two very simple and conservative reconstruction methods, as the objective of the part is to investigate the interval DFT algorithm on signals with missing data and not to find a particularly suitable reconstruction method. The gaps are reconstructed with intervals to obtain a certain margin of safety. The interval DFT algorithm is then tested in several case studies: the sensitivity to the interval uncertainty in general, the influence of the number of missing data, the influence of the gap size and the distribution of the missing data within the signal are investigated.

The results show that the interval DFT algorithm can cope with missing data signals. In general, the algorithm can transform any signal, regardless of the interval uncertainty or the size of the missing data gap. However, it is important to ensure that appropriate methods and/or interval widths are used to reconstruct the missing data. If the interval uncertainty is too large, the bounds in the frequency domain are correspondingly large as well, such that the quantification of the uncertainties is correct but no longer applicable in practical examples or realistic cases. However, if the reconstruction methods and the interval uncertainties are chosen appropriately for the given problem, the algorithm provides results that are useful in practice. A quantification of the uncertainties of an input signal is therefore enabled by the interval DFT algorithm.

Assessing the severity of missing data problems with the interval discrete Fourier transform algorithm

Marco Behrendt^a, Marco de Angelis^b, Liam Comerford^b, Michael Beer^{a,b,c}

^aInstitute for Risk and Reliability, Leibniz Universität Hannover, Germany

^bInstitute for Risk and Uncertainty, University of Liverpool, United Kingdom

^cInternational Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University, Shanghai, China

Published in *Proceedings of the 32nd European Safety and Reliability Conference (ESREL 2022)* in August 2022

Abstract

The interval discrete Fourier transform (DFT) algorithm can propagate in polynomial time signals carrying interval uncertainty. By computing the exact theoretical bounds on a signal with missing data, the algorithm can be used to assess the worst-case scenario in terms of maximum or minimum power, and to provide insights into the amplitude spectrum bands of the transformed signal. The uncertainty width of the spectrum bands can also be interpreted as an indicator of the quality of the reconstructed signal. This strategy must however, assume upper and lower values for the missing data present in the signal. While this may seem arbitrary, there are a number of existing techniques that can be used to obtain reliable bounds in the time domain, for example Kriging regressor or interval predictor models. Alternative heuristic strategies based on variable (as opposed to fixed) bounds can also be explored, thanks to the flexibility and efficiency of the interval DFT algorithm. This is illustrated by means of numerical examples and sensitivity analyses.

Keywords: Missing data, Exact bounds, Interval discrete Fourier transform, Power spectral density estimation, Interval uncertainty, Uncertainty quantification.

8.1 Introduction

The consideration and quantification of uncertainties in real data are of paramount importance for the design and simulation of buildings and structures. Even small measurement errors can lead to a wrong consideration of the input data and result in a disastrous interpretation of the simulation results, e.g. if an actually catastrophic result is shifted into an acceptable range by not taking uncertainties into account. Uncertainties should therefore be considered in any case and included in the simulation, also in order to determine possible safety margins. In order to safely design or to assess the reliability and robustness of buildings and structures that are subject to environmental processes such as wind, earthquakes or waves and thus exhibit dynamic behaviour, simulations are indispensable. The discrete Fourier transform (DFT) is an important tool in this field to determine the frequency components and their amplitude of the environmental processes. Consideration of uncertainties in the data, such as missing data, should be combined with the DFT to obtain reliable results.

Some approaches for estimating power spectral density (PSD) functions from signals with missing data have already been proposed. In particular, approaches treating missing data as normal

distributed random variables and propagating them through the DFT [77, 105]. Another approach was presented by [213], where the fast Fourier transform (FFT) and convolution were studied for signals with interval and fuzzy uncertainty. An algorithm to propagate interval signals through the DFT to obtain exact bounds on the Fourier amplitude was proposed by the authors of this work in [214], while the algorithm is described in details and applied to an example involving dynamic structural analysis in [221]. The algorithm enables the quantification of uncertainties in time signals and to project them into the frequency domain. No assumptions are made about the dependence and distribution of the error over the time steps. Thus, a bounded Fourier amplitude and a PSD function can be computed, which can be used to analyse system responses in the frequency domain, taking into account these uncertainties. The objective of this work is to investigate the capabilities of the interval DFT in missing data problems. It also aims to determine the severity of the missing data and the possible impact on the interval DFT algorithm. The quantity used to measure uncertainty in this work is the area between the upper and lower bounds. Contrarily, a Fourier amplitude without uncertainty will have such an area equal to zero.

This work is organised as follows: Some theoretical background that is relevant for this work is provided in Section 8.2. The problem of missing data is elaborated in Section 8.4. In Section 8.3 the interval DFT algorithm is described briefly. The capabilities of said algorithm in combination with missing data problems are explored in Section 8.5, while the final conclusions are given in Section 8.6.

8.2 Preliminaries

This section introduces some fundamental theoretical concepts that will be required in this work.

8.2.1 Power spectrum estimation

Given a signal x_n , represented as a zero mean stochastic process. To examine the signal for its frequency components, it can be transformed into the frequency domain using the periodogram. The periodogram is the squared absolute value of the Fourier transform and reads as follows

$$\hat{S}_X(\omega_k) = \frac{\Delta t^2}{T} \left| \sum_{n=0}^{N-1} x_n \cdot e^{-\frac{2\pi i k n}{N}} \right|^2, \quad (8.1)$$

where Δt is the time step size, T is the total length of the record, n describes the data point index in the record, N is the total number of data points in the signal and k is the frequency number of $\omega_k = \frac{2\pi k}{T}$.

8.2.2 Generation of artificial time signals

To generate an artificial time signal for simulation purposes, the Spectral Representation Method (SRM) can be utilised [84]. The SRM generates a time signal X_t based on an underlying PSD

function S_X and carries their characteristics. The SRM is

$$X_t = \sum_{m=0}^{M-1} \sqrt{4S_X(\omega_m)\Delta\omega} \cos(\omega_m t + \varphi_m), \quad (8.2)$$

with $\omega_m = m\Delta\omega$, $m = 0, 1, 2, \dots, M-1$, where M is the total number of frequency points, t as time vector and φ_m as uniformly distributed random phase angles in the range $[0, 2\pi]$.

As the underlying PSD function, a spectrum derived within the Joint North Sea Wave Observation Project (JONSWAP) [89] will be used throughout this work. The JONSWAP spectrum is an extension of the Pierson-Moskowitz PSD function [88] and is utilised to describe the dynamic behaviour of sea waves in the frequency domain.

8.3 The interval DFT algorithm

To transform signals into the frequency domain, the DFT can be employed. The DFT converts a signal $x = x_0, x_1, \dots, x_{N-1}$ to a Fourier sequence $z = z_0, z_1, \dots, z_{N-1}$ for $k = 0, \dots, N-1$. Since many signals are subject to missing data, these must be taken into account during the transformation in order to obtain reliable results. One possibility is to reconstruct the data before the transformation. However, since the DFT is very sensitive to changes in the signal, as shown in Section 8.4, it is more reasonable to fill the missing data gaps with intervals and propagate them through the DFT. However, since the DFT is not able to transform such uncertainties, an algorithm was proposed that is capable to propagate interval uncertainties through the DFT and thus calculate exact bounds on the Fourier amplitude. This *interval DFT algorithm* is briefly described here, for a detailed explanation and examples the reader is referred to [221] and [214]. Based on the *interval extension* of the DFT, obtained by replacing the real signal with their interval values for each frequency number k

$$\bar{z}_k = \sum_{n=0}^{N-1} \bar{x}_n e^{-ia_k} = \sum_{n=0}^{N-1} \bar{x}_n \cdot [\cos a_k - i \cdot \sin a_k], \quad (8.3)$$

with $a_k = 2\pi kn/N$, the algorithm computes two vertices for each iteration n , resulting from the interval values of the n -th data point of the signal. In each iteration step, the vertices are added to the previous vertices. The convex hull is calculated from these. The points of the convex hull are passed on to the next iteration step, while the remaining vertices have no influence on the calculation and are discarded. Once all points of the signal have been iterated, the minimum and maximum distance of the convex hull to the origin of the coordinate system is determined, which defines the interval bounds of the absolute value of the transform

$$\bar{A}_k = |\bar{z}_k| = \sqrt{\left[\sum_{n=0}^{N-1} \bar{x}_n \cos a_k \right]^2 + \left[\sum_{n=0}^{N-1} \bar{x}_n \sin a_k \right]^2}. \quad (8.4)$$

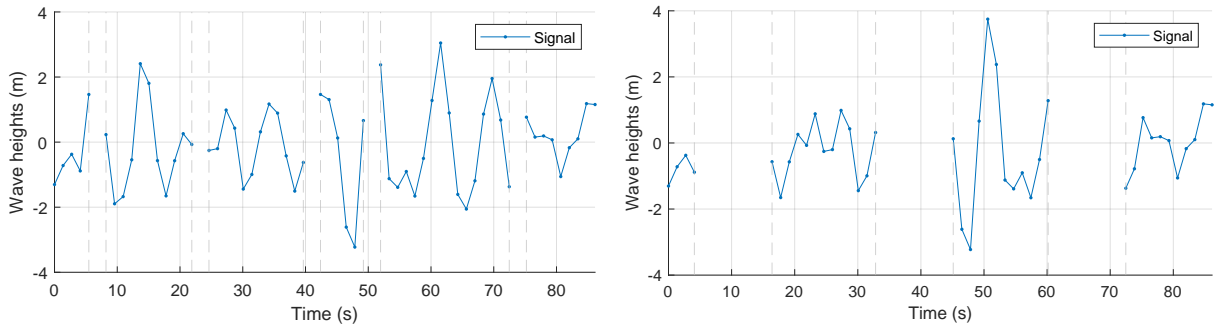


Figure 8.1: The signal under investigation with two examples each with missing data.

The absolute values of the points in the convex hull are calculated for this purpose. If the origin of the coordinate system is within the convex hull, the lower bound is 0, otherwise it is defined by the minimum absolute value. The upper bound is always determined by the maximum absolute value. Thus, an upper and lower bound of the Fourier amplitude can be computed for each frequency number k .

8.4 Missing data

A common problem when using real data records is that of missing data. The causes of missing data range from simple measurement errors to total sensor failure. It is possible that the sensor is damaged by the event it is supposed to record, e.g. an earthquake, and makes incorrect recordings or stops recording completely. In addition, the sensors may be temporarily unavailable due to maintenance. If the period of unavailability is sufficiently short, intervals can be used to bridge this gap. These causes introduce uncertainty into the data series. Although there are various reconstruction methods, e.g. least squares method, compressive sensing or autoregressive methods, the method for reconstructing the signal in case of missing data will not be considered further. Here, the focus is on the performance of the proposed algorithm rather than the reconstruction method. The reconstructed data are represented by intervals, accounting for uncertainties induced through the reconstruction. Thus, the reconstructed signal is passed to the interval DFT algorithm as an interval signal. Fig. 8.1 shows the signal under investigation in this work with two examples each with missing data. In this work, the same signal is used throughout to ensure maximum comparability of the different cases. If a signal in time domain is certainly known, it can be transformed to the frequency domain via the DFT without loss of information. In fact, the DFT is sensitive to small changes in the signal. To demonstrate the sensitivity, the signal in Fig. 8.1, consisting of 64 data points, is investigated. The target spectrum, i.e. the Fourier amplitude of the signal without missing data computed with Eq. 8.1, is depicted with the Fourier amplitudes of the same signal with 5%, 10% and 25% missing data, which are reconstructed by linear interpolation between the two adjacent non-missing data points, see Fig. 8.2. The position of the missing data is randomly chosen. The interpolated values are treated as discrete values instead of intervals first. Although linear interpolation is not

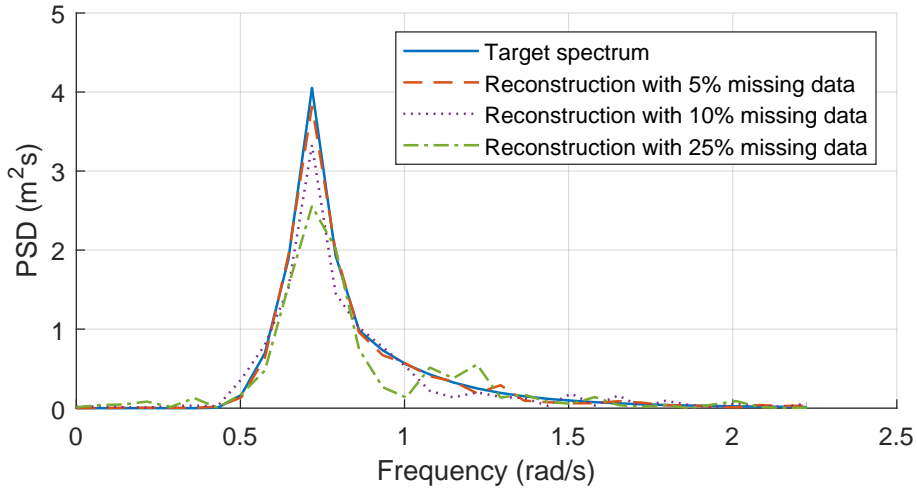


Figure 8.2: Influence of the linear interpolation on the amplitude of the DFT.

considered as a reconstruction method in this work, it can be used to illustrate the sensitivity. It can be clearly seen that the transformations have the same shape and peak frequency, but are in part very different from the target spectrum and are not as smooth. Since reconstructed data accordingly do not allow a reliable transformation into the frequency domain and do not take uncertainties into account, it is reasonable to derive bounds in which the actual spectrum may be located. The algorithm presented in 8.3 is applicable for this purpose.

Two reconstruction methods are utilised in this paper:

1. A method based on artificial inflation of the “true” data point using the sample standard deviation s of the entire signal. An interval of height $[-s, s]$ replaces the missing data.
2. A method that replaces the missing data by an interval determined by the minimum and maximum value of the entire signal.

The sample standard deviation of the signal is defined as:

$$s = \sqrt{\frac{\sum_{n=0}^{N-1} (x_n - \tilde{x})^2}{N - 1}},$$

where \tilde{x} is the sample mean of the signal.

8.5 Case studies

In this section, the influence of missing data on the bounds of the estimated PSD is investigated. Specifically, interval width, the number of missing data, the gap length, and the distribution of missing data within the signal are examined. The signal under investigation is generated by SRM (Eq. 8.2) with the underlying PSD function given in [89]. The positions of the missing data in the signal are artificially generated in random order. It is assumed that the missing data is uniformly distributed within the signal. A study is also conducted to investigate the

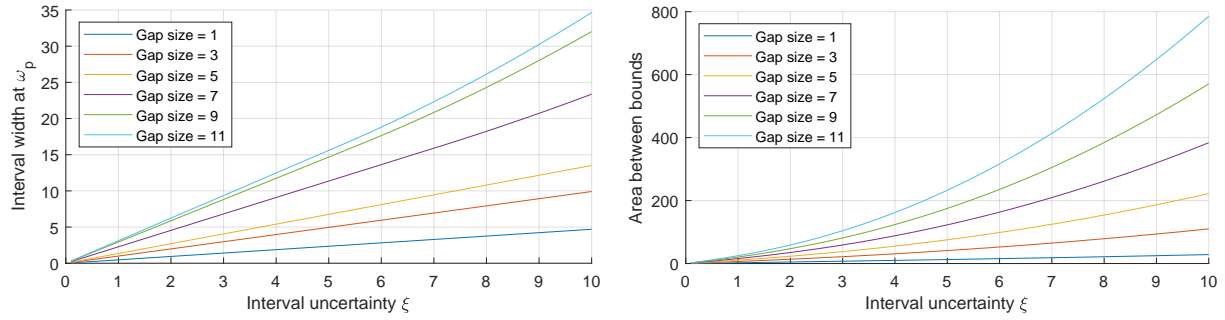


Figure 8.3: Interval width of the bounded PSD function at the peak frequency ω_p (top) and the area between upper and lower bound (bottom) for increasing interval uncertainty ξ and different lengths of the gap $l_g \in \{1, 3, 5, 7, 9, 11\}$.

influence of the position of the missing data, comparing the uniformly distributed missing data with binomially distributed missing data. In order to obtain the best possible comparison, the same signal is used in all studies of this work.

8.5.1 Sensitivity to interval uncertainty

Let ξ be the height of the interval gap, a.k.a. interval uncertainty. To investigate the sensitivity of the interval uncertainty ξ in time domain to the interval uncertainty in the frequency domain, the signal is randomly equipped with missing data gaps of length $l_g \in \{1, 3, 5, 7, 9, 11\}$, where the gap length is given as the number of missing time points.

The interval uncertainty ξ of these gaps is successively increased from 0.1 to 10. To determine the sensitivity, the interval width of the Fourier amplitude at the peak frequency ω_p , as well as the area between the upper and lower bound are determined. The results are depicted in Fig. 8.3. For smaller gaps with low interval uncertainty, a linear trend in the interval width at ω_p appears at the beginning, which later changes to a non-linear trend. This occurs as soon as the lower bound of the Fourier amplitude reaches 0 and only the upper bound contributes to the interval width. The increase is nevertheless moderate and not extremely strong. The area between the bounds, on the other hand, has a non-linear trend even with low interval uncertainty and small gaps. This non-linearity becomes stronger the larger the gap becomes. This is due to the fact that the entire frequency range is considered instead of only the peak frequency ω_p . At many frequency points, the lower bound has already reached 0, while it is still higher at the peak frequency. For larger gaps, the lower bound is mostly zero, which explains why in Fig. 3 the start of the non-linear behaviour is appreciated for lower interval uncertainty.

8.5.2 Number of missing data

In the following example, the interval uncertainty has been kept constant and corresponds to the sample standard deviation s of the signal. The number of missing data points, on the other hand, has been gradually increased to investigate the influence of the number of missing data on the bounds of the PSD. In Fig. 8.4, the reconstructed signals and the the bounds of the

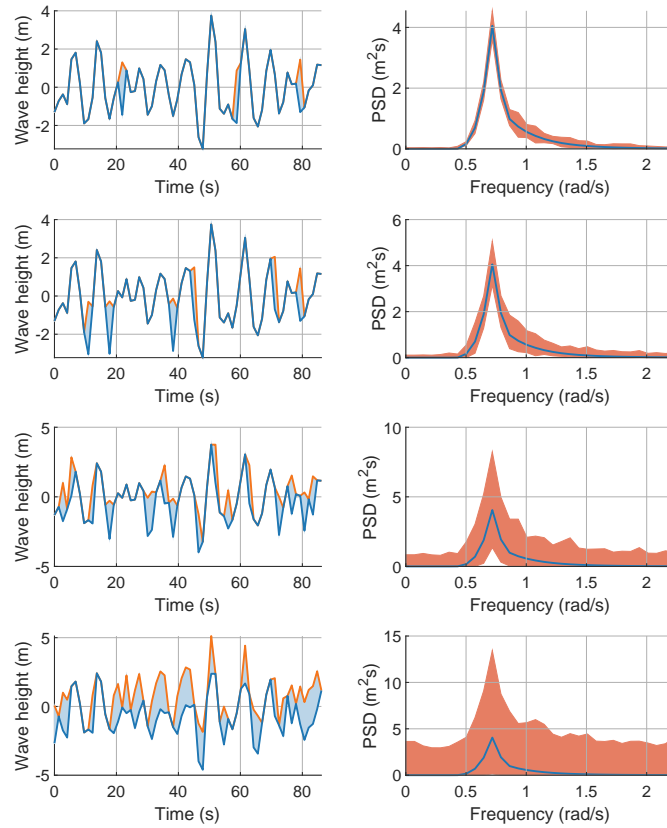


Figure 8.4: Signal with 5%, 10%, 25% and 50% missing data reconstructed using method (1) and corresponding bounded PSD functions.

estimated PSDs are shown for 5%, 10%, 25% and 50% missing data in the signal. The results show that a small amount of missing data (e.g. 5% or 10%) can be captured well with the interval DFT algorithm. The bounds enclose the estimated PSD function of the discrete signal relatively tightly and are therefore very useful for quantifying the uncertainties. Also, the bounds of the PSD for a higher amount of missing data in the signal (up to 50% in this example) can still be considered, despite the relatively wide bounds, e.g. for a worst-case consideration where only the upper bound is used.

In the following, the same example is shown, but the data was reconstructed using method (2), see Fig. 8.5 for the reconstructed signals and the bounds of the PSDs in frequency domain.

The results also show here that small amounts of missing data can be mapped well in the frequency domain even with reconstruction method (2). With higher numbers of gaps, however, the determination of the bounds in the frequency domain reaches its limitation, as the computed bounds are very high and can no longer be used for practical purposes. For example, the bounds from the previous example with 50% missing data have a lower interval uncertainty than the signal with 25% missing data in this example. This yields in particular that if there is little missing data, reconstruction can be carried out conservatively with wide intervals. Conversely, if the number of missing data is large, a method with a more accurate reconstruction is required.

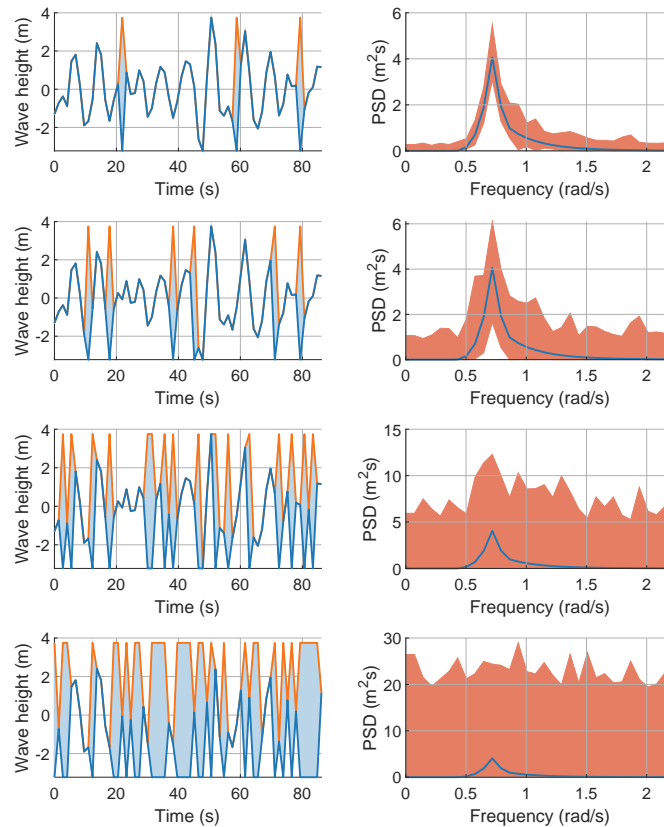


Figure 8.5: Signal with 5%, 10%, 25% and 50% missing data reconstructed using method (2) and corresponding bounded PSD functions.

As a measure for uncertainty, the area between upper and lower bound is utilised. Fig. 8.6 shows this for an increasing number of missing data recovered with the two reconstruction methods. Due to possible random fluctuations, as the position of missing data is randomly chosen, this simulation was carried out 100 times in order to average out these fluctuations. As expected, there is a significantly higher area between the bounds when using reconstruction method (2) compared to reconstruction method (1).

8.5.3 Gap size of missing data

Recall that gap size is given as the number of missing time points, and it is also referred to as gap length. To determine the influence of the gap length, different scenarios were evaluated. The gap lengths $l_g \in \{1, 20, 40, 60\}$ were artificially inserted into the signal. The gaps were first reconstructed with method (1). The signals and the corresponding transformations are shown in Fig. 8.7. It can be seen that small gaps filled with the intervals provide a good transformation and the bounds are relatively tight around the exact spectrum. The interval DFT algorithm can also handle larger gaps well, although the bounds of the transformation are comparatively large. Nevertheless, these can be used, for example, to design for a worst-case when only the upper bound with the largest energy content is used for planning and simulation.

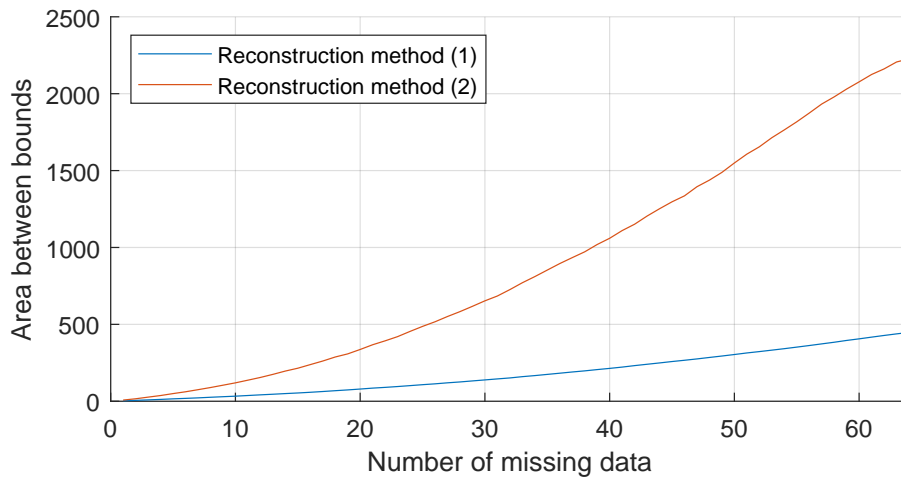


Figure 8.6: Area between upper and lower bound investigated for the number of missing data.

For these investigations, all examples with reconstruction method (2) are omitted, since it has been shown that large gaps already lead to extremely large bounds with reconstruction method (1) and are practically no longer useful. Since the length of the gap naturally corresponds to the number of missing data, no significant differences between Fig. 8.8 and Fig. 8.6 in the previous section can be detected. This indicates that the position of the missing data has a minor role in determining the uncertainty, but the number has a major role. Nevertheless, for completeness the influence of the distribution of the missing data will be investigated in the next section.

8.5.4 Distribution of missing data

This example is used to investigate the influence of the distribution of the missing data within the signal. For the sake of brevity, only the results for reconstruction method (1) are shown. In addition, it has been shown in the previous sections reconstruction method (2) cannot be used for real phenomena if the number of missing data is sufficiently high, as the bounds are extremely large. For the investigations a uniform distribution and a binomial distribution were utilised to randomly generate the missing data and to investigate their influence on the transformation to the frequency domain. The interval PSDs of the reconstructed signal with 4, 8, 16 and 32 missing data are depicted in Fig. 8.9. It can be seen that the influence of the position of the missing data is of minor relevance. Although the transformed signals shown are only specific cases, they are nevertheless representative of the general case. This statement can be supported by the fact that this simulation has been carried out several times, but the results are always identical. The transforms look almost identical in each case, regardless of the distribution of the missing data. In addition, the interval widths at the respective frequencies, such as the peak frequency, are very similar and the area between the bounds are also almost identical.

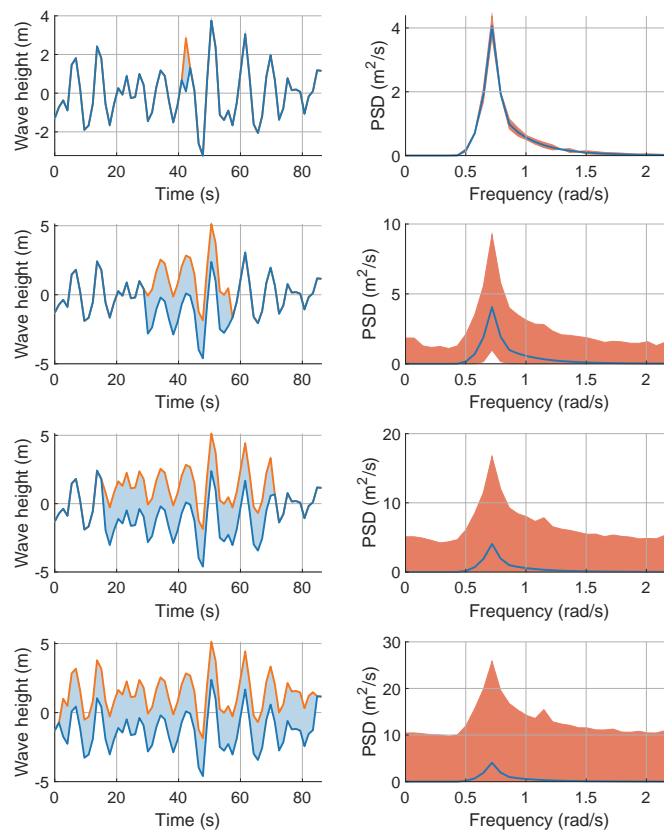


Figure 8.7: Signals with a gap of length $l_g \in \{1, 20, 40, 60\}$ of missing data reconstructed by method (1) and corresponding bounded PSD functions.

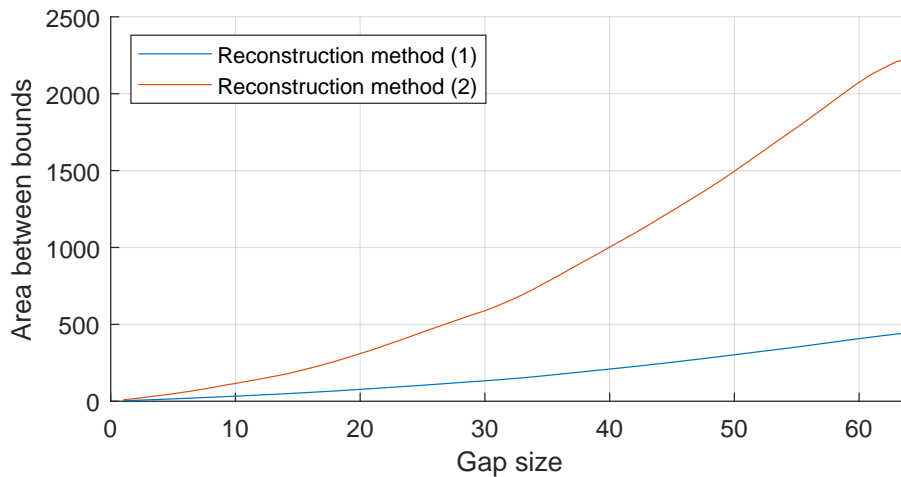


Figure 8.8: Area between upper and lower bound investigated for the length of the gap.

8.6 Conclusions

In this work, the interval DFT algorithm has been investigated for its ability to transform signals with missing data reconstructed by intervals. Different scenarios have been considered, such as the influence of the interval width, the number of missing data, the length of the gap of missing data and the distribution of the missing data in the signal. It was shown, that the largest influence was exerted by the interval uncertainty in the signal and the number of missing data, while the distribution of the data and their position is of minor importance. In addition, no indications could be found of an influence whether the data are missing at individual points or appear as a large gap. It was found that too large intervals often lead to extremely wide bounds, which are usually no longer usable for practical purposes. If the number of missing data is sufficiently small, however, a good transformation can be computed even with a conservative estimation of the intervals, in which the bounds are close to the actual spectrum. With a larger number of missing data or larger gaps, it is also possible to plan for the worst-case by considering only the upper bound, provided that the interval width is reasonably chosen. It has also been shown that the potential energy content of the PSD can change significantly depending on the choice of intervals. In summary, the interval DFT algorithm provides excellent results for uncertain data. However, it should be noted that the results are highly dependent on the reconstruction of the data. Thus, it is highly recommended that in the case of missing data, the interval DFT algorithm should be employed with an advanced reconstruction method in order to obtain reliable results.

Replicability

The software for computing the interval DFT can be accessed in a single instance via GitHub at: <https://github.com/interval-fourier-transform/application-to-missing-data>.

The code, examples and numerical results presented in this paper are therefore fully replicable.

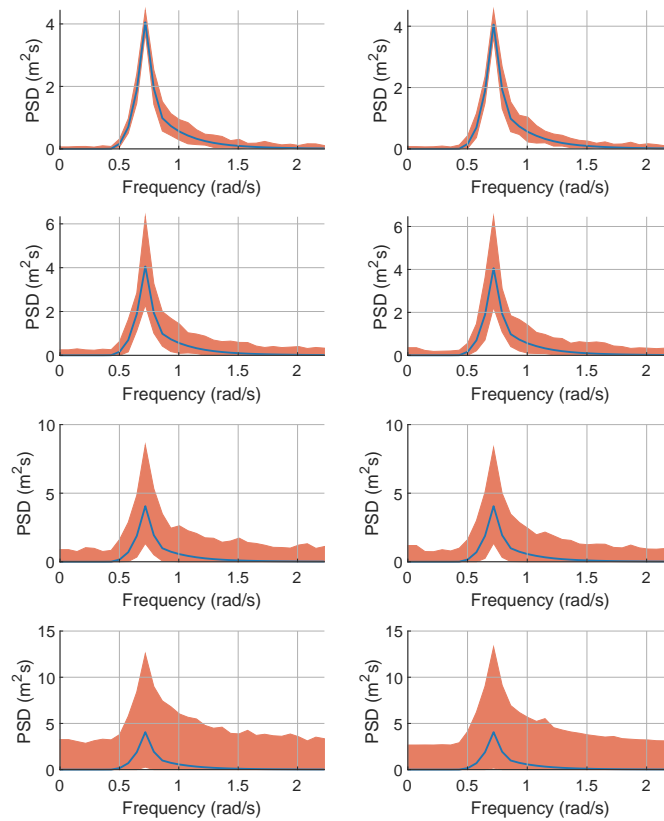


Figure 8.9: Influence of the distribution of missing data within the signal for 4, 8, 16 and 32 missing data reconstructed with method (1).

Acknowledgement

This research is funded by the Engineering & Physical Sciences Research Council (EPSRC) with grant no. EP/R006768/1. The EPSRC is greatly acknowledged for their funding and support.

9 | Conclusions and Outlook

In this concluding chapter, the main developments of the work are summarised and useful extensions and possible improvements are suggested.

9.1 Conclusions

The quantification of uncertainties is a very important issue in reliability analysis simulations and in determining the dynamic system behaviour of buildings and structures. If uncertainties are not adequately taken into account, this can have fatal consequences, so that a structure is considered safe even though it is at high risk of damage or collapse. Especially when using real data, uncertainties are present and need to be quantified and, if possible, reduced. This work therefore focuses on the quantification of uncertainties in the frequency domain.

The novel methods presented in this thesis contribute to the quantification of uncertainties that occur in the data records and are induced by the PSD estimation process. The methods are based on uncertainty quantification in the frequency domain, resulting in novel load models that, depending on the amount and appearance of the data, result in a probabilistic or an interval-based model. These models have proven useful in quantifying uncertainties and thus providing reliable simulation results. Since the system responses have the same appearance as the input models, i.e. probabilistic or interval-based, reliable ranges instead of deterministic values can be determined in which the actual system behaviour is located, taking into account the corresponding uncertainties. This has the advantage that the system response can thus be determined more reliably. A conclusion without taking uncertainties into account may be difficult and misleading.

The first approach presented is based on the assumption that multiple data records are available that have similar characteristics in the frequency domain. From these similar data, a probabilistic model, the relaxed PSD, is derived that is characterised by a PDF at each individual frequency. This has the advantage that reliable ranges in which the true PSD lies can be determined by subjective probabilities. It is robust against spectral outliers, which can lead to non-representative results. The probabilistic nature of this model makes it possible to directly sample individual PSDs, which can be used to generate artificial stochastic processes. In another part of this thesis, the possibilities of generating stochastic processes are examined, as these are often influenced by random variables. Overall, it can be stated that the relaxed PSD is able to generate stochastic processes with suitable methods. It is also possible to reduce the number of random variables if suitable methods such as the SHF are used.

In the case where only few but similar data are available, it is not possible to derive a reliable probabilistic model. The derived PDFs and associated probabilities can then be highly misleading. In such cases, it may be useful to derive an interval-based model instead of using probabilities. The advantage of this approach is that bounds are derived which constrain the data, and no information about the distribution of the data within these bounds needs to be

known. First, a model was presented that optimises reliable bounds via the computed basis functions of a radial basis function network. This approach takes into account the physics of the underlying process and reflects the dependencies between frequencies. A second proposed model calculates confidence intervals of the data at different confidence levels, which are transformed into Fuzzy numbers for each frequency. Both representations result in non-probabilistic models, thus they can be used to calculate reliable ranges of system response and to robustly determine critical system behaviour.

Since the models presented are based on the assumption that the data used are similar in frequency domain, it is necessary to first verify the similarity of these data before deriving the respective load models. In all cases, arbitrarily used data can lead to load models that do not adequately represent the actual underlying process. In a probabilistic model, the PDFs can have a very high variance, and in an interval-based model the bounds can be very wide, making these models less accurate. This may lead to less meaningful system responses in subsequent system analyses, as the aim is to narrow the actual system response as much as possible, rather than measuring a wide range. A prior determination of the data for spectral similarity is therefore absolutely necessary. An approach addressing this issue is presented in this thesis. Based on distances, measured with the Bhattacharyya distance, and the k -means algorithm, a data set with high spectral variance can be effectively classified into groups of similar data. An algorithm for determining the optimal number of groups based on the Silhouette method was presented in addition. This approach can not only be used to classify data, but also enables the reduction of variance in the data set and in the simulation results. Individual load models can be generated from the respective groups, which can be applied to the structure under investigation in order to obtain more accurate simulation results. For an approximation of the failure probability, the empirical CDF can be derived. A weighted CDF can be calculated from the individual load models taking into account the probability of occurrence.

The final contribution of this thesis is an extension of the DFT. Signals are typically subject to uncertainties, so it can be useful to assume reliable bounds for such a signal. Instead of a discrete-valued signal, an interval-valued signal can subsequently be obtained, which is characterised by an upper and lower bound for each point in time. In this case an exact statement about the position of the true signal cannot be made. The presented interval DFT algorithm characterises the geometry of the united set and thus provides exact bounds instead of outer bounds. It also addresses the repeated variables problem, a typical problem in interval arithmetic that causes an artificial inflation of the bounds. The presented algorithm thus allows the propagation of an interval signal through the DFT whereby exact bounds of the Fourier amplitude can be obtained. The algorithm enables the propagation of signals with missing data, provided that the data have been reconstructed appropriately beforehand. Missing data signals can be a major problem, as no information about the actual position of the data point is known. Therefore, a reliable reconstruction and subsequent derivation of bounds in the frequency domain is essential and contributes to the efficient quantification of uncertainties.

Overall, the methods presented in this thesis provide a significant contribution to the quantification of uncertainties in PSD estimation. A wide range of different cases are covered, namely multiple data resulting in a probabilistic model; limited data, allowing interval and fuzzy-based models to be generated; single uncertain time signals, which led to the interval DFT algorithm and the computation of exact bounds on the Fourier amplitude; and the classification of data in the frequency domain to define the spectral similarity and reduce the variance in the results and input data. In addition, the missing data problem was addressed by reconstructing the data with reliable intervals and calculating reliable Fourier amplitude bounds using the interval DFT algorithm. The developments in this work thus provide a framework for quantifying uncertainties in the PSD estimation and in determining the reliability of buildings and structures. This has opened up new avenues and possibilities in the treatment of uncertainties in the frequency domain. Since uncertainties are present in time signals and further uncertainties are induced by the PSD estimation process, a quantification of these uncertainties directly in the frequency domain seems to be reasonable. Thus, this framework offers a significant contribution to uncertainty quantification and closes a gap in this research area.

9.2 Outlook

The methods presented in this paper have proven to be capable of handling uncertainties in environmental processes and developing corresponding load models in the frequency domain. Nevertheless, further useful extensions and improvements have already been identified during the development, which will be further discussed here.

The main focus of the thesis is on modelling uncertainties. However, the propagation of uncertainties in the simulation should not be neglected. While relatively simple propagation methods, such as Monte Carlo, were used in this work, advanced approaches should be used in future works so that, for example, the failure probability can be determined efficiently in the context of reliability analyses. Advanced sampling methods such as Subset Sampling or Directional Importance Sampling can be used to generate adequate stochastic processes efficiently, the Probability Density Evolution Method (PDEM) can be applied to obtain a probabilistic system response which develops over time, and Change of Measure (COM) may be implemented to efficiently propagate PSD functions and stochastic processes. These methods lead to a robust and efficient propagation of uncertainties through a system and provide reliable system responses. Within this scope, any attempt should be made to reduce the number of random variables in the derived load models and generated stochastic processes. Random variables introduce further uncertainty in the sense of stochastic variation into the processes, such that a relatively high number of generated processes is required to average out the influence of the random variables. Therefore, efficient methods to reduce the random variables are necessary. Initial studies on this have already been carried out in Chapter 3.

When developing load models, it is important to consider correlations and dependencies between frequencies. For the imprecise PSD in Chapter 4, this has already been realised through the

use of basis functions and an approximation of the basis spectrum. For the relaxed PSD in Chapter 2 the dependencies between frequencies are not considered. However, this can be achieved by including a full covariance matrix of the spectrum process. Correlations identified in the ensemble spectrum can have a significant impact on the model description of the process. For example, the individual PSD functions for each data set in an ensemble might be comparably different, but their integral sums might remain constant. In this case, not taking spectrum covariance into account may result in a model that contains uncharacteristic extreme values of high and low power.

Since this work aims at modelling environmental processes considering uncertainties, it is important to prove the feasibility with real data sets and real-world problems. The load models developed in Chapter 4 and 6 demonstrate that they can be employed for this purpose. Nevertheless, an extension of the other developed models is necessary. Nowadays, thanks to numerous databases, it is possible to obtain real data easily. It is assumed that the usability of real data is not a major challenge, provided that adequate pre-processing of the data has taken place. Since real data are usually subject to uncertainties, they are often affected by strong variations in the frequency domain if no suitable PSD estimators are used. However, if estimators for PSDs are used in an averaged sense, for instance, the difficulties in deriving the proposed load models are diminished. This has already been demonstrated in Chapter 4, where the use of weak estimators and inadequately pre-processed data led to unacceptable results. This was counteracted by the utilisation of appropriate estimation techniques.

Another important step is the extension of the load models to the non-stationary case, as the estimation of evolutionary PSD functions also takes into account temporal changes in frequency. Thus, the process behaviour will be represented more realistically. In particular, earthquakes as short-term processes have a relevant transient behaviour instead of a stationary response to dynamic accelerations characterised by the response to a PSD function. Since only the stationary case is considered in this work, it is necessary to extend the developed load models to the non-stationary case, using the short-time Fourier transform (STFT) or the harmonic wavelet transform to estimate the frequency in the time-frequency domain. This is expected to be straightforward in all cases, except for minor adjustments, as the methods developed are independent of the dimension of the given problem. However, two problems can already be identified in the estimation of evolutionary PSDs that need to be addressed. The first challenge is to determine the relative position in time of each process record as part of an ensemble. For processes that have a strict timestamp for comparison, this is not a problem. However, when records from independent environmental processes are used as part of an ensemble, establishing a strict beginning, end and duration for each record is not a simple task. To account for this, an optimisation problem must be solved over the temporal shift of the spectrum that minimises the total spectral variance over the ensemble. Such a problem could be solved iteratively by defining an initial mean spectrum, then optimising the shifts of each data set to reduce the variance of the ensemble, and eventually recalculating the mean spectrum. Such a framework

would prove to be a useful tool in a number of areas outside of civil engineering where non-stationary ensembles of data are analysed. The second problem to be solved is the occurrence of "end-effects" in the time-frequency domain, or specifically in the context of wavelets referred to as "cone-of-influence", which are artefacts in wavelet or STFT representations of the process estimated from data in time domain. The problem arises from the fact that numerical wavelet transforms and the discrete Fourier transform (for STFT) involve cyclic convolutions, which causes the spectral power to spill over from one end of the signal to the other. This problem is usually mitigated in STFT by applying a smooth window when truncating the data set, but the orthogonality of the transformation is no longer preserved. With wavelets, the problem can be solved by artificially expanding the signal, e.g. by padding with zeros or mirrored copies of the signal.

By incorporating the above-mentioned considerations, the load models presented in this thesis can be further improved and a more realistic description can be derived. The resulting simulations, for example in the context of reliability analyses, thus become more accurate and robust.

10 | List of Publications

Journal articles:

- Behrendt, M.; Bittner, M.; Comerford, L.; Beer, M.; Chen, J.B. (2022): Relaxed power spectrum estimation from multiple data records utilising subjective probabilities, *Mechanical Systems and Signal Processing*, 165, Article 108346.
DOI: 10.1016/j.ymsp.2021.108346
- Behrendt, M.; de Angelis, M.; Comerford, L.A.; Zhang, Y.J.; Beer, M. (2022): Projecting interval uncertainty through the discrete Fourier transform: an application to time signals with poor precision, *Mechanical Systems and Signal Processing*, Volume 172, 1 June 2022, 108920.
DOI: 10.1016/j.ymsp.2022.108920
- Behrendt, M.; Kitahara, M.; Kitahara, T.; Comerford, L.; Beer, M. (2022): Data-driven reliability assessment of dynamic structures based on power spectrum classification, *Engineering Structures*, Volume 268, 1 October 2022, 114648.
DOI: 10.1016/j.engstruct.2022.114648
- Behrendt, M.; Faes, M. GR; Valdebenito, M. A.; Beer, M.: Estimation of an imprecise power spectral density function with optimised bounds from scarce data for epistemic uncertainty quantification, *submitted to Mechanical Systems and Signal Processing* on 9 September 2022.

Conference paper:

- Behrendt, M.; Bittner, M.; Beer, M. (2022): Stochastic process generation from relaxed power spectra utilising stochastic harmonic functions, *Proceedings of the 8th International Symposium on Reliability Engineering and Risk Management (ISRERM 2022)*, Hannover, Germany.
DOI: 10.3850/978-981-18-5184-1_MS-01-220-cd
- Behrendt, M.; Kitahara, M.; Kitahara, T.; Comerford, L.; Beer, M. (2022): Classification of power spectra from data sets with high spectral variance for reliability analysis of dynamic structures, *Proceedings of the 8th International Symposium on Reliability Engineering and Risk Management (ISRERM 2022)*, Hannover, Germany.
DOI: 10.3850/978-981-18-5184-1_MS-11-160-cd
- Behrendt, M.; de Angelis, M.; Comerford, L.; Beer, M. (2022): Assessing the severity of the missing data problem with the interval DFT algorithm, *Proceedings of the 32nd European Safety and Reliability Conference (ESREL 2022)*, Dublin, Ireland.
DOI: 10.3850/978-981-18-5183-4_S14-05-243-cd

- Behrendt, M.; Faes, M. GR; Valdebenito, M. A.; Beer, M. (2022): Capturing Epistemic Uncertainties in the Power Spectral Density for Limited Data Sets, *Probabilistic Safety Assessment and Management (PSAM 16)*, Honolulu, USA.
- Behrendt, M.; Bittner, M.; Comerford, L.; Broggi, M.; Beer, M. (2020): Parameter Investigation of Relaxed Uncertain Power Spectra for Stochastic Dynamic Systems, *Proceedings of the XI International Conference on Structural Dynamics (EURODYN 2020)*, Athens, Greece.
DOI: 10.47964/1120.9311.18861
- Behrendt, M., Comerford, L., Beer, M. (2019): Development of a Relaxed Stationary Power Spectrum using Imprecise Probabilities with Application to High-rise Buildings., *IEEE Symposium Series on Computational Intelligence (SSCI)*.
DOI: 10.1109/SSCI44817.2019.9002899
- Behrendt, M., Comerford, L., Beer, M., (2019): Relaxed Stationary Power Spectrum Model Using Imprecise Probabilities, *COMPADYN Proceedings*, 1, pp. 592-599.
DOI: 10.7712/120119.6941.19045
- Behrendt, M.; Comerford, L.; Beer, M. (2019): Stochastic Processes Identification from Data Ensembles via Power Spectrum Classification., *13th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP13)*.
DOI: 10.22725/ICASP13.407

Bibliography

- [1] Armen Der Kiureghian and Ove Ditlevsen. “Aleatory or epistemic? Does it matter?” In: *Structural Safety* 31.2 (2009). Risk Acceptance and Risk Communication, pages 105–112. ISSN: 0167-4730. DOI: <https://doi.org/10.1016/j.strusafe.2008.06.020>. URL: <https://www.sciencedirect.com/science/article/pii/S0167473008000556>.
- [2] Matthias G.R. Faes, Marcos A. Valdebenito, David Moens, and Michael Beer. “Operator norm theory as an efficient tool to propagate hybrid uncertainties and calculate imprecise probabilities”. In: *Mechanical Systems and Signal Processing* 152 (2021), page 107482. ISSN: 0888-3270. DOI: <https://doi.org/10.1016/j.ymsp.2020.107482>. URL: <https://www.sciencedirect.com/science/article/pii/S0888327020308682>.
- [3] Chao Dang, Pengfei Wei, Matthias G.R. Faes, and Michael Beer. “Bayesian probabilistic propagation of hybrid uncertainties: Estimation of response expectation function, its variable importance and bounds”. In: *Computers & Structures* 270 (2022), page 106860. ISSN: 0045-7949. DOI: <https://doi.org/10.1016/j.compstruc.2022.106860>. URL: <https://www.sciencedirect.com/science/article/pii/S0045794922001201>.
- [4] Efstratios Nikolaidis, Dan M Ghiocel, and Suren Singhal. *Engineering Design Reliability Handbook*. 1st edition. CRC press, 2004. DOI: <https://doi.org/10.1201/9780203483930>.
- [5] G.I. Schuëller. “On the treatment of uncertainties in structural mechanics and analysis”. In: *Computers & Structures* 85.5 (2007). Computational Stochastic Mechanics, pages 235–243. ISSN: 0045-7949. DOI: <https://doi.org/10.1016/j.compstruc.2006.10.009>. URL: <https://www.sciencedirect.com/science/article/pii/S0045794906003348>.
- [6] Colleen Murphy, Paolo Gardoni, and Charles E Harris. “Classification and Moral Evaluation of Uncertainties in Engineering Modeling”. In: *Science and Engineering Ethics* 17.3 (2011), pages 553–570. DOI: <https://doi.org/10.1007/s11948-010-9242-2>.
- [7] Sifeng Bi, Matteo Broggi, Pengfei Wei, and Michael Beer. “The Bhattacharyya distance: Enriching the P-box in stochastic sensitivity analysis”. In: *Mechanical Systems and Signal Processing* 129 (2019), pages 265–281. ISSN: 0888-3270. DOI: <https://doi.org/10.1016/j.ymsp.2019.04.035>. URL: <https://www.sciencedirect.com/science/article/pii/S0888327019302717>.
- [8] Jianbing Chen and Zhiqiang Wan. “A compatible probabilistic framework for quantification of simultaneous aleatory and epistemic uncertainty of basic parameters of structures by synthesizing the change of measure and change of random variables”. In: *Structural Safety* 78 (2019), pages 76–87. ISSN: 0167-4730. DOI: <https://doi.org/10.1016/j.strusafe.2019.01.001>. URL: <http://www.sciencedirect.com/science/article/pii/S0167473018303059>.

- [9] Mircea Grigoriu. *Stochastic Calculus: Applications in Science and Engineering*. Springer, 2002. DOI: <https://doi.org/10.1007/978-0-8176-8228-6>.
- [10] Mircea Grigoriu. *Stochastic systems: Uncertainty Quantification and Propagation*. Springer Science & Business Media, 2012. DOI: <https://doi.org/10.1007/978-1-4471-2327-9>.
- [11] Maurice Bertram Priestley. *Spectral analysis and time series: probability and mathematical statistics*. 04; QA280, P7. 1981.
- [12] David Edward Newland. *An introduction to random vibrations, spectral & wavelet analysis*. Courier Corporation, 2012.
- [13] Matthias G.R. Faes and David Moens. “Recent Trends in the Modeling and Quantification of Non-probabilistic Uncertainty”. In: *Archives of Computational Methods in Engineering* 27 (2020), pages 633–671. DOI: <https://doi.org/10.1007/s11831-019-09327-x>. URL: <https://link.springer.com/article/10.1007/s11831-019-09327-x>.
- [14] J.C. Helton, J.D. Johnson, and W.L. Oberkampf. “An exploration of alternative approaches to the representation of uncertainty in model predictions”. In: *Reliability Engineering & System Safety* 85.1 (2004). Alternative Representations of Epistemic Uncertainty, pages 39–71. ISSN: 0951-8320. DOI: <https://doi.org/10.1016/j.res.2004.03.025>. URL: <https://www.sciencedirect.com/science/article/pii/S0951832004000511>.
- [15] Jon C. Helton, Jay D. Johnson, William L. Oberkampf, and Cédric J. Sallaberry. “Representation of analysis results involving aleatory and epistemic uncertainty”. In: *International Journal of General Systems* 39.6 (2010), pages 605–646. DOI: 10.1080/03081079.2010.486664. eprint: <https://doi.org/10.1080/03081079.2010.486664>. URL: <https://doi.org/10.1080/03081079.2010.486664>.
- [16] Bernd Möller and Michael Beer. “Engineering computation under uncertainty - Capabilities of non-traditional models”. In: *Computers & Structures* 86.10 (2008). Uncertainty in Structural Analysis - Their Effect on Robustness, Sensitivity and Design, pages 1024–1041. ISSN: 0045-7949. DOI: <https://doi.org/10.1016/j.compstruc.2007.05.041>. URL: <https://www.sciencedirect.com/science/article/pii/S0045794907002210>.
- [17] Michael Beer, Scott Ferson, and Vladik Kreinovich. “Imprecise probabilities in engineering analyses”. In: *Mechanical Systems and Signal Processing* 37.1 (2013), pages 4–29. ISSN: 0888-3270. DOI: <https://doi.org/10.1016/j.ymsp.2013.01.024>. URL: <http://www.sciencedirect.com/science/article/pii/S0888327013000812>.
- [18] A. Gray, A. Wimbush, M. de Angelis, P.O. Hristov, D. Calleja, E. Miralles-Dolz, and R. Rocchetta. “From inference to design: A comprehensive framework for uncertainty quantification in engineering with limited information”. In: *Mechanical Systems and Signal Processing* 165 (2022), page 108210. ISSN: 0888-3270. DOI: <https://doi.org/10.1016/j>

- .ymssp.2021.108210. URL: <https://www.sciencedirect.com/science/article/pii/S0888327021005859>.
- [19] Kari Sentz and Scott Ferson. “Combination of Evidence in Dempster-Shafer Theory”. In: (Apr. 2002). DOI: 10.2172/800792. URL: <https://www.osti.gov/biblio/800792>.
- [20] Kurt Weichselberger. “The theory of interval-probability as a unifying concept for uncertainty”. In: *International Journal of Approximate Reasoning* 24.2 (2000), pages 149–170. ISSN: 0888-613X. DOI: [https://doi.org/10.1016/S0888-613X\(00\)00032-3](https://doi.org/10.1016/S0888-613X(00)00032-3). URL: <https://www.sciencedirect.com/science/article/pii/S0888613X00000323>.
- [21] Chao Dang, Pengfei Wei, Matthias G.R. Faes, Marcos A. Valdebenito, and Michael Beer. “Parallel adaptive Bayesian quadrature for rare event estimation”. In: *Reliability Engineering & System Safety* 225 (2022), page 108621. ISSN: 0951-8320. DOI: <https://doi.org/10.1016/j.res.2022.108621>. URL: <https://www.sciencedirect.com/science/article/pii/S0951832022002617>.
- [22] Enrico 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>.
- [23] Siu-Kui Au and James L. Beck. “Estimation of small failure probabilities in high dimensions by subset simulation”. In: *Probabilistic Engineering Mechanics* 16.4 (2001), pages 263–277. ISSN: 0266-8920. DOI: [https://doi.org/10.1016/S0266-8920\(01\)00019-4](https://doi.org/10.1016/S0266-8920(01)00019-4). URL: <https://www.sciencedirect.com/science/article/pii/S0266892001000194>.
- [24] Mauricio A. Misraji, Marcos A. Valdebenito, Héctor A. Jensen, and C. Franco Mayorga. “Application of directional importance sampling for estimation of first excursion probabilities of linear structural systems subject to stochastic Gaussian loading”. In: *Mechanical Systems and Signal Processing* 139 (2020), page 106621. ISSN: 0888-3270. DOI: <https://doi.org/10.1016/j.ymssp.2020.106621>. URL: <https://www.sciencedirect.com/science/article/pii/S0888327020300078>.
- [25] Abraham M. Hasofer and Niels C. Lind. “Exact and Invariant Second-Moment Code Format”. In: *Journal of the Engineering Mechanics Division* 100.1 (1974), pages 111–121. DOI: 10.1061/JMCEA3.0001848. eprint: <https://ascelibrary.org/doi/pdf/10.1061/JMCEA3.0001848>. URL: <https://ascelibrary.org/doi/abs/10.1061/JMCEA3.0001848>.
- [26] Rüdiger Rackwitz and Bernd Flessler. “Structural reliability under combined random load sequences”. In: *Computers & Structures* 9.5 (1978), pages 489–494. ISSN: 0045-7949. DOI: [https://doi.org/10.1016/0045-7949\(78\)90046-9](https://doi.org/10.1016/0045-7949(78)90046-9). URL: <https://www.sciencedirect.com/science/article/pii/0045794978900469>.

- [27] Behrooz Keshtegar and Zeng Meng. “A hybrid relaxed first-order reliability method for efficient structural reliability analysis”. In: *Structural Safety* 66 (2017), pages 84–93. ISSN: 0167-4730. DOI: <https://doi.org/10.1016/j.strusafe.2017.02.005>. URL: <https://www.sciencedirect.com/science/article/pii/S0167473017300607>.
- [28] Armen Der Kiureghian and Mario De Stefano. “Efficient Algorithm for Second-Order Reliability Analysis”. In: *Journal of Engineering Mechanics* 117.12 (1991), pages 2904–2923. DOI: 10.1061/(ASCE)0733-9399(1991)117:12(2904). eprint: <https://ascelibrary.org/doi/pdf/10.1061/%28ASCE%290733-9399%281991%29117%3A12%282904%29>. URL: <https://ascelibrary.org/doi/abs/10.1061/%28ASCE%290733-9399%281991%29117%3A12%282904%29>.
- [29] Xianzhen Huang, Yuxiong Li, Yimin Zhang, and Xufang Zhang. “A new direct second-order reliability analysis method”. In: *Applied Mathematical Modelling* 55 (2018), pages 68–80. ISSN: 0307-904X. DOI: <https://doi.org/10.1016/j.apm.2017.10.026>. URL: <https://www.sciencedirect.com/science/article/pii/S0307904X17306431>.
- [30] Mårten Olsson Zhangli Hu Rami Mansour and Xiaoping Du. “Second-order reliability methods: a review and comparative study”. In: *Structural and Multidisciplinary Optimization* 64 (2021), pages 3233–3263. DOI: 10.1007/s00158-021-03013-y. URL: <https://doi.org/10.1007/s00158-021-03013-y>.
- [31] Yan-Gang Zhao and Zhao-Hui Lu. *Structural reliability: approaches from perspectives of statistical moments*. John Wiley & Sons, 2021.
- [32] Jianbing Chen, Jiashu Yang, and Hector Jensen. “Structural optimization considering dynamic reliability constraints via probability density evolution method and change of probability measure”. In: *Structural and Multidisciplinary Optimization* 62.5 (2020), pages 2499–2516.
- [33] Jiashu Yang, Hector Jensen, and Jianbing Chen. “Structural optimization under dynamic reliability constraints utilizing probability density evolution method and metamodels in augmented input space”. In: *Structural and Multidisciplinary Optimization* 65.4 (2022), pages 1–26.
- [34] Jorge E Hurtado and Diego A Alvarez. “Neural-network-based reliability analysis: a comparative study”. In: *Computer Methods in Applied Mechanics and Engineering* 191.1 (2001). Micromechanics of Brittle Materials and Stochastic Analysis of Mechanical Systems, pages 113–132. ISSN: 0045-7825. DOI: [https://doi.org/10.1016/S0045-7825\(01\)00248-1](https://doi.org/10.1016/S0045-7825(01)00248-1). URL: <https://www.sciencedirect.com/science/article/pii/S0045782501002481>.
- [35] A. Hosni Elhewy, E. Mesbahi, and Y. Pu. “Reliability analysis of structures using neural network method”. In: *Probabilistic Engineering Mechanics* 21.1 (2006), pages 44–53. ISSN: 0266-8920. DOI: <https://doi.org/10.1016/j.pro bengmech.2005.07.002>. URL: <https://www.sciencedirect.com/science/article/pii/S026689200500041X>.

- [36] Claudio M. Rocco and José Alí Moreno. “Fast Monte Carlo reliability evaluation using support vector machine”. In: *Reliability Engineering & System Safety* 76.3 (2002), pages 237–243. ISSN: 0951-8320. DOI: [https://doi.org/10.1016/S0951-8320\(02\)00015-7](https://doi.org/10.1016/S0951-8320(02)00015-7). URL: <https://www.sciencedirect.com/science/article/pii/S0951832002000157>.
- [37] Jorge E. Hurtado. “An examination of methods for approximating implicit limit state functions from the viewpoint of statistical learning theory”. In: *Structural Safety* 26.3 (2004), pages 271–293. ISSN: 0167-4730. DOI: <https://doi.org/10.1016/j.strusafe.2003.05.002>. URL: <https://www.sciencedirect.com/science/article/pii/S0167473003000420>.
- [38] Bruno Sudret. “Global sensitivity analysis using polynomial chaos expansions”. In: *Reliability Engineering & System Safety* 93.7 (2008). Bayesian Networks in Dependability, pages 964–979. ISSN: 0951-8320. DOI: <https://doi.org/10.1016/j.res.2007.04.002>. URL: <https://www.sciencedirect.com/science/article/pii/S0951832007001329>.
- [39] Thierry Crestaux, Olivier Le Maître, and Jean-Marc Martinez. “Polynomial chaos expansion for sensitivity analysis”. In: *Reliability Engineering & System Safety* 94.7 (2009). Special Issue on Sensitivity Analysis, pages 1161–1172. ISSN: 0951-8320. DOI: <https://doi.org/10.1016/j.res.2008.10.008>. URL: <https://www.sciencedirect.com/science/article/pii/S0951832008002561>.
- [40] Mohammad Hadigol and Alireza Doostan. “Least squares polynomial chaos expansion: A review of sampling strategies”. In: *Computer Methods in Applied Mechanics and Engineering* 332 (2018), pages 382–407. ISSN: 0045-7825. DOI: <https://doi.org/10.1016/j.cma.2017.12.019>. URL: <https://www.sciencedirect.com/science/article/pii/S0045782517307697>.
- [41] G. Alefeld and J. Herzberger. *Introduction to Interval Computation*. Computer Science and Applied Mathematics. Elsevier Science, 2012. ISBN: 9780080916361.
- [42] Ramon E Moore. *Interval analysis*. Volume 4. Prentice-Hall Englewood Cliffs, 1966.
- [43] Ramon E Moore. *Methods and applications of interval analysis*. SIAM, 1979.
- [44] R.E. Moore, R.B. Kearfott, and M.J. Cloud. *Introduction to Interval Analysis*. Cambridge University Press, 2009. ISBN: 9780898716696.
- [45] Suhuan Chen, Huadong Lian, and Xiaowei Yang. “Interval static displacement analysis for structures with interval parameters”. In: *International Journal for Numerical Methods in Engineering* 53.2 (2002), pages 393–407. DOI: <https://doi.org/10.1002/nme.281>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.281>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/nme.281>.

- [46] Zhiping Qiu and Isaac Elishakoff. “Antioptimization of structures with large uncertain-but-non-random parameters via interval analysis”. In: *Computer Methods in Applied Mechanics and Engineering* 152.3 (1998), pages 361–372. ISSN: 0045-7825. DOI: [https://doi.org/10.1016/S0045-7825\(96\)01211-X](https://doi.org/10.1016/S0045-7825(96)01211-X). URL: <https://www.sciencedirect.com/science/article/pii/S004578259601211X>.
- [47] B.Y. Ni, C. Jiang, P.G. Wu, Z.H. Wang, and W.Y. Tian. “A sequential simulation strategy for response bounds analysis of structures with interval uncertainties”. In: *Computers & Structures* 266 (2022), page 106785. ISSN: 0045-7949. DOI: <https://doi.org/10.1016/j.compstruc.2022.106785>. URL: <https://www.sciencedirect.com/science/article/pii/S0045794922000451>.
- [48] Robin R.P. Callens, Matthias G.R. Faess, and David Moens. “Multilevel quasi-monte carlo for interval analysis”. In: *International Journal for Uncertainty Quantification* 12.4 (2022), pages 1–19. ISSN: 2152-5080.
- [49] Peihua Ni, Danko J. Jerez, Vasileios C. Frangkoulis, Matthias G. R. Faes, Marcos A. Valdebenito, and Michael Beer. “Operator Norm-Based Statistical Linearization to Bound the First Excursion Probability of Nonlinear Structures Subjected to Imprecise Stochastic Loading”. In: *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* 8.1 (2022), page 04021086. DOI: 10.1061/AJRUA6.0001217. eprint: <https://ascelibrary.org/doi/pdf/10.1061/AJRUA6.0001217>. URL: <https://ascelibrary.org/doi/abs/10.1061/AJRUA6.0001217>.
- [50] Pengfei Wei, Jingwen Song, Sifeng Bi, Matteo Broggi, Michael Beer, Zhenzhou Lu, and Zhufeng Yue. “Non-intrusive stochastic analysis with parameterized imprecise probability models: I. Performance estimation”. In: *Mechanical Systems and Signal Processing* 124 (2019), pages 349–368. ISSN: 0888-3270. DOI: <https://doi.org/10.1016/j.ymsp.2019.01.058>. URL: <https://www.sciencedirect.com/science/article/pii/S0888327019300743>.
- [51] Pengfei Wei, Jingwen Song, Sifeng Bi, Matteo Broggi, Michael Beer, Zhenzhou Lu, and Zhufeng Yue. “Non-intrusive stochastic analysis with parameterized imprecise probability models: II. Reliability and rare events analysis”. In: *Mechanical Systems and Signal Processing* 126 (2019), pages 227–247. ISSN: 0888-3270. DOI: <https://doi.org/10.1016/j.ymsp.2019.02.015>. URL: <https://www.sciencedirect.com/science/article/pii/S0888327019300986>.
- [52] Jie Li and Jianbing Chen. *Stochastic dynamics of structures*. John Wiley & Sons, 2009.
- [53] T.T. Soong and M. Grigoriu. *Random Vibration of Mechanical and Structural Systems*. PTR Prentice Hall, 1993. ISBN: 9780137523610.
- [54] Alan Powell and SH Crandall. *Random Vibration*. The Technology Press of the Massachusetts Institute of Technology, Cambridge, 1958.

- [55] Yu-Kweng Lin and Guo-Qiang Cai. *Probabilistic structural dynamics: advanced theory and applications*. McGraw-Hill New York, 1995.
- [56] Petre Stoica, Randolph L Moses, et al. “Spectral analysis of signals”. In: (2005).
- [57] Yi-Ming Zhang, Zifeng Huang, and Yong Xia. “An adaptive multi-taper spectral estimation for stationary processes”. In: *Mechanical Systems and Signal Processing* 183 (2023), page 109629. ISSN: 0888-3270. DOI: <https://doi.org/10.1016/j.ymssp.2022.109629>. URL: <https://www.sciencedirect.com/science/article/pii/S0888327022007178>.
- [58] George W Housner. “Characteristics of strong-motion earthquakes”. In: *Bulletin of the Seismological Society of America* 37.1 (1947), pages 19–31.
- [59] Kiyoshi Kanai. “Semi-empirical formula for the seismic characteristics of the ground”. In: *Bulletin of the Earthquake Research Institute* 35 (1957), pages 309–325.
- [60] Arthur Schuster. “On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena”. In: *Terrestrial Magnetism* 3.1 (1898), pages 13–41.
- [61] Richard A Muller and Gordon J MacDonald. *Ice ages and astronomical causes: data, spectral analysis and mechanisms*. Springer Science & Business Media, 2002.
- [62] P.D. Spanos, I.A. Kougioumtzoglou, and C. Soize. “On the determination of the power spectrum of randomly excited oscillators via stochastic averaging: An alternative perspective”. In: *Probabilistic Engineering Mechanics* 26.1 (2011). Special Issue: Stochastic Methods in Mechanics — Status and Challenges, pages 10–15. ISSN: 0266-8920. DOI: <https://doi.org/10.1016/j.probengmech.2010.06.001>. URL: <https://www.sciencedirect.com/science/article/pii/S0266892010000391>.
- [63] M. S. Bartlett. “Smoothing periodograms from time-series with continuous spectra”. In: *Nature* 161.4096 (1948), pages 686–687.
- [64] M. S. Bartlett. “Periodogram analysis and continuous spectra”. In: *Biometrika* 37.1-2 (June 1950), pages 1–16. ISSN: 0006-3444. DOI: [10.1093/biomet/37.1-2.1](https://doi.org/10.1093/biomet/37.1-2.1). eprint: <https://academic.oup.com/biomet/article-pdf/37/1-2/1/486591/37-1-2-1.pdf>. URL: <https://doi.org/10.1093/biomet/37.1-2.1>.
- [65] Peter Welch. “The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms”. In: *IEEE Transactions on audio and electroacoustics* 15.2 (1967), pages 70–73.
- [66] Ralph Beebe Blackman and John Wilder Tukey. “The Measurement of Power Spectra from the Point of View of Communications Engineering — Part I”. In: *Bell System Technical Journal* 37.1 (1958), pages 185–282. DOI: <https://doi.org/10.1002/j.1538-7305.1958.tb03874.x>.

- [67] Ralph Beebe Blackman and John Wilder Tukey. “The Measurement of Power Spectra from the Point of View of Communications Engineering — Part II”. In: *Bell System Technical Journal* 37.2 (1958), pages 485–569. DOI: <https://doi.org/10.1002/j.1538-7305.1958.tb01530.x>.
- [68] N.R. Lomb. “Least-squares frequency analysis of unequally spaced data”. In: *Astrophys Space Sci* 39 (1976), pages 447–462. DOI: <https://doi.org/10.1007/BF00648343>.
- [69] Jeffrey D Scargle. “Studies in astronomical time series analysis. II-Statistical aspects of spectral analysis of unevenly spaced data”. In: *The Astrophysical Journal* 263 (1982), pages 835–853.
- [70] Maurice B Priestley. “Evolutionary spectra and non-stationary processes”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 27.2 (1965), pages 204–229.
- [71] MB Priestley. “Power spectral analysis of non-stationary random processes”. In: *Journal of Sound and Vibration* 6.1 (1967), pages 86–97.
- [72] David E Newland. “Harmonic wavelet analysis”. In: *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences* 443.1917 (1993), pages 203–225.
- [73] David E Newland. “Harmonic and musical wavelets”. In: *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences* 444.1922 (1994), pages 605–620.
- [74] D. E. Newland. “Ridge and Phase Identification in the Frequency Analysis of Transient Signals by Harmonic Wavelets”. In: *Journal of Vibration and Acoustics* 121.2 (Apr. 1999), pages 149–155. ISSN: 1048-9002. DOI: 10.1115/1.2893957. URL: <https://doi.org/10.1115/1.2893957>.
- [75] Sanjit Kumar Mitra and Yonghong Kuo. *Digital signal processing: a computer-based approach*. Volume 2. McGraw-Hill New York, 2006.
- [76] Liam Comerford, Ioannis A Kougioumtzoglou, and Michael Beer. “An artificial neural network approach for stochastic process power spectrum estimation subject to missing data”. In: *Structural Safety* 52 (2015), pages 150–160. DOI: <https://doi.org/10.1016/j.strusafe.2014.10.001>. URL: <https://www.sciencedirect.com/science/article/pii/S0167473014000915>.
- [77] Liam Comerford, Ioannis A Kougioumtzoglou, and Michael Beer. “On quantifying the uncertainty of stochastic process power spectrum estimates subject to missing data”. In: *International Journal of Sustainable Materials and Structural Systems* 2.1-2 (2015), pages 185–206. DOI: 10.1504/IJSMSS.2015.078358. eprint: <https://www.inderscienceonline.com/doi/pdf/10.1504/IJSMSS.2015.078358>. URL: <https://www.inderscienceonline.com/doi/abs/10.1504/IJSMSS.2015.078358>.

- [78] Liam Comerford, Ioannis A Kougioumtzoglou, and Michael Beer. “Compressive sensing based stochastic process power spectrum estimation subject to missing data”. In: *Probabilistic Engineering Mechanics* 44 (2016), pages 66–76. DOI: <https://doi.org/10.1016/j.pro bengmech.2015.09.015>. URL: <https://www.sciencedirect.com/science/article/pii/S0266892015300436>.
- [79] Guy P Nason, Rainer Von Sachs, and Gerald Kroisandt. “Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 62.2 (2000), pages 271–292.
- [80] Idris A Eckley, Guy P Nason, and Robert L Treloar. “Locally stationary wavelet fields with application to the modelling and analysis of image texture”. In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 59.4 (2010), pages 595–616.
- [81] Pol D Spanos, Jale Tezcan, and Petros Tratskas. “Stochastic processes evolutionary spectrum estimation via harmonic wavelets”. In: *Computer Methods in Applied Mechanics and Engineering* 194.12-16 (2005), pages 1367–1383.
- [82] P.D. Spanos and I.A. Kougioumtzoglou. “Harmonic wavelets based statistical linearization for response evolutionary power spectrum determination”. In: *Probabilistic Engineering Mechanics* 27.1 (2012). The IUTAM Symposium on Nonlinear Stochastic Dynamics and Control, pages 57 –68. ISSN: 0266-8920. DOI: <https://doi.org/10.1016/j.pro bengmech.2011.05.008>. URL: <http://www.sciencedirect.com/science/article/pii/S0266892011000294>.
- [83] Pol D Spanos and Giuseppe Failla. “Wavelets: Theoretical concepts and vibrations related applications”. In: *The Shock and vibration digest* 37.5 (2005), pages 359–376.
- [84] Masanobu Shinozuka and George Deodatis. “Simulation of stochastic processes by spectral representation”. In: *Applied Mechanics Reviews* 44.4 (1991), pages 191–204. DOI: 10.1115/1.3119501.
- [85] M Shinozuka and G Deodatis. “Response variability of stochastic finite element systems”. In: *Journal of Engineering Mechanics* 114.3 (1988), pages 499–519.
- [86] Hiroshi Tajimi. “A statistical method of determining the maximum response of a building structure during an earthquake”. In: *Proceedings of the 2nd world conference of earthquake engineering*. Volume 11. 1960, pages 781–797.
- [87] RW Clough and J Penzien. *Dynamics of Structures*. McGraw-Hill, 1975.
- [88] Willard J. Pierson Jr. and Lionel Moskowitz. “A proposed spectral form for fully developed wind seas based on the similarity theory of S. A. Kitaigorodskii”. In: *Journal of Geophysical Research (1896-1977)* 69.24 (1964), pages 5181–5190. DOI: <https://doi.org/10.1029/JZ069i024p05181>.

- [89] Klaus F Hasselmann, Tim P Barnett, E Bouws, H Carlson, David E Cartwright, K Eake, JA Euring, A Gienapp, et al. “Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP)”. In: *Ergaenzungsheft zur Deutschen Hydrographischen Zeitschrift, Reihe A* (1973).
- [90] M.D. Shields and G. Deodatis. “Estimation of evolutionary spectra for simulation of non-stationary and non-Gaussian stochastic processes”. In: *Computers & Structures* 126 (2013). Uncertainty Quantification in structural analysis and design: To commemorate Professor Gerhart I. Schueller for his life-time contribution in the area of computational stochastic mechanics, pages 149–163. ISSN: 0045-7949. DOI: <https://doi.org/10.1016/j.compstruc.2013.02.007>. URL: <https://www.sciencedirect.com/science/article/pii/S0045794913000552>.
- [91] Jianbing Chen, Weiling Sun, Jie Li, and Jun Xu. “Stochastic Harmonic Function Representation of Stochastic Processes”. In: *Journal of Applied Mechanics* 80.1 (July 2012). 011001. ISSN: 0021-8936. DOI: 10.1115/1.4006936. eprint: https://asmedigitalcollection.asme.org/appliedmechanics/article-pdf/80/1/011001/6076229/jam_80_1_011001.pdf. URL: <https://doi.org/10.1115/1.4006936>.
- [92] Jianbing Chen, Fan Kong, and Yongbo Peng. “A stochastic harmonic function representation for non-stationary stochastic processes”. In: *Mechanical Systems and Signal Processing* 96 (2017), pages 31–44. ISSN: 0888-3270. DOI: <https://doi.org/10.1016/j.ymsp.2017.03.048>. URL: <http://www.sciencedirect.com/science/article/pii/S0888327017301784>.
- [93] Jianbing Chen, Liam Comerford, Yongbo Peng, Michael Beer, and Jie Li. “Reduction of random variables in the Stochastic Harmonic Function representation via spectrum-relative dependent random frequencies”. In: *Mechanical Systems and Signal Processing* 141 (2020), page 106718. ISSN: 0888-3270. DOI: <https://doi.org/10.1016/j.ymsp.2020.106718>. URL: <http://www.sciencedirect.com/science/article/pii/S0888327020301047>.
- [94] Jianwen Liang, Samit Ray Chaudhuri, and Masanobu Shinozuka. “Simulation of non-stationary stochastic processes by spectral representation”. In: *Journal of Engineering Mechanics* 133.6 (2007), pages 616–627.
- [95] M. Shinozuka and G. Deodatis. “Stochastic process models for earthquake ground motion”. In: *Probabilistic Engineering Mechanics* 3.3 (1988), pages 114–123. ISSN: 0266-8920. DOI: [https://doi.org/10.1016/0266-8920\(88\)90023-9](https://doi.org/10.1016/0266-8920(88)90023-9). URL: <https://www.sciencedirect.com/science/article/pii/0266892088900239>.
- [96] Mircea Grigoriu. “On the spectral representation method in simulation”. In: *Probabilistic Engineering Mechanics* 8.2 (1993), pages 75–90. ISSN: 0266-8920. DOI: [https://doi.org/10.1016/0266-8920\(93\)90002-D](https://doi.org/10.1016/0266-8920(93)90002-D). URL: <https://www.sciencedirect.com/science/article/pii/026689209390002D>.

- [97] Mircea Grigoriu. “Evaluation of Karhunen-Loève, Spectral, and Sampling Representations for Stochastic Processes”. In: *Journal of Engineering Mechanics* 132.2 (2006), pages 179–189. DOI: 10.1061/(ASCE)0733-9399(2006)132:2(179). eprint: <https://ascelibrary.org/doi/pdf/10.1061/%28ASCE%290733-9399%282006%29132%3A2%28179%29>. URL: <https://ascelibrary.org/doi/abs/10.1061/%28ASCE%290733-9399%282006%29132%3A2%28179%29>.
- [98] Quanshun Ding, Ledong Zhu, and Haifan Xiang. “An efficient ergodic simulation of multivariate stochastic processes with spectral representation”. In: *Probabilistic Engineering Mechanics* 26.2 (2011), pages 350–356. ISSN: 0266-8920. DOI: <https://doi.org/10.1016/j.probengmech.2010.09.006>. URL: <https://www.sciencedirect.com/science/article/pii/S0266892010000810>.
- [99] He Zhang, Marius Bittner, and Michael Beer. “Method to generate artificial earthquake accelerations with time domain enhancement and attenuation characteristics”. In: *Ain Shams Engineering Journal* 13.3 (2022), page 101606. ISSN: 2090-4479. DOI: <https://doi.org/10.1016/j.asej.2021.09.031>. URL: <https://www.sciencedirect.com/science/article/pii/S2090447921003713>.
- [100] Ning Zhao, Guoqing Huang, Ahsan Kareem, Yongle Li, and Liuliu Peng. “Simulation of ergodic multivariate stochastic processes: An enhanced spectral representation method”. In: *Mechanical Systems and Signal Processing* 161 (2021), page 107949. ISSN: 0888-3270. DOI: <https://doi.org/10.1016/j.ymsp.2021.107949>. URL: <https://www.sciencedirect.com/science/article/pii/S0888327021003447>.
- [101] G. Muscolino, F. Genovese, G. Biondi, and E. Cascone. “Generation of fully non-stationary random processes consistent with target seismic accelerograms”. In: *Soil Dynamics and Earthquake Engineering* 141 (2021), page 106467. ISSN: 0267-7261. DOI: 10.1016/j.soildyn.2020.106467.
- [102] G. G. Fahlman and T. J. Ulrych. “A new method for estimating the power spectrum of gapped data”. In: *Monthly Notices of the Royal Astronomical Society* 199.1 (May 1982), pages 53–65. ISSN: 0035-8711. DOI: 10.1093/mnras/199.1.53. eprint: <https://academic.oup.com/mnras/article-pdf/199/1/53/4167828/mnras199-0053.pdf>. URL: <https://doi.org/10.1093/mnras/199.1.53>.
- [103] Stefan Baisch and Götz H.R. Bokelmann. “Spectral analysis with incomplete time series: an example from seismology”. In: *Computers & Geosciences* 25.7 (1999), pages 739–750. ISSN: 0098-3004. DOI: [https://doi.org/10.1016/S0098-3004\(99\)00026-6](https://doi.org/10.1016/S0098-3004(99)00026-6). URL: <https://www.sciencedirect.com/science/article/pii/S0098300499000266>.
- [104] Piet M.T. Broersen. “Automatic spectral analysis with missing data”. In: *Digital Signal Processing* 16.6 (2006), pages 754–766. ISSN: 1051-2004. DOI: <https://doi.org/10.1016/j.dsp.2006.01.001>. URL: <https://www.sciencedirect.com/science/article/pii/S1051200406000030>.

- [105] Y. Zhang, L. Comerford, I. A Kougoumtzoglou, E. Patelli, and M. Beer. “Uncertainty quantification of power spectrum and spectral moments estimates subject to missing data”. In: *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* 3.4 (2017), page 04017020. DOI: 10.1061/AJRUA6.0000925. eprint: <https://ascelibrary.org/doi/pdf/10.1061/AJRUA6.0000925>. URL: <https://ascelibrary.org/doi/abs/10.1061/AJRUA6.0000925>.
- [106] L Comerford, HA Jensen, F Mayorga, M Beer, and IA Kougoumtzoglou. “Compressive sensing with an adaptive wavelet basis for structural system response and reliability analysis under missing data”. In: *Computers & Structures* 182 (2017), pages 26–40. DOI: <https://doi.org/10.1016/j.compstruc.2016.11.012>. URL: <https://www.sciencedirect.com/science/article/pii/S0045794916304618>.
- [107] Jui-Chung Hung. “A genetic algorithm approach to the spectral estimation of time series with noise and missed observations”. In: *Information Sciences* 178.24 (2008), pages 4632–4643. ISSN: 0020-0255. DOI: <https://doi.org/10.1016/j.ins.2008.08.018>. URL: <https://www.sciencedirect.com/science/article/pii/S0020025508003496>.
- [108] Johan Karlsson and Tryphon T. Georgiou. “Uncertainty Bounds for Spectral Estimation”. In: *IEEE Transactions on Automatic Control* 58.7 (2013), pages 1659–1673. DOI: 10.1109/TAC.2013.2251775.
- [109] Giuseppe Muscolino, Federica Genovese, and Alba Sofi. “Reliability Bounds for Structural Systems Subjected to a Set of Recorded Accelerograms Leading to Imprecise Seismic Power Spectrum”. In: *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* 8.2 (2022), page 04022009. DOI: 10.1061/AJRUA6.0001215.
- [110] G. I. Schuëller. “Efficient Monte Carlo simulation procedures in structural uncertainty and reliability analysis - recent advances”. In: *Structural Engineering and Mechanics* 32.1 (2009), pages 1–20.
- [111] Loren D Lutes and Shahram Sarkani. *Random vibrations: analysis of structural and mechanical systems*. Butterworth-Heinemann, 2004.
- [112] T. Kishida, V. Contreras, Y. Bozorgnia, N. A. Abrahamson, Ahdi, T.D. S.K. Ancheta, D.M. Boore, K.W. Campbell, et al. “NGA-Sub Ground Motion Database”. In: *Proceedings of the Eleventh U.S. National Conference on Earthquake Engineering* (2018).
- [113] Atsushi Nozu. “A super asperity model for the 2011 off the Pacific coast of Tohoku earthquake”. In: *Journal of Japan Association for Earthquake Engineering* 14.6 (2014), pages 36–55.

- [114] Jie Li and Jianbing Chen. “The principle of preservation of probability and the generalized density evolution equation”. In: *Structural Safety* 30.1 (2008), pages 65–77. ISSN: 0167-4730. DOI: <https://doi.org/10.1016/j.strusafe.2006.08.001>. URL: <http://www.sciencedirect.com/science/article/pii/S0167473006000518>.
- [115] Zhiqiang Wan, Jianbing Chen, Jie Li, and Alfredo H.-S. Ang. “An efficient new PDEM-COM based approach for time-variant reliability assessment of structures with monotonically deteriorating materials”. In: *Structural Safety* 82 (2020), page 101878. ISSN: 0167-4730. DOI: <https://doi.org/10.1016/j.strusafe.2019.101878>. URL: <https://www.sciencedirect.com/science/article/pii/S0167473019300207>.
- [116] G. Kost, T. Tellkamp, H. Kamil, A. Gantayat, and F. Weber. “Automated generation of spectrum-compatible artificial time histories”. In: *Nuclear Engineering and Design* 45.1 (1978), pages 243–249. ISSN: 0029-5493. DOI: [https://doi.org/10.1016/0029-5493\(78\)90119-X](https://doi.org/10.1016/0029-5493(78)90119-X). URL: <http://www.sciencedirect.com/science/article/pii/002954937890119X>.
- [117] L. F. Shampine and M. W. Reichelt. “The MATLAB ODE Suite”. In: *SIAM Journal on Scientific Computing* 18.1 (1997), pages 1–22. DOI: 10.1137/S1064827594276424. eprint: <https://doi.org/10.1137/S1064827594276424>. URL: <https://doi.org/10.1137/S1064827594276424>.
- [118] R Bouc. “Forced vibrations of mechanical systems with hysteresis”. In: *Proc. of the Fourth Conference on Nonlinear Oscillations, Prague, 1967*. 1967.
- [119] Yi-Kwei Wen. “Method for random vibration of hysteretic systems”. In: *Journal of the engineering mechanics division* 102.2 (1976), pages 249–263.
- [120] Thomas T Baber and Yi-Kwei Wen. “Random vibration hysteretic, degrading systems”. In: *Journal of the Engineering Mechanics Division* 107.6 (1981), pages 1069–1087.
- [121] Thomas T Baber and Mohammad N Noori. “Random vibration of degrading, pinching systems”. In: *Journal of Engineering Mechanics* 111.8 (1985), pages 1010–1026.
- [122] T. T. Baber and M. N. Noori. “Modeling General Hysteresis Behavior and Random Vibration Application”. In: *ASME. J. Vib., Acoust., Stress, and Reliab.* 108.4 (1986), pages 411–420.
- [123] F. Ma, H. Zhang, A. Bockstedte, G. C. Foliente, and P. Paevere. “Parameter Analysis of the Differential Model of Hysteresis ”. In: *Journal of Applied Mechanics* 71.3 (June 2004), pages 342–349. ISSN: 0021-8936. DOI: 10.1115/1.1668082. eprint: https://asmedigitalcollection.asme.org/appliedmechanics/article-pdf/71/3/342/5471142/342_1.pdf. URL: <https://doi.org/10.1115/1.1668082>.
- [124] John Brian Roberts and Pol D Spanos. *Random vibration and statistical linearization*. Courier Corporation, 2003.

- [125] A. K. Chopra. *Dynamics of structures: Theory and applications to earthquake engineering*. Englewood Cliffs, N.J: Prentice Hall, 1995.
- [126] Gerhart I Schuëller and Masanobu Shinozuka. *Stochastic Methods in Structural Dynamics*. Volume 10. Mechanics: Dynamical Systems. Springer Science & Business Media, 2012. ISBN: 9789400936812.
- [127] Marco Behrendt, Marius Bittner, Liam Comerford, Michael Beer, and Jianbing Chen. “Relaxed power spectrum estimation from multiple data records utilising subjective probabilities”. In: *Mechanical Systems and Signal Processing* 165 (2022), page 108346. ISSN: 0888-3270. DOI: <https://doi.org/10.1016/j.ymsp.2021.108346>. URL: <https://www.sciencedirect.com/science/article/pii/S0888327021007020>.
- [128] M. Shinozuka and C.-M. Jan. “Digital simulation of random processes and its applications”. In: *Journal of Sound and Vibration* 25.1 (1972), pages 111–128. ISSN: 0022-460X. DOI: [https://doi.org/10.1016/0022-460X\(72\)90600-1](https://doi.org/10.1016/0022-460X(72)90600-1).
- [129] J. Chen, Y. Chen, Y. Peng, S. Zhu, M. Beer, and . Comerford. “Stochastic harmonic function based wind field simulation and wind-induced reliability of super high-rise buildings”. In: *Mechanical Systems and Signal Processing* 133 (2019), page 106264. ISSN: 0888-3270. DOI: <https://doi.org/10.1016/j.ymsp.2019.106264>.
- [130] Tadahiro Kishida, Victor Contreras, Yousef Bozorgnia, Norman A Abrahamson, Sean K Ahdi, Timothy D Ancheta, David M Boore, Kenneth W Campbell, et al. “NGA-sub ground motion database”. In: *UCLA* (2021). URL: <https://escholarship.org/uc/item/3bn528xc>.
- [131] Marco Behrendt, Masaru Kitahara, Takeshi Kitahara, Liam Comerford, and Michael Beer. “Data-driven reliability assessment of dynamic structures based on power spectrum classification”. In: *Engineering Structures* 268 (2022), page 114648. ISSN: 0141-0296. DOI: <https://doi.org/10.1016/j.engstruct.2022.114648>.
- [132] Marco de Angelis, Edoardo Patelli, and Michael Beer. “Advanced Line Sampling for efficient robust reliability analysis”. In: *Structural Safety* 52 (2015). Engineering Analyses with Vague and Imprecise Information, pages 170–182. ISSN: 0167-4730. DOI: <https://doi.org/10.1016/j.strusafe.2014.10.002>. URL: <https://www.sciencedirect.com/science/article/pii/S0167473014000927>.
- [133] Matthias G.R. Faes, Marco Daub, Stefano Marelli, Edoardo Patelli, and Michael Beer. “Engineering analysis with probability boxes: A review on computational methods”. In: *Structural Safety* 93 (2021), page 102092. ISSN: 0167-4730. DOI: <https://doi.org/10.1016/j.strusafe.2021.102092>. URL: <https://www.sciencedirect.com/science/article/pii/S0167473021000187>.

- [134] Luis G. Crespo, Brendon K. Colbert, Sean P. Kenny, and Daniel P. Giesy. “On the quantification of aleatory and epistemic uncertainty using Sliced-Normal distributions”. In: *Systems & Control Letters* 134 (2019), page 104560. ISSN: 0167-6911. DOI: <https://doi.org/10.1016/j.sysconle.2019.104560>. URL: <https://www.sciencedirect.com/science/article/pii/S0167691119301707>.
- [135] Brendon K. Colbert, Luis G. Crespo, and Matthew M. Peet. “A Convex Optimization Approach to Improving Suboptimal Hyperparameters of Sliced Normal Distributions”. In: *2020 American Control Conference (ACC)*. 2020, pages 4478–4483. DOI: [10.23919/ACC45564.2020.9147403](https://doi.org/10.23919/ACC45564.2020.9147403).
- [136] Luis G. Crespo, Brendon K. Colbert, Tanner Slager, and Sean P. Kenny. “Robust Estimation of Sliced-Exponential Distributions”. In: *2021 60th IEEE Conference on Decision and Control (CDC)*. 2021, pages 6742–6748. DOI: [10.1109/CDC45484.2021.9683584](https://doi.org/10.1109/CDC45484.2021.9683584).
- [137] M.C. Campi, G. Calafiore, and S. Garatti. “Interval predictor models: Identification and reliability”. In: *Automatica* 45.2 (2009), pages 382–392. ISSN: 0005-1098. DOI: <https://doi.org/10.1016/j.automatica.2008.09.004>. URL: <https://www.sciencedirect.com/science/article/pii/S0005109808004664>.
- [138] Roberto Rocchetta, Qi Gao, and Milan Petkovic. “Soft-constrained interval predictor models and epistemic reliability intervals: A new tool for uncertainty quantification with limited experimental data”. In: *Mechanical Systems and Signal Processing* 161 (2021), page 107973. ISSN: 0888-3270. DOI: <https://doi.org/10.1016/j.ymsp.2021.107973>. URL: <https://www.sciencedirect.com/science/article/pii/S088832702100368X>.
- [139] Jonathan Sadeghi, Marco de Angelis, and Edoardo Patelli. “Efficient training of interval Neural Networks for imprecise training data”. In: *Neural Networks* 118 (2019), pages 338–351. ISSN: 0893-6080. DOI: <https://doi.org/10.1016/j.neunet.2019.07.005>. URL: <https://www.sciencedirect.com/science/article/pii/S0893608019301935>.
- [140] A. Gray, A. Wimbush, M. de Angelis, P.O. Hristov, D. Calleja, E. Miralles-Dolz, and R. Rocchetta. “From inference to design: A comprehensive framework for uncertainty quantification in engineering with limited information”. In: *Mechanical Systems and Signal Processing* 165 (2022), page 108210. ISSN: 0888-3270. DOI: <https://doi.org/10.1016/j.ymsp.2021.108210>. URL: <https://www.sciencedirect.com/science/article/pii/S0888327021005859>.
- [141] Matthias G.R. Faes, Marcos A. Valdebenito, David Moens, and Michael Beer. “Bounding the first excursion probability of linear structures subjected to imprecise stochastic loading”. In: *Computers & Structures* 239 (2020), page 106320. ISSN: 0045-7949. DOI: <https://doi.org/10.1016/j.compstruc.2020.106320>. URL: <https://www.sciencedirect.com/science/article/pii/S0045794920301231>.

- [142] David S Broomhead and David Lowe. *Radial basis functions, multi-variable functional interpolation and adaptive networks*. Technical report. Royal Signals and Radar Establishment Malvern (United Kingdom), 1988.
- [143] Martin D. Buhmann. *Radial Basis Functions: Theory and Implementations*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2003, pages 240–242. DOI: [10.1017/CB09780511543241](https://doi.org/10.1017/CB09780511543241).
- [144] J. Ghosh and A. Nag. “An Overview of Radial Basis Function Networks”. In: *Radial Basis Function Networks 2: New Advances in Design*. Heidelberg: Physica-Verlag HD, 2001, pages 1–36. ISBN: 978-3-7908-1826-0. DOI: [10.1007/978-3-7908-1826-0_1](https://doi.org/10.1007/978-3-7908-1826-0_1).
- [145] S Chen, CFN Cowan, and PM Grant. “Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks”. In: *IEEE Transactions on neural networks* 2.2 (1991), page 303.
- [146] Jooyoung Park and Irwin W. Sandberg. “Approximation and Radial-Basis-Function Networks”. In: *Neural Computation* 5.2 (Mar. 1993), pages 305–316. ISSN: 0899-7667. DOI: [10.1162/neco.1993.5.2.305](https://doi.org/10.1162/neco.1993.5.2.305). eprint: <https://direct.mit.edu/neco/article-pdf/5/2/305/812543/neco.1993.5.2.305.pdf>. URL: <https://doi.org/10.1162/neco.1993.5.2.305>.
- [147] Pramod L. Narasimha, Walter H. Delashmit, Michael T. Manry, Jiang Li, and Francisco Maldonado. “An integrated growing-pruning method for feedforward network training”. In: *Neurocomputing* 71.13 (2008). Artificial Neural Networks (ICANN 2006) / Engineering of Intelligent Systems (ICEIS 2006), pages 2831–2847. ISSN: 0925-2312. DOI: <https://doi.org/10.1016/j.neucom.2007.08.026>.
- [148] Matthias G.R. Faes and Marcos A. Valdebenito. “Fully decoupled reliability-based optimization of linear structures subject to Gaussian dynamic loading considering discrete design variables”. In: *Mechanical Systems and Signal Processing* 156 (2021), page 107616. ISSN: 0888-3270. DOI: <https://doi.org/10.1016/j.ymssp.2021.107616>. URL: <https://www.sciencedirect.com/science/article/pii/S088832702100011X>.
- [149] Christine A Goulet, Tadahiro Kishida, Timothy D Ancheta, Chris H Cramer, Robert B Darragh, Walter J Silva, Youssef MA Hashash, Joseph Harmon, et al. “PEER NGA-East database”. In: *Earthquake Spectra* 37.1_suppl (2021), pages 1331–1353. DOI: [10.1177/87552930211015695](https://doi.org/10.1177/87552930211015695). eprint: <https://doi.org/10.1177/87552930211015695>. URL: <https://doi.org/10.1177/87552930211015695>.
- [150] Michael Beer. “Engineering quantification of inconsistent information”. In: *International Journal of Reliability and Safety* 3.1-3 (2009), pages 174–200. DOI: [10.1504/IJRS.2009.02684](https://doi.org/10.1504/IJRS.2009.02684). eprint: <https://www.inderscienceonline.com/doi/pdf/10.1504/IJRS.2009.02684>. URL: <https://www.inderscienceonline.com/doi/abs/10.1504/IJRS.2009.02684>.

- [151] L. A. Zadeh. “Fuzzy sets”. In: *Information and Control* 8.3 (1965), pages 338–353.
- [152] H. Bandemer and S. Gottwald. *Fuzzy sets, fuzzy logic, fuzzy methods*. Wiley Chichester, 1995.
- [153] H.-J. Zimmermann. *Fuzzy set theory - and its applications*. Springer Science & Business Media, 2011.
- [154] J. Chen and J. Li. “Optimal determination of frequencies in the spectral representation of stochastic processes”. In: *Computational Mechanics* 51.5 (2013), pages 791–806. DOI: <https://doi.org/10.1007/s00466-012-0764-0>.
- [155] J.H. Lin, Y.H. Zhang, and Y. Zhao. “Pseudo Excitation Method and Some Recent Developments”. In: *Procedia Engineering* 14 (2011). The Proceedings of the Twelfth East Asia-Pacific Conference on Structural Engineering and Construction, pages 2453–2458. ISSN: 1877-7058. DOI: <https://doi.org/10.1016/j.proeng.2011.07.308>. URL: <https://www.sciencedirect.com/science/article/pii/S1877705811013853>.
- [156] Rune Brincker and Lingmi Zhang and Palle Andersen. “Modal identification of output-only systems using frequency domain decomposition”. In: *Smart Materials and Structures* 10.3 (2001), pages 441–445. DOI: 10.1088/0964-1726/10/3/303. URL: <https://doi.org/10.1088/0964-1726/10/3/303>.
- [157] Bart Peeters and Guido De Roeck. “Stochastic System Identification for Operational Modal Analysis: A Review”. In: *Journal of Dynamic Systems, Measurement, and Control* 123.4 (Feb. 2001), pages 659–667. ISSN: 0022-0434. DOI: 10.1115/1.1410370. eprint: https://asmedigitalcollection.asme.org/dynamicsystems/article-pdf/123/4/659/5514360/659_1.pdf. URL: <https://doi.org/10.1115/1.1410370>.
- [158] Svend Gade. “Frequency domain techniques for operational modal analysis”. In: *The Shock and Vibration Digest* 38.6 (2006), pages 537–538. URL: <https://link.gale.com/apps/doc/A152997313/AONE?u=anon~9b6dd377&sid=googleScholar&xid=2a59f38d>.
- [159] T. Wang, O. Celik, F.N. Catbas, and L.M. Zhang. “A frequency and spatial domain decomposition method for operational strain modal analysis and its application”. In: *Engineering Structures* 114 (2016), pages 104–112. ISSN: 0141-0296. DOI: <https://doi.org/10.1016/j.engstruct.2016.02.011>. URL: <https://www.sciencedirect.com/science/article/pii/S0141029616000997>.
- [160] Çağlayan Hızal and Engin Aktaş. “Probabilistic investigation of error propagation in frequency domain decomposition-based operational modal analysis”. In: *Structural Control and Health Monitoring* 28.8 (2021), e2759. DOI: <https://doi.org/10.1002/stc.2759>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/stc.2759>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/stc.2759>.

- [161] Bruce R. Ellingwood. “Earthquake risk assessment of building structures”. In: *Reliability Engineering & System Safety* 74.3 (2001), pages 251–262. ISSN: 0951-8320. DOI: [https://doi.org/10.1016/S0951-8320\(01\)00105-3](https://doi.org/10.1016/S0951-8320(01)00105-3). URL: <https://www.sciencedirect.com/science/article/pii/S0951832001001053>.
- [162] Price Code. “Eurocode 8: Design of structures for earthquake resistance-part 1: general rules, seismic actions and rules for buildings”. In: *Brussels: European Committee for Standardization* (2005).
- [163] Dimitrios Vamvatsikos and C. Allin Cornell. “Incremental dynamic analysis”. In: *Earthquake Engineering & Structural Dynamics* 31.3 (2002), pages 491–514. DOI: <https://doi.org/10.1002/eqe.141>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/eqe.141>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/eqe.141>.
- [164] Dimitrios Vamvatsikos and C. Allin Cornell. “Applied Incremental Dynamic Analysis”. In: *Earthquake Spectra* 20.2 (2004), pages 523–553. DOI: 10.1193/1.1737737. eprint: <https://doi.org/10.1193/1.1737737>. URL: <https://doi.org/10.1193/1.1737737>.
- [165] Dimitrios Vamvatsikos and Michalis Fragiadakis. “Incremental dynamic analysis for estimating seismic performance sensitivity and uncertainty”. In: *Earthquake Engineering & Structural Dynamics* 39.2 (2010), pages 141–163. DOI: <https://doi.org/10.1002/eqe.935>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/eqe.935>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/eqe.935>.
- [166] Keith Porter and Robert Kennedy and Robert Bachman. “Creating Fragility Functions for Performance-Based Earthquake Engineering”. In: *Earthquake Spectra* 23.2 (2007), pages 471–489. DOI: 10.1193/1.2720892. eprint: <https://doi.org/10.1193/1.2720892>. URL: <https://doi.org/10.1193/1.2720892>.
- [167] M. G. Sfahani, Hong Guan, and Yew-Chaye Loo. “Seismic Reliability and Risk Assessment of Structures Based on Fragility Analysis - A Review”. In: *Advances in Structural Engineering* 18.10 (2015), pages 1653–1669. DOI: 10.1260/1369-4332.18.10.1653. eprint: <https://doi.org/10.1260/1369-4332.18.10.1653>. URL: <https://doi.org/10.1260/1369-4332.18.10.1653>.
- [168] Irmela Zentner, Max Gündel, and Nicolas Bonfils. “Fragility analysis methods: Review of existing approaches and application”. In: *Nuclear Engineering and Design* 323 (2017), pages 245–258. ISSN: 0029-5493. DOI: <https://doi.org/10.1016/j.nucengdes.2016.12.021>. URL: <https://www.sciencedirect.com/science/article/pii/S0029549316305209>.
- [169] Ghosh, Swarup and Ghosh, Shyamal and Chakraborty, Subrata. “Seismic fragility analysis in the probabilistic performance-based earthquake engineering framework: an overview”. In: *International Journal of Advances in Engineering Sciences and Applied Mathematics* 13.1 (2021), pages 122–135. DOI: 10.1007/s12572-017-0200-y. URL: <https://doi.org/10.1007/s12572-017-0200-y>.

- [170] Sheenthoi Rajkumari, Kanan Thakkar, and Harsh Goyal. “Fragility analysis of structures subjected to seismic excitation: A state-of-the-art review”. In: *Structures* 40 (2022), pages 303–316. ISSN: 2352-0124. DOI: <https://doi.org/10.1016/j.istruc.2022.04.023>. URL: <https://www.sciencedirect.com/science/article/pii/S2352012422002843>.
- [171] Giorgio Lupoi, Paolo Franchin, Alessio Lupoi, and Paolo E. Pinto. “Seismic Fragility Analysis of Structural Systems”. In: *Journal of Engineering Mechanics* 132.4 (2006), pages 385–395. DOI: 10.1061/(ASCE)0733-9399(2006)132:4(385).
- [172] Nikos D. Lagaros, Yiannis Tsompanakis, Prodromos N. Psarropoulos, and Evaggelos C. Georgopoulos. “Computationally efficient seismic fragility analysis of geostructures”. In: *Computers & Structures* 87.19 (2009). Civil-Comp Special Issue, pages 1195–1203. ISSN: 0045-7949. DOI: <https://doi.org/10.1016/j.compstruc.2008.12.001>. URL: <https://www.sciencedirect.com/science/article/pii/S0045794908002721>.
- [173] Swarup Ghosh and Subrata Chakraborty. “Seismic fragility analysis of structures based on Bayesian linear regression demand models”. In: *Probabilistic Engineering Mechanics* 61 (2020), page 103081. ISSN: 0266-8920. DOI: <https://doi.org/10.1016/j.probengech.2020.103081>. URL: <https://www.sciencedirect.com/science/article/pii/S0266892020300680>.
- [174] A.H.M. Muntasir Billah and M. Shahria Alam. “Seismic fragility assessment of highway bridges: a state-of-the-art review”. In: *Structure and Infrastructure Engineering* 11.6 (2015), pages 804–832. DOI: 10.1080/15732479.2014.912243. eprint: <https://doi.org/10.1080/15732479.2014.912243>. URL: <https://doi.org/10.1080/15732479.2014.912243>.
- [175] Mohammad Amin Hariri-Ardebili and Victor E. Saouma. “Seismic fragility analysis of concrete dams: A state-of-the-art review”. In: *Engineering Structures* 128 (2016), pages 374–399. ISSN: 0141-0296. DOI: <https://doi.org/10.1016/j.engstruct.2016.09.034>. URL: <https://www.sciencedirect.com/science/article/pii/S014102961630654X>.
- [176] Sushreyo Misra and Jamie Ellen Padgett. “Seismic Fragility of Railway Bridge Classes: Methods, Models, and Comparison with the State of the Art”. In: *Journal of Bridge Engineering* 24.12 (2019), page 04019116. DOI: 10.1061/(ASCE)BE.1943-5592.0001485.
- [177] Xu Chen, Nailiang Xiang, Zhongguo Guan, and Jianzhong Li. “Seismic vulnerability assessment of tall pier bridges under mainshock-aftershock-like earthquake sequences using vector-valued intensity measure”. In: *Engineering Structures* 253 (2022), page 113732. ISSN: 0141-0296. DOI: <https://doi.org/10.1016/j.engstruct.2021.113732>. URL: <https://www.sciencedirect.com/science/article/pii/S0141029621018186>.

- [178] Abdoul R. Ghotbi and Ertugrul Taciroglu. “Ground motion selection based on a multi-intensity-measure conditioning approach with emphasis on diverse earthquake contents”. In: *Earthquake Engineering & Structural Dynamics* 50.5 (2021), pages 1378–1394. DOI: <https://doi.org/10.1002/eqe.3383>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/eqe.3383>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/eqe.3383>.
- [179] Matthieu Perrault and Philippe Guéguen. “Correlation between Ground Motion and Building Response using California Earthquake Records”. In: *Earthquake Spectra* 31.4 (2015), pages 2027–2046. DOI: 10.1193/062413EQS168M. eprint: <https://doi.org/10.1193/062413EQS168M>. URL: <https://doi.org/10.1193/062413EQS168M>.
- [180] 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”. In: *Earthquake Spectra* 37.1_suppl (2021), pages 1331–1353. DOI: 10.1177/87552930211015695. URL: <https://doi.org/10.1177/87552930211015695>.
- [181] Berry, Michael and Parrish, Myles and Eberhard, Marc. “PEER structural performance database user’s manual (version 1.0)”. In: *University of California, Berkeley* (2004).
- [182] Keith 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, pages 1–8.
- [183] *National Research Institute for Earth Science and Disaster Resilience, NIED K-NET, KiK-net*. 2019. DOI: 10.17598/NIED.0004.
- [184] Evangelos I. Katsanos, Anastasios G. Sextos, and George D. Manolis. “Selection of earthquake ground motion records: A state-of-the-art review from a structural engineering perspective”. In: *Soil Dynamics and Earthquake Engineering* 30.4 (2010), pages 157–169. ISSN: 0267-7261. DOI: <https://doi.org/10.1016/j.soildyn.2009.10.005>. URL: <https://www.sciencedirect.com/science/article/pii/S0267726109001602>.
- [185] Subash Ghimire and Philippe Guéguen and Ariana Astorga. “Analysis of the efficiency of intensity measures from real earthquake data recorded in buildings”. In: *Soil Dynamics and Earthquake Engineering* 147 (2021), page 106751. ISSN: 0267-7261. DOI: <https://doi.org/10.1016/j.soildyn.2021.106751>. URL: <https://www.sciencedirect.com/science/article/pii/S0267726121001731>.
- [186] Burks, Lynne S. and Baker, Jack W. “Validation of Ground-Motion Simulations through Simple Proxies for the Response of Engineered Systems”. In: *Bulletin of the Seismological Society of America* 104.4 (July 2014), pages 1930–1946. ISSN: 0037-1106. DOI: 10.1785/0120130276. eprint: <https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/104/4/1930/2676473/1930.pdf>. URL: <https://doi.org/10.1785/0120130276>.

- [187] Anil K. Jain. “Data clustering: 50 years beyond K-means”. In: *Pattern Recognition Letters* 31.8 (2010). Award winning papers from the 19th International Conference on Pattern Recognition (ICPR), pages 651–666. ISSN: 0167-8655. DOI: <https://doi.org/10.1016/j.patrec.2009.09.011>. URL: <https://www.sciencedirect.com/science/article/pii/S0167865509002323>.
- [188] Mehrdad Honarkhah and Jef Caers. “Stochastic simulation of patterns using distance-based pattern modeling”. In: *Mathematical Geosciences* 42.5 (2010), pages 487–517. DOI: 10.1007/s11004-010-9276-7.
- [189] H. Bolton Seed, Celso Ugas, and John Lysmer. “Site-dependent spectra for earthquake-resistant design”. In: *Bulletin of the Seismological Society of America* 66.1 (Feb. 1976), pages 221–243. ISSN: 0037-1106. DOI: 10.1785/BSSA0660010221. eprint: <https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/66/1/221/5320247/bssa0660010221.pdf>. URL: <https://doi.org/10.1785/BSSA0660010221>.
- [190] Saman Yaghmaei-Sabegh. “A novel approach for classification of earthquake ground-motion records”. In: *Journal of Seismology* 21.4 (2017), pages 885–907. DOI: 10.1007/s10950-017-9642-8.
- [191] William B. Joyner and David M. Boore. “Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 imperial valley, California, earthquake”. In: *Bulletin of the Seismological Society of America* 71.6 (Dec. 1981), pages 2011–2038. ISSN: 0037-1106. DOI: 10.1785/BSSA0710062011. eprint: <https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/71/6/2011/5330048/bssa0710062011.pdf>. URL: <https://doi.org/10.1785/BSSA0710062011>.
- [192] Yanqiong Ding, Yongbo Peng, and Jie Li. “A stochastic semi-physical model of seismic ground motions in time domain”. In: *Journal of Earthquake and Tsunami* 12.03 (2018), page 1850006. DOI: 10.1142/S1793431118500069. eprint: <https://doi.org/10.1142/S1793431118500069>. URL: <https://doi.org/10.1142/S1793431118500069>.
- [193] Yanqiong Ding, Yongbo Peng, and Jie Li. “Cluster Analysis of Earthquake Ground-Motion Records and Characteristic Period of Seismic Response Spectrum”. In: *Journal of Earthquake Engineering* 24.6 (2020), pages 1012–1033. DOI: 10.1080/13632469.2018.1453420. eprint: <https://doi.org/10.1080/13632469.2018.1453420>. URL: <https://doi.org/10.1080/13632469.2018.1453420>.
- [194] Arzhang Alimoradi, Shahram Pezeshk, Farzad Naeim, and Hichem Frigui. “Fuzzy pattern classification of strong ground motion records”. In: *Journal of Earthquake Engineering* 9.3 (2005), pages 307–332. DOI: 10.1080/13632460509350544. eprint: <https://doi.org/10.1080/13632460509350544>. URL: <https://doi.org/10.1080/13632460509350544>.

- [195] Alireza Azarbakht and Zahra Minaei. “Adaptive Fuzzy C-Mean Clustering of Ground Motion Prediction Equations”. In: *Journal of Seismology and Earthquake Engineering* 19.1 (2017), pages 25–35. ISSN: 1735-1669. eprint: http://www.jsee.ir/article_240753_0a1a7663be54d1d446a2b60802644cb2.pdf. URL: http://www.jsee.ir/article_240753.html.
- [196] Anil Bhattacharyya. “On a measure of divergence between two statistical populations defined by their probability distributions”. In: *Bull. Calcutta Math. Soc.* 35 (1943), pages 99–109.
- [197] A. Bhattacharyya. “On a Measure of Divergence between Two Multinomial Populations”. In: *Sankhyā: The Indian Journal of Statistics (1933-1960)* 7.4 (1946), pages 401–406. ISSN: 00364452. URL: <http://www.jstor.org/stable/25047882>.
- [198] Sifeng Bi, Saurabh Prabhu, Scott Cogan, and Sez Atamturktur. “Uncertainty Quantification Metrics with Varying Statistical Information in Model Calibration and Validation”. In: *AIAA Journal* 55.10 (2017), pages 3570–3583. DOI: 10.2514/1.J055733. eprint: <https://doi.org/10.2514/1.J055733>. URL: <https://doi.org/10.2514/1.J055733>.
- [199] Peter J. Rousseeuw. “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis”. In: *Journal of Computational and Applied Mathematics* 20 (1987), pages 53–65. ISSN: 0377-0427. DOI: [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7). URL: <https://www.sciencedirect.com/science/article/pii/0377042787901257>.
- [200] Leonard Kaufman and Peter J Rousseeuw. *Finding groups in data: an introduction to cluster analysis*. Volume 344. John Wiley & Sons, 2009. DOI: 10.1002/9780470316801.
- [201] Channamma Patil and Ishwar Baidari. “Estimating the Optimal Number of Clusters k in a Dataset Using Data Depth”. In: *Data Science and Engineering* 4.2 (2019), pages 132–140. DOI: 10.1007/s41019-019-0091-y.
- [202] Japan Road Association. “Specifications for highway bridges, Part V: Seismic Design”. In: *Maruzen, Tokyo, Japan* (2016).
- [203] Japan Road Association. “Manual on bearings for highway bridges (in Japanese)”. In: *Maruzen, Tokyo, Japan* (2004).
- [204] Toshikazu Takeda, Mete A. Sozen, and N. Norby Nielsen. “Reinforced Concrete Response to Simulated Earthquakes”. In: *Journal of the Structural Division* 96.12 (1970), pages 2557–2573. DOI: 10.1061/JSDEAG.0002765. eprint: <https://ascelibrary.org/doi/pdf/10.1061/JSDEAG.0002765>. URL: <https://ascelibrary.org/doi/abs/10.1061/JSDEAG.0002765>.
- [205] I.N. Sneddon. *Fourier Transforms*. Dover Books on Mathematics. Dover Publications, 1995. ISBN: 9780486685229.

- [206] J. F. James. *A Student's Guide to Fourier Transforms: With Applications in Physics and Engineering*. 3rd edition. Student's Guides. Cambridge University Press, 2011. DOI: 10.1017/CB09780511762307.
- [207] James W. Cooley and John W. Tukey. "An Algorithm for the Machine Calculation of Complex Fourier Series". In: *Mathematics of Computation* 19.90 (1965), pages 297–301. ISSN: 00255718, 10886842. URL: <http://www.jstor.org/stable/2003354>.
- [208] Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to algorithms*. MIT press, 2009.
- [209] Josef Stoer and Roland Bulirsch. *Introduction to numerical analysis*. Volume 12. Springer Science & Business Media, 2013.
- [210] Scott Ferson, Vladik Kreinovich, Janos Hajagos, William Louis Oberkampf, and Lev Ginzburg. "Experimental uncertainty estimation and statistics for data having interval uncertainty". In: (May 2007). DOI: 10.2172/910198.
- [211] Andrea Wiencierz. "Regression analysis with imprecise data". PhD thesis. Ludwig-Maximilians-Universität München, 2013. DOI: 10.5282/edoc.16678.
- [212] T. Augustin, F.P.A. Coolen, G. de Cooman, and M.C.M. Troffaes. *Introduction to Imprecise Probabilities*. Wiley Series in Probability and Statistics. Wiley, 2014. ISBN: 9781118763148.
- [213] Guoqing Liu and Vladik Kreinovich. "Fast convolution and Fast Fourier Transform under interval and fuzzy uncertainty". In: *Journal of Computer and System Sciences* 76.1 (2010). Special Issue on Intelligent Data Analysis, pages 63–76. ISSN: 0022-0000. DOI: <https://doi.org/10.1016/j.jcss.2009.05.006>.
- [214] Marco De Angelis, Marco Behrendt, Liam Comerford, Yuanjin Zhang, and Michael Beer. "Forward interval propagation through the discrete Fourier transform". In: *The 9th international workshop on Reliable Engineering Computing*. 2021, pages 39–52. arXiv: 2012.09778 [eess.SP].
- [215] M De Angelis, V Ricciardi, and E Dalmau. "Uncertainty estimation of road-dust emissions via interval statistics". In: *Journal of Physics: Conference Series* 1065 (2018), page 212023. DOI: 10.1088/1742-6596/1065/21/212023.
- [216] Steffen Schön and Hansjörg Kutterer. "Using Zonotopes for Overestimation-Free Interval Least-Squares—Some Geodetic Applications". In: *Reliable Computing* 11.2 (2005), pages 137–155. DOI: <https://doi.org/10.1007/s11155-005-3034-4>.
- [217] Olga Kosheleva and Vladik Kreinovich. "Low-Complexity Zonotopes Can Enhance Uncertainty Quantification (UQ)". In: *Proceedings of the 4th International Conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP 2021)*. 2021, pages 26–34. DOI: <https://doi.org/10.7712/120221.8019.18856>.

- [218] Ander Gray, Marco De Angelis, Scott Ferson, and Edoardo Patelli. “What’s Z-X, when Z= X+Y? Dependency tracking in interval arithmetic with bivariate sets”. In: *The 9th international workshop on Reliable Engineering Computing*. 2021, pages 27–38.
- [219] Scott Ferson. “What Monte Carlo methods cannot do”. In: *Human and Ecological Risk Assessment: An International Journal* 2.4 (1996), pages 990–1007. DOI: 10.1080/10807039609383659.
- [220] Marco De Angelis, Scott Ferson, Edoardo Patelli, and Vladik Kreinovich. “Black-box propagation of failure probabilities under epistemic uncertainty”. In: *Proceedings of the 3rd International Conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP 2019)*. 2019, pages 713–723. DOI: <https://doi.org/10.7712/120219.6373.18699>.
- [221] Marco Behrendt, Marco de Angelis, Liam Comerford, Yuanjin Zhang, and Michael Beer. “Projecting interval uncertainty through the discrete Fourier transform: An application to time signals with poor precision”. In: *Mechanical Systems and Signal Processing* 172 (2022), page 108920. ISSN: 0888-3270. DOI: <https://doi.org/10.1016/j.ymssp.2022.108920>. URL: <https://www.sciencedirect.com/science/article/pii/S0888327022001054>.

Curriculum Vitae

Personal details

Name	Marco Behrendt
Date of birth	18 May 1989
Place of birth	Neustadt am Rübenberge, Germany
Nationality	German

Education

since 01/2018	Ph.D. student Institute for Risk and Reliability, Leibniz Universität Hannover
10/2014 – 12/2016	Master of Science in Computational Engineering Leibniz Universität Hannover Thesis: <i>Coupling of Domain Decomposition Algorithms with the Particle Swarm Optimization</i>
10/2009 – 09/2014	Bachelor of Science in Computational Engineering Leibniz Universität Hannover Thesis: <i>Implementation and Generation of Heterogeneous Parameter Fields for a Two-Phase Flow Model</i>

Work experience

since 01/2018	Research assistant Institute for Risk and Reliability, Leibniz Universität Hannover Research project: <i>Uncertainty modelling in power spectrum estimation of environmental processes subject to high rise building performance evaluation</i>
09/2018 – 11/2021	Research associate Institute for Risk and Uncertainty, University of Liverpool, United Kingdom
03/2017 – 12/2017	Student assistant Institute for Risk and Reliability, Leibniz Universität Hannover