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]

[a] *Institute for Risk and Reliability, Leibniz Universität Hannover, Callinstraße 34, 30167 Hannover, Germany*
[b] *Chair for Reliability Engineering, TU Dortmund University, Leonhard-Euler-Straße 5, 44227 Dortmund, Germany*
[c] *Institute for Risk and Uncertainty, University of Liverpool, Peach Street, Liverpool L69 7ZF, United Kingdom*
[d] *International Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University, 1239 Siping Road, Shanghai 200092, China*

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.

---

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 [1, 2, 3, 4]. Buildings and structures are subject to random vibrations induced, for example, by environmental processes such as earthquakes or wind loads [5, 6, 7]. 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 components 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 [8, 9]. 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 [10, 11] and other advanced sampling techniques such as subset sampling [12], line sampling [13], directional importance sampling [14] 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 [15, 16]. Transformations based on estimations, such as a PSD estimation, can introduce additional uncertainties, as some of these estimators may provide results of poor quality [8]. 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 [17]. 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 [4, 18], non-probabilistic models [19] or imprecise probabilistic models [20]. Specific methods are, for instance, p-boxes [21], which are used to bound the cumulative distribution function of an uncertain parameter, sliced-normal [22, 23] or sliced-exponential [24] 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 [25, 26] which can also be combined with interval neural networks [27]. A framework for uncertainty quantification with limited information is given in [28]. Other works use operator norm theory to reliably determine first passage problems under imprecise loads [29, 30, 31].

Some approaches to estimate the PSD functions that account for uncertainties in the data have already been presented. For example, in [32, 33] 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 [34], 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 [35]. 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 [36] 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 as the shape is approximated to represent the overall distribution and magnitude of the individual frequencies. In particular, this means that peak frequencies, for instance, are adequately represented. The approximation by means of the bounds is able to represent this behaviour. Dependencies between the frequency components are also taken into account by this approach. Discontinuities between two neighbouring frequencies are unlikely, but these can occur when estimating the PSD, especially when only limited data is available. By approximating the bounds using an optimisation, these discontinuities are avoided. In addition, individual smooth PSD functions can be generated from the weights and basis functions of the RBF network to represent the data set. Since it is very unlikely that the spectral densities of a PSD function alternate between two frequencies between the upper and lower bounds, discontinuities are thus avoided. The premise for this approach is data similarity. A method for determining the spectral similarity for such a data set is given in [37]. 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 2. The proposed imprecise PSD model is described in Section 3. This approach is illustrated by means of two academic examples in Section 4 and a set of real data records in Section 5. The paper concludes with Section 6.

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.

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., [38].

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) [39]. 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

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

where

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

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 power of the PSD function [39]. 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 [3, 9], which is formed by the squared absolute value of the discrete Fourier transform of the signal $x(t)$. The periodogram reads as follows

$$\hat{S}_X(\omega_k) = \frac{1}{N_t} \left| \sum_{j=0}^{N_t-1} x(j) e^{-\frac{i2\pi}{N_t} kj} \right|^2 , \tag{3}$$

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 $j = 0, \ldots, N_t - 1$, 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 [40]. 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 $x_1(t) = x(t^*)$, $x_2(t) = x(t^* + D)$, $\ldots$, $x_K(t) = x(t^* + (K-1)D)$ with $t^* = 0, 1, \ldots, L-1$, 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:

$$P_k(\omega_m) = \frac{1}{L}\left|\sum_{t^*=0}^{L-1} x_k(t^*)W(t^*)e^{-2\pi i m t^*/L}\right|^2 \tag{4}$$

with $k = 1, \ldots, K$ and ω_m analogous to ω_k in Eq. 3. The resulting modified periodograms are averaged to obtain the estimated smoother PSD function.

$$\hat{S}_x^W(\omega_m) = \frac{1}{K}\sum_{k=1}^{K} P_k(\omega_m) \tag{5}$$

The selection of the window function can be chosen according to the PSD estimation requirements. Two window functions are suggested in [40], which are

$$W_1(j) = 1 - \left(\frac{j - \frac{L-1}{2}}{\frac{L+1}{2}}\right)^2 \tag{6}$$

and

$$W_1(j) = 1 - \left|\frac{j - \frac{L-1}{2}}{\frac{L+1}{2}}\right|, \tag{7}$$

with $j = 0, 1, \ldots, L-1$. 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.

2.2. Radial basis function networks

An RBF network is a class of artificial neural networks [36]. 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 $y : \mathbb{R}^{N_\omega} \to \mathbb{R}$.

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 radial basis functions, which are characterised by the fact that they are symmetrical around their assigned centre c_i. In this work, the RBF

$$\phi_i(x) = e^{-\left(||x-c_i||\cdot b_{\phi_i}\right)^2} \tag{8}$$

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

The function values of the radial basis functions 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

$$y(x) = \sum_{i=1}^{N_B} w_i \; \phi_i(||x - c_i|| \cdot b_{\phi_i}) + b_0 \qquad x \in \mathbb{R}^{N_\omega}. \tag{9}$$

For an exact interpolation of a function, the number of basis functions 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 basis functions 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 [41, 42, 43, 44] and the references therein.

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 optimisation 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 steps given in Fig. 1. These steps will be discussed in the subsequent sections in details.

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

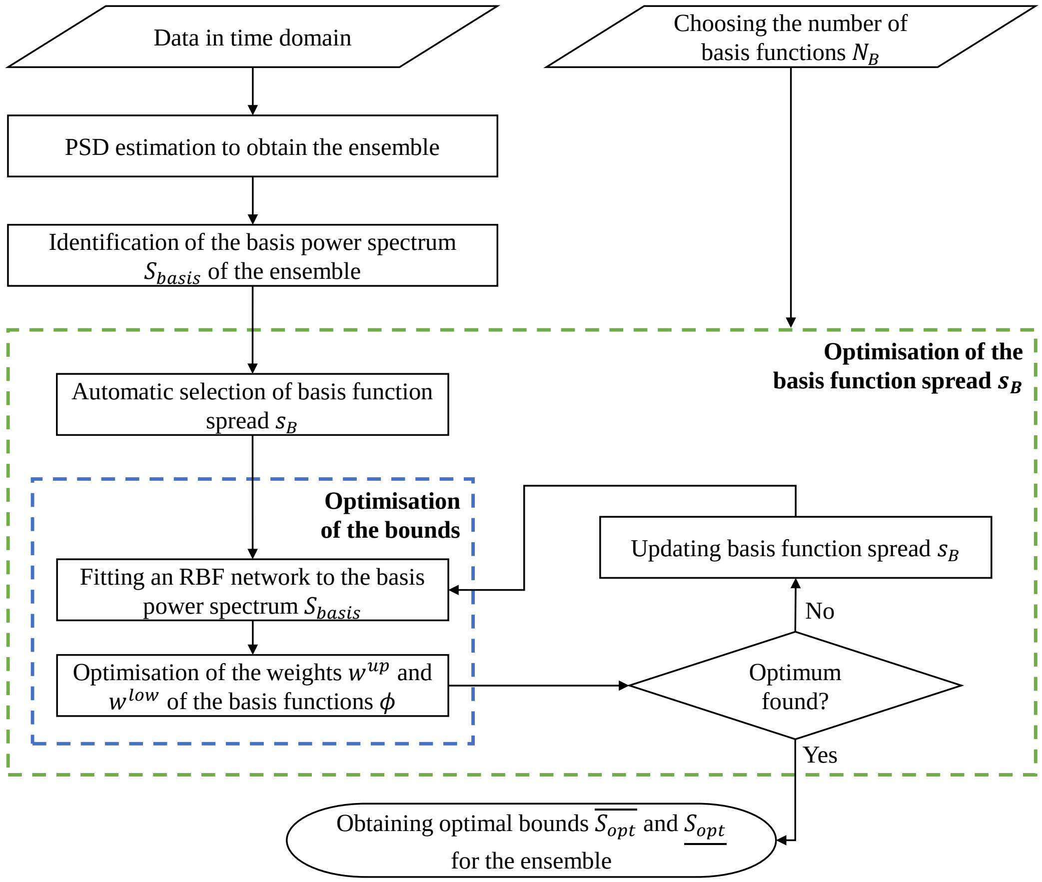

Figure 1: Scheme for computing the optimal bounds.

spectrum. The mean spectrum can be obtained by

$$S_{mean}(\omega_n) = \frac{1}{R}\sum_{i=1}^{R} S^{(i)}(\omega_n), \tag{10}$$

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

$$S_{midpoint}(\omega_n) = \frac{1}{2}\left(S_{max}(\omega_n) + S_{min}(\omega_n)\right). \tag{11}$$

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.

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 results 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 the 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 constrains 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., [43, 45, 42] and references therein. Since the fitting of the RBF network is followed by the optimisation of the bounds, 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 4.2 the influence of different hyperparameters on the resulting bounds is discussed and in Section 3.4 an optimisation of the hyperparameters is suggested.

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 $w^{up} \in \mathbb{R}^{N_B}$ and $w^{low} \in \mathbb{R}^{N_B}$ 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. 9) must be adapted for the following optimisation problem. The upper bound thus results in

$$\overline{S_{opt}}(\omega_n; w^{up}) = \sum_{i=1}^{N_B} w_i^{up} \phi_i + b_0 \tag{12}$$

and the lower bound is

$$\underline{S_{opt}}(\omega_n; w^{low}) = \sum_{i=1}^{N_B} w_i^{low} \phi_i + b_0 \tag{13}$$

with ω_n and n as defined in Eq. 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 $|w^{up}| + \left|w^{low}\right| = 2N_B$, where $|\cdot|$ is the cardinality.

To ensure that representative and optimal bounds are derived for the data set, the Euclidean norm of the difference between the upper and lower bound will be the objective function for the optimisation. This optimisation is subject to the conditions, 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 as negative values are not possible in terms of power spectral densities. Since the weights are to be used as intervals in subsequent simulations, it also must be ensured that the weights for the lower bound are smaller than those for the upper bound. Thus, the optimisation problem results as follows

$$
\begin{aligned}
\min \quad & \left|\left|\overline{S_{opt}}(\omega_n; w^{up}) - \underline{S_{opt}}(\omega_n; w^{low})\right|\right| \\
\text{s.t.} \quad & \overline{S_{opt}}(\omega_n; w^{up}) \geq S_{max}(\omega_n) \\
& \underline{S_{opt}}(\omega_n; w^{low}) \leq S_{min}(\omega_n) \\
& \underline{S_{opt}}(\omega_n; w^{low}) \geq 0 \\
& w^{low} \leq w^{up}
\end{aligned}
\tag{14}
$$

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

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 [46] for the hyperparameters and a non-linear constrained optimisation for the bounds, see for instance [47]. However, various problems arise, for example that several local minima exist, which makes it difficult even for advanced algorithms 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 [48]. 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 more data enclosing bounds, fewer basis functions for a more efficient interval propagation – is left to the analyst.

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

$$S_1(\omega) = S_0 \frac{1 + 4\xi^2 \frac{\omega^2}{\omega_p^2}}{\left(1 - \frac{\omega^2}{\omega_p^2}\right)^2 + 4\xi^2 \frac{\omega^2}{\omega_p^2}} \tag{15}$$

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 [49, 50]. 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 [39].

$$S_2(\omega) = \frac{1}{4}\sigma^2 b^3 \omega^2 e^{-b|\omega|} \tag{16}$$

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 [51, 39]. 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. 1, which were then transformed back into the frequency domain using the periodogram as in Eq. 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. 2. The ensembles utilised aim to illustrate the capabilities of the imprecise PSD in dealing with different shapes of datasets, such as a narrower and a more variant dataset, i.e. with low and high spectral variation, respectively. Throughout this work, the respective data sets will be called ensemble A, which was generated from the analytical expression of the Kanai-Tajimi

PSD function in Eq. 15, and ensemble B, which was generated from the PSD function given in Eq. 16.

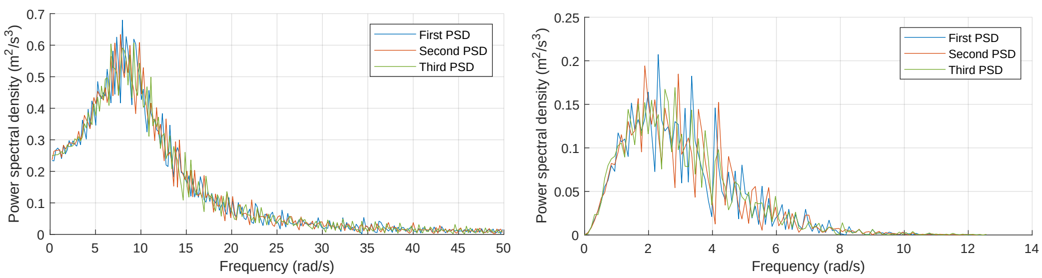

Figure 2: Ensemble A (left), generated from the Kanai-Tajimi PSD function (Eq. 15), and ensemble B (right), generated from the PSD function (Eq. 16), consisting of 3 PSD functions each, utilised to estimate the imprecise PSD function.

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}.

4.1. 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 basis 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. 3. In addition, the calculated basis power spectra (target) and the basis power spectra approximated via the basis functions (output) are given.

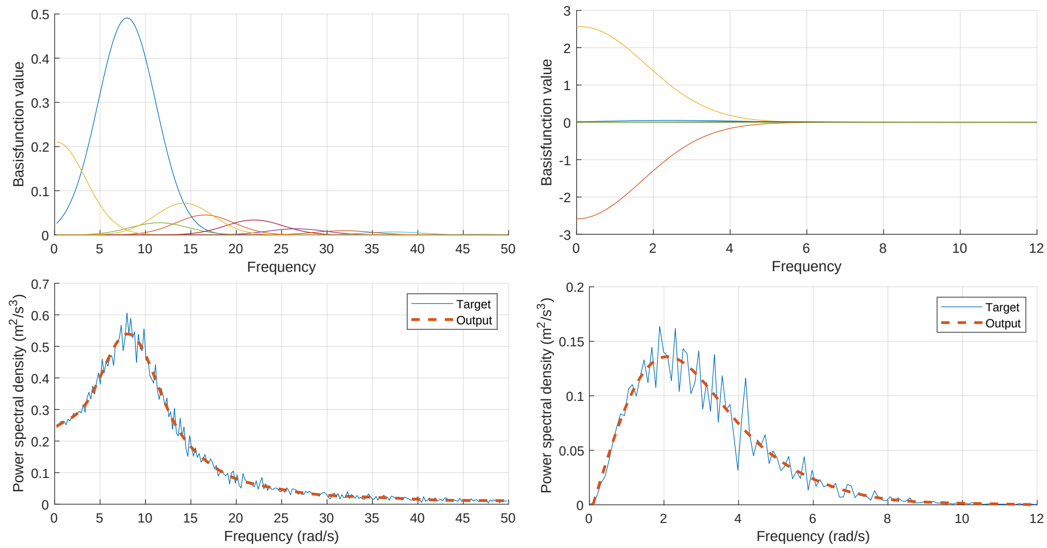

Figure 3: Weighted basis functions used to approximate the basis power spectrum.

The imprecise PSD function for the ensembles given in Fig. 2 are shown in Fig. 4. The final objective function value of the optimisation of the bounds, i.e. the norm between upper bound $\overline{S_{opt}}$ and lower bound $\underline{S_{opt}}$, for ensemble A is 1.825 and for ensemble B 0.885. For comparison, the smallest possible objective function values are 0.8470 for ensemble A and 0.3723 for ensemble B, as this corresponds to the norm between the maximum and minimum values of the ensembles. However, this would require that the number of basis functions is equal to the number of frequency points. In such a case it would be an interpolation rather than an approximation, which is not the aim of this study. The objective function value can therefore also be understood as an indicator of the quality of the optimisation. These objective function values will be of importance for Section 4.2, where the influence of the hyperparameters on the resulting bounds is investigated.

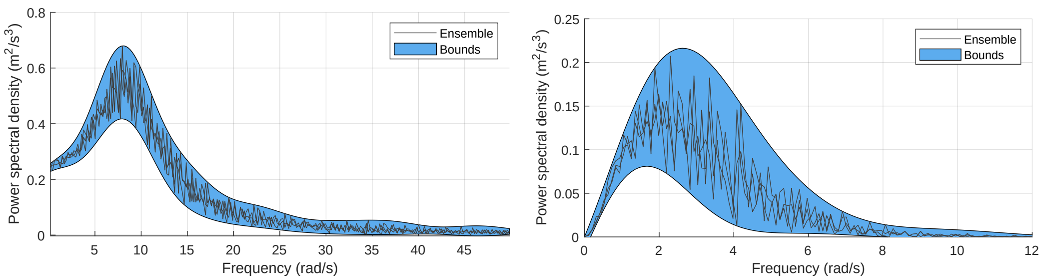

Figure 4: Bounds of the estimated imprecise PSD function.

In some cases, PSD values may lie (by a small margin) outside the optimised bounds or the bounds may take values smaller than 0. 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. Thus, values smaller than zero are an artefact of the optimisation. However, due to tolerances on the optimisation constraints this is justifiable, since the general shape of the PSD and the underlying physics were nevertheless captured very well. Furthermore, these tolerance exceedances often only occur at low PSD values, the influence of which is of minor significance. However, this is more of an implementation problem than a theoretical issue. In the general case, the chosen optimisation schemes lead to robust results. In special cases, other optimisation methods may lead to improved results.

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 function values for ensemble A are depicted in Fig. 5, while those for ensemble B can be obtained in Fig. 6. In both figures, the colour scale is adjusted to reasonable objective function values, i.e. the optimised bounds with such an objective function 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 ensemble can be better captured. This is clearly reflected in the figures as for a high number of basis functions the objective function value decreases, which accordingly means that the bounds are tighter. However, since a lower number of basis functions is desirable in terms of interval propagation, as already stated in Sections 3.4, 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 function 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.

The choice of hyperparameters has a direct influence on the objective function value, as shown in Fig. 5 and Fig. 6, and thus consequently on the

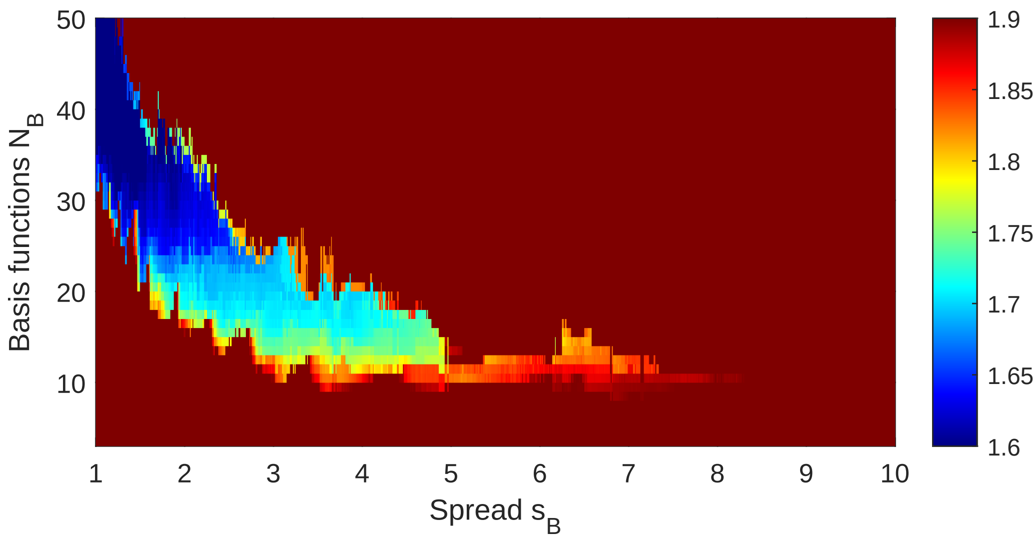

Figure 5: objective function values for ensemble A.

bounds. A high number of basis functions N_B can, with the appropriate selection of the spread s_B, usually represent the ensemble better, so that the bounds enclose the data more closely. In this case, the objective function value will be smaller. A lower number of basis functions, on the other hand, results in a smoother approximation of the ensemble, which also increases the objective function value. For both cases, it can be argued why these are preferable, as the optimisation of the bounds is case dependent and is significantly influenced by the shape of the data.

However, it is also important to note that not every combination of number of basis functions and spread leads to reasonable results, as these are also in direct relation to each other. A high number of basis functions usually needs a lower spread, because significantly more basis functions cover the entire frequency range. A spread that is too high would overlap too many basis functions, making the optimisation more complex. Fig. 5, for example,

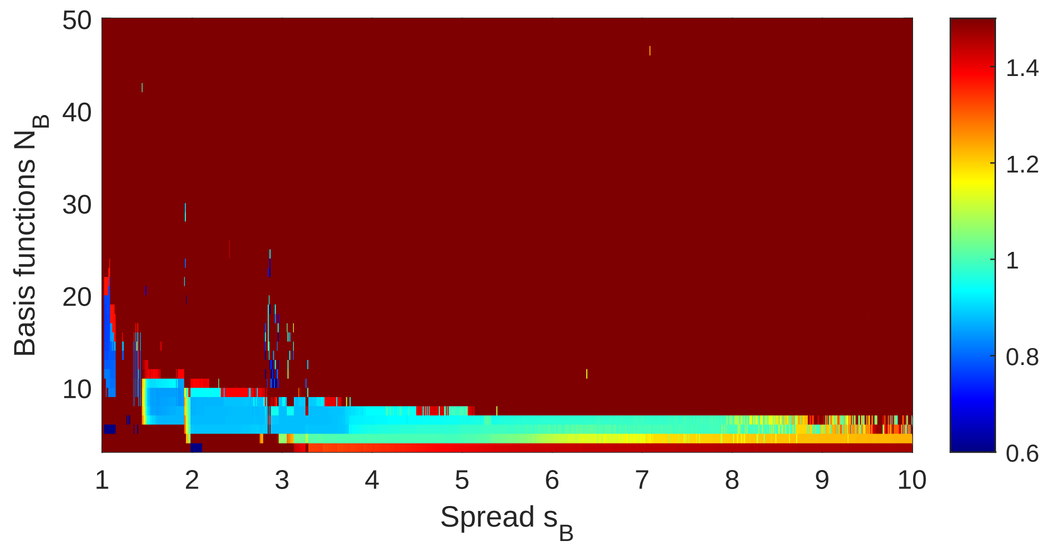

Figure 6: objective function values for ensemble B.

shows that a combination of $N_B = 50$ and $s_B = 10$ leads to undesirable results.

It can be concluded that it is a highly complicated task and challenging for the analyst to find optimal parameters, which motivates to incorporate the optimisation for identifying the parameters such as described in Section 3.4.

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 [52, 53] 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

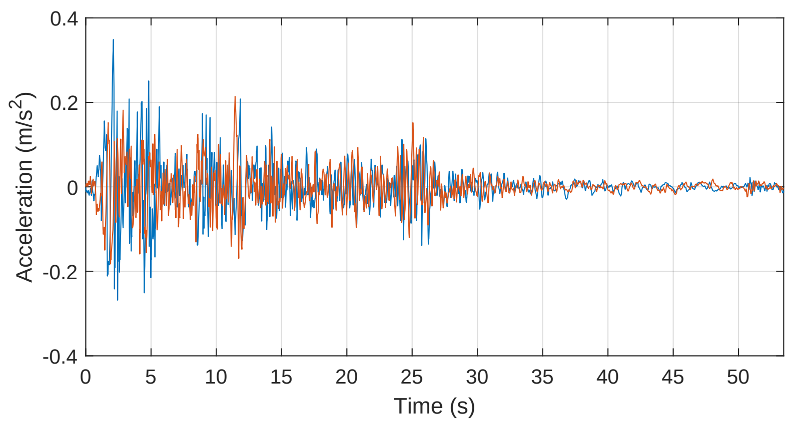

Figure 7: Time records of the El Centro Earthquake.

shown in Fig. 7. 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 ground motions 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. 3), which leads to the PSD functions given in Fig. 8. Using this data set for optimisation poses some problems. The data set shows a high spectral variation, a problem that arises from the use of the periodogram, as already mentioned in Section 2. Due to the high variation, 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 (Eq. 4 and 5) can be used, which usually leads to smoother results

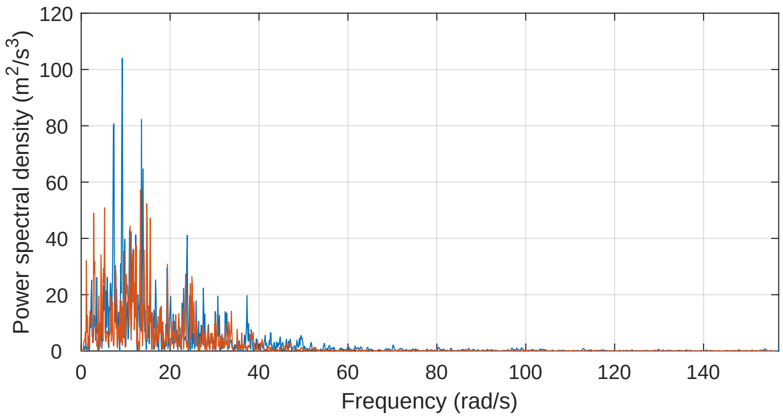

Figure 8: El Centro earthquake records transformed to frequency domain with the periodogram.

by averaging and windowing the input signals. This can be appreciated in Fig. 9.

Another problem is that although the spectral variation 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 [39], it is suggested that ω_U be chosen such that 99% of the total power is still contained in the PSD function. Here, the cut-off frequency is set at 95% of the total power, which is $\omega_U = 50$ rad/s, because, as it can be seen in Fig. 9, a very large amount of frequency components are close to zero.

After appropriate pre-processing of the data set, the bounds can be opti-

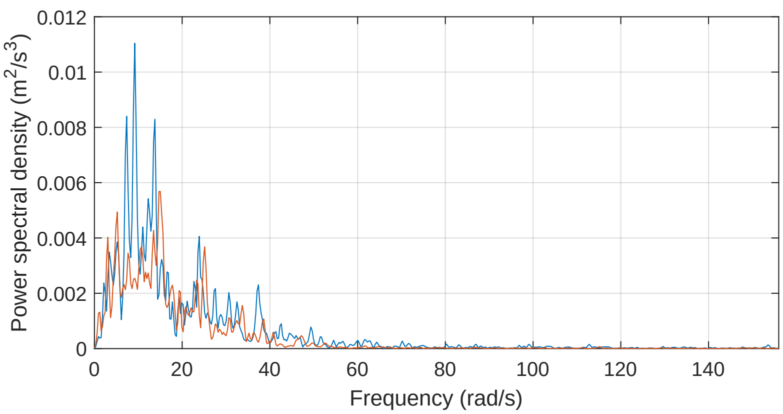

Figure 9: El Centro earthquake records transformed to frequency domain with Welch's method.

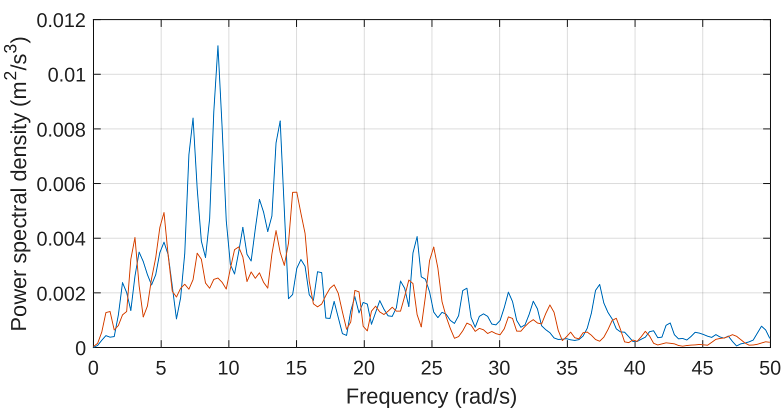

Figure 10: El Centro earthquake records transformed to frequency domain with Welch's method and definition of suitable cut-off frequency.

mised 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. 11. From the optimised bounds it can be seen that the more basis functions are used, the smaller the norm 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 norm between upper and lower bound in exchange to a lower number of basis functions. For comparison, the corresponding optimised spreads s_B and objective function values for the optimised bounds are given in Table 1.

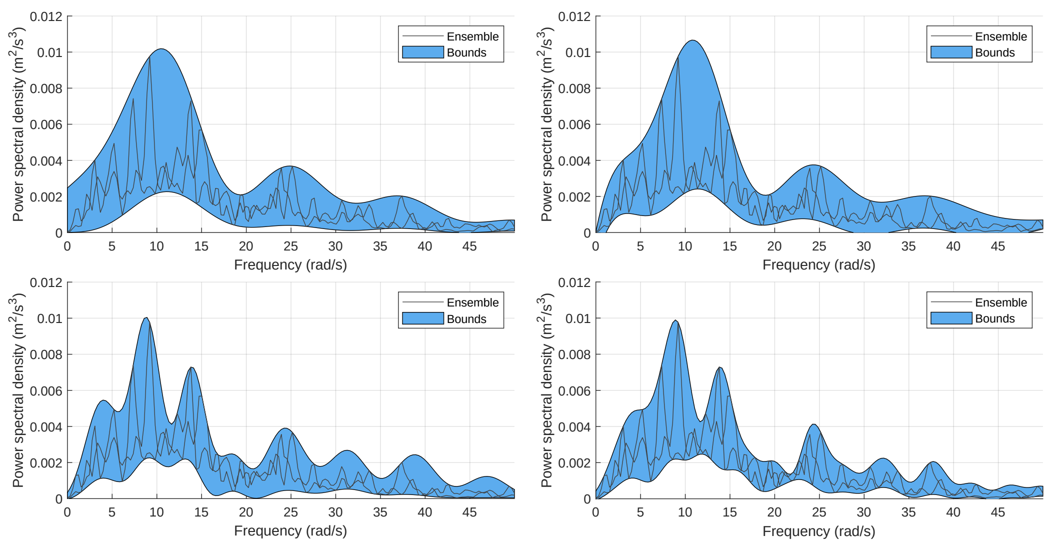

Figure 11: Optimised bounds of the real data set for $N_B \in \{8, 10, 15, 20\}$ basis functions.

To draw the reader's attention to the importance of selecting suitable pa-

Table 1: Optimised spread and resulting objective function value depending on the number of basis functions.

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

rameters 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 highly unrepresentative results. The optimisation 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. 12 (left). For the second counter-examples the optimisation was carried out for the ensemble given in Fig. 8, where the signals were transformed to frequency domain using the periodogram 3. The counter-example is depicted in Fig. 12 (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.

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

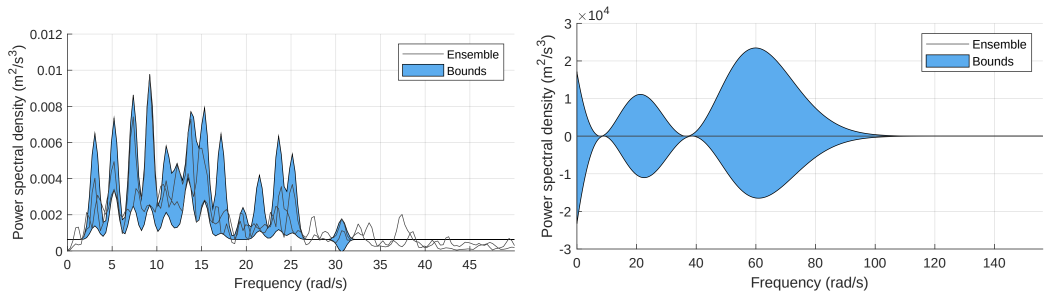

Figure 12: Counter-examples: a poor choice of hyperparameters (left) and a poorly pre-processed data set (right).

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 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.

References

[1] Y.-K. Lin, G.-Q. Cai, Probabilistic Structural Dynamics: Advanced Theory and Applications, McGraw-Hill New York, 1995.

[2] A. K. Chopra, Dynamics of Structures: Theory and Applications to Earthquake Engineering, Prentice-Hall, 1995.

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

[4] M. Grigoriu, Stochastic Calculus: Applications in Science and Engineering, Springer, 2002. doi:https://doi.org/10.1007/978-0-8176-8228-6.

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

[6] L. D. Lutes, S. Sarkani, Random Vibrations: Analysis of Structural and Mechanical Systems, Butterworth-Heinemann, 2004.

[7] T. T. Soong, M. Grigoriu, Random Vibration of Mechanical and Structural Systems, Prentice-Hall, 1993.

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

[9] D. E. Newland, An Introduction to Random Vibrations, Spectral & Wavelet Analysis, Courier Corporation, 2012.

[10] G. I. Schuëller, Efficient monte carlo simulation procedures in structural uncertainty and reliability analysis - recent advances, Structural Engineering and Mechanics 32 (2009) 1–20.

[11] E. Zio, Monte carlo simulation: The method, in: The Monte Carlo simulation method for system reliability and risk analysis, Springer, 2013, pp. 19–58.

[12] S.-K. Au, J. L. Beck, Estimation of small failure probabilities in high dimensions by subset simulation, Probabilistic Engineering Mechanics 16 (2001) 263–277. URL: https://www.sciencedirect.com/science/article/pii/S0266892001000194. doi:https://doi.org/10.1016/S0266-8920(01)00019-4.

[13] M. de Angelis, E. Patelli, M. Beer, Advanced line sampling for efficient robust reliability analysis, Structural Safety 52

(2015) 170–182. URL: https://www.sciencedirect.com/science/article/pii/S0167473014000927. doi:https://doi.org/10.1016/j.strusafe.2014.10.002, engineering Analyses with Vague and Imprecise Information.

[14] M. A. Misraji, M. A. Valdebenito, H. A. Jensen, C. F. Mayorga, Application of directional importance sampling for estimation of first excursion probabilities of linear structural systems subject to stochastic gaussian loading, Mechanical Systems and Signal Processing 139 (2020) 106621. URL: https://www.sciencedirect.com/science/article/pii/S0888327020300078. doi:https://doi.org/10.1016/j.ymssp.2020.106621.

[15] C. Murphy, P. Gardoni, C. E. Harris, Classification and moral evaluation of uncertainties in engineering modeling, Science and Engineering Ethics 17 (2011) 553–570. doi:https://doi.org/10.1007/s11948-010-9242-2.

[16] E. Nikolaidis, D. M. Ghiocel, S. Singhal, Engineering Design Reliability Handbook, 1 ed., CRC press, 2004. doi:https://doi.org/10.1201/9780203483930.

[17] A. D. Kiureghian, O. Ditlevsen, Aleatory or epistemic? does it matter?, Structural Safety 31 (2009) 105–112. URL: https://www.sciencedirect.com/science/article/pii/S0167473008000556. doi:https://doi.org/10.1016/j.strusafe.2008.06.020, risk Acceptance and Risk Communication.

[18] M. Grigoriu, Stochastic systems: Uncertainty Quantification and Propagation, Springer Science & Business Media, 2012. doi:https://doi.org/10.1007/978-1-4471-2327-9.

[19] M. G. Faes, D. Moens, Recent trends in the modeling and quantification of non-probabilistic uncertainty, Archives of Computational Methods in Engineering 27 (2020) 633–671. URL: https://link.springer.com/article/10.1007/s11831-019-09327-x. doi:https://doi.org/10.1007/s11831-019-09327-x.

[20] M. Beer, S. Ferson, V. Kreinovich, Imprecise probabilities in engineering analyses, Mechanical Systems and Signal Processing 37 (2013) 4 – 29. doi:10.1016/j.ymssp.2013.01.024.

[21] M. G. Faes, M. Daub, S. Marelli, E. Patelli, M. Beer, Engineering analysis with probability boxes: A review on computational methods, Structural Safety 93 (2021) 102092. URL: https://www.sciencedirect.com/science/article/pii/S0167473021000187. doi:https://doi.org/10.1016/j.strusafe.2021.102092.

[22] L. G. Crespo, B. K. Colbert, S. P. Kenny, D. P. Giesy, On the quantification of aleatory and epistemic uncertainty using sliced-normal distributions, Systems & Control Letters 134 (2019) 104560. URL: https://www.sciencedirect.com/science/article/pii/S0167691119301707. doi:https://doi.org/10.1016/j.sysconle.2019.104560.

[23] B. K. Colbert, L. G. Crespo, M. M. Peet, A convex optimization ap-

proach to improving suboptimal hyperparameters of sliced normal distributions, in: 2020 American Control Conference (ACC), 2020, pp. 4478–4483. doi:10.23919/ACC45564.2020.9147403.

[24] L. G. Crespo, B. K. Colbert, T. Slager, S. P. Kenny, Robust estimation of sliced-exponential distributions, in: 2021 60th IEEE Conference on Decision and Control (CDC), 2021, pp. 6742–6748. doi:10.1109/CDC45484.2021.9683584.

[25] M. Campi, G. Calafiore, S. Garatti, Interval predictor models: Identification and reliability, Automatica 45 (2009) 382–392. URL: https://www.sciencedirect.com/science/article/pii/S0005109808004664. doi:https://doi.org/10.1016/j.automatica.2008.09.004.

[26] R. Rocchetta, Q. Gao, M. Petkovic, Soft-constrained interval predictor models and epistemic reliability intervals: A new tool for uncertainty quantification with limited experimental data, Mechanical Systems and Signal Processing 161 (2021) 107973. URL: https://www.sciencedirect.com/science/article/pii/S088832702100368X. doi:https://doi.org/10.1016/j.ymssp.2021.107973.

[27] J. Sadeghi, M. de Angelis, E. Patelli, Efficient training of interval neural networks for imprecise training data, Neural Networks 118 (2019) 338–351. URL: https://www.sciencedirect.com/science/article/pii/S0893608019301935. doi:https://doi.org/10.1016/j.neunet.2019.07.005.

[28] A. Gray, A. Wimbush, M. de Angelis, P. Hristov, D. Calleja, E. Miralles-Dolz, R. Rocchetta, From inference to design: A comprehensive framework for uncertainty quantification in engineering with limited information, Mechanical Systems and Signal Processing 165 (2022) 108210. URL: https://www.sciencedirect.com/science/article/pii/S0888327021005859. doi:https://doi.org/10.1016/j.ymssp.2021.108210.

[29] M. G. Faes, M. A. Valdebenito, D. Moens, M. Beer, Bounding the first excursion probability of linear structures subjected to imprecise stochastic loading, Computers & Structures 239 (2020) 106320. URL: https://www.sciencedirect.com/science/article/pii/S0045794920301231. doi:https://doi.org/10.1016/j.compstruc.2020.106320.

[30] M. G. Faes, M. A. Valdebenito, D. Moens, M. Beer, Operator norm theory as an efficient tool to propagate hybrid uncertainties and calculate imprecise probabilities, Mechanical Systems and Signal Processing 152 (2021) 107482. URL: https://www.sciencedirect.com/science/article/pii/S0888327020308682. doi:https://doi.org/10.1016/j.ymssp.2020.107482.

[31] P. Ni, D. J. Jerez, V. C. Fragkoulis, M. G. R. Faes, M. A. Valdebenito, M. Beer, Operator Norm-Based Statistical Linearization to Bound the First Excursion Probability of Nonlinear Structures Subjected to Imprecise Stochastic Loading, ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil En-

gineering 8 (2022) 04021086. URL: https://ascelibrary.org/doi/abs/10.1061/AJRUA6.0001217. doi:10.1061/AJRUA6.0001217. arXiv:https://ascelibrary.org/doi/pdf/10.1061/AJRUA6.0001217.

[32] L. Comerford, I. A. Kougioumtzoglou, M. Beer, On quantifying the uncertainty of stochastic process power spectrum estimates subject to missing data, International Journal of Sustainable Materials and Structural Systems 2 (2015) 185–206. doi:10.1504/IJSMSS.2015.078358.

[33] Y. Zhang, L. Comerford, I. A. Kougioumtzoglou, E. Patelli, M. Beer, Uncertainty quantification of power spectrum and spectral moments estimates subject to missing data, ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 3 (2017) 04017020. doi:10.1061/AJRUA6.0000925.

[34] G. Muscolino, F. Genovese, A. Sofi, Reliability Bounds for Structural Systems Subjected to a Set of Recorded Accelerograms Leading to Imprecise Seismic Power Spectrum, ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 8 (2022) 04022009. doi:10.1061/AJRUA6.0001215.

[35] M. Behrendt, M. Bittner, L. Comerford, M. Beer, J. Chen, Relaxed power spectrum estimation from multiple data records utilising subjective probabilities, Mechanical Systems and Signal Processing 165 (2022) 108346. doi:10.1016/j.ymssp.2021.108346.

[36] D. S. Broomhead, D. Lowe, Radial basis functions, multi-variable func-

tional interpolation and adaptive networks, Technical Report, Royal Signals and Radar Establishment Malvern (United Kingdom), 1988.

[37] M. Behrendt, M. Kitahara, T. Kitahara, L. Comerford, M. Beer, Data-driven reliability assessment of dynamic structures based on power spectrum classification, Engineering Structures 268 (2022) 114648. doi:https://doi.org/10.1016/j.engstruct.2022.114648.

[38] M. Priestley, Spectral Analysis and Time Series, Probability and mathematical statistics : A series of monographs and textbooks, Academic Press, 1982.

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

[40] P. Welch, The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms, IEEE Transactions on audio and electroacoustics 15 (1967) 70–73.

[41] M. D. Buhmann, Radial Basis Functions: Theory and Implementations, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, 2003. doi:10.1017/CBO9780511543241.

[42] J. Ghosh, A. Nag, An Overview of Radial Basis Function Networks, Physica-Verlag HD, Heidelberg, 2001, pp. 1–36. doi:10.1007/978-3-7908-1826-0_1.

[43] S. Chen, C. Cowan, P. Grant, Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks, IEEE Transactions on neural networks 2 (1991) 303.

[44] J. Park, I. W. Sandberg, Approximation and radial-basis-function networks, Neural Computation 5 (1993) 305–316. URL: https://doi.org/10.1162/neco.1993.5.2.305. doi:10.1162/neco.1993.5.2.305. arXiv:https://direct.mit.edu/neco/article-pdf/5/2/305/812543/neco.1993.5.2.305

[45] P. L. Narasimha, W. H. Delashmit, M. T. Manry, J. Li, F. Maldonado, An integrated growing-pruning method for feedforward network training, Neurocomputing 71 (2008) 2831–2847. doi:https://doi.org/10.1016/j.neucom.2007.08.026, artificial Neural Networks (ICANN 2006) / Engineering of Intelligent Systems (ICEIS 2006).

[46] D. Jones, M. Schonlau, W. Welch, Efficient global optimization of expensive black-box functions, Journal of Global Optimization 13 (1998) 455–492. doi:https://doi.org/10.1023/A:1008306431147.

[47] S. S. Rao, Engineering optimization: theory and practice, John Wiley & Sons, 2019.

[48] M. G. Faes, M. A. Valdebenito, Fully decoupled reliability-based optimization of linear structures subject to gaussian dynamic loading considering discrete design variables, Mechanical Systems and Signal Processing 156 (2021) 107616. URL: https://www.sciencedirect.com/science/article/pii/S088832702100011X. doi:https://doi.org/10.1016/j.ymssp.2021.107616.

[49] K. Kanai, Semi-empirical formula for the seismic characteristics of the ground, Bulletin of the Earthquake Research Institute 35 (1957) 309–325.

[50] H. 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, pp. 781–797.

[51] M. Shinozuka, G. Deodatis, Response variability of stochastic finite element systems, Journal of Engineering Mechanics 114 (1988) 499–519.

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

[53] C. A. Goulet, T. Kishida, T. D. Ancheta, C. H. Cramer, R. B. Darragh, W. J. Silva, Y. M. Hashash, J. Harmon, G. A. Parker, J. P. Stewart, R. R. Youngs, PEER NGA-East database, Earthquake Spectra 37 (2021) 1331–1353. URL: https://doi.org/10.1177/87552930211015695. doi:10.1177/87552930211015695. arXiv:https://doi.org/10.1177/87552930211015695.