13th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP13
Seoul, South Korea, May 26-30, 2019

Stochastic Processes Identification from Data Ensembles via Power Spectrum Classification

Marco Behrendt
Institute for Risk and Reliability, Leibniz University Hannover, Germany
Institute for Risk and Uncertainty, School of Engineering, University of Liverpool, United Kingdom

Liam Comerford
Institute for Risk and Uncertainty, School of Engineering, University of Liverpool, United Kingdom

Michael Beer
Institute for Risk and Reliability, Leibniz University Hannover, Germany
Institute for Risk and Uncertainty, School of Engineering, University of Liverpool, United Kingdom
Deptartment of Structural Engineering & Shanghai Institute for Disaster Prevention, College of Civil Engineering, Tongji University, Shanghai, P.R. China

ABSTRACT: Modern approaches to solve dynamic problems where random vibration is of significance will in most of cases rely upon the fundamental concept of the power spectrum as a core model for excitation and response process representation. This is partly due to the practicality of spectral models for frequency domain analysis, as well as their ease of use for generating compatible time domain samples. Such samples may be utilised for numerical performance evaluation of structures, those represented by complex non-linear models. Utilisation of ensemble statistics will be considered first for stationary processes only. For a stationary stochastic process, its power spectrum can be estimated statistically across all time or for a single window in time across an ensemble of records.
In this work, it is first shown that ensemble characteristics can be utilised to improve the resulting power spectra by using estimations of the median instead of the mean of multiple data records. The improved power spectrum will be more robust in the presence of spectral outliers. The median spectrum will result in more reliable response statistics, particularly when source ensemble records contain low power spectra that are significantly below the mean. A weighted median spectrum will also be utilised, based upon the spectral distance of each record from the median, which will shift the estimated spectrum in the direction of the closest samples.
In some cases, the data records exhibit high spectral variance so such an extent that a single power spectrum estimate is insufficient to adequately model the process statistics. In such cases, a more realistic representation of the spectral range of the process is captured by estimating two or more power spectra. This is done by classifying individual process records based upon their individual spectral estimates' distance from each other, and therefore the only parameterisation required is to choose the number of spectrum models to be defined.

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 excited by earthquake and wind loads, or offshore platforms in the ocean excited by waves. To determine the influence of environmental processes on the structures by running simulations, it is necessary to record these influences and apply them to a model with system excitation/response process (Comerford et al. (2015)). For this purpose, e.g. the earthquake ground motion is recorded and utilised for system analysis. However, these real data records 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 earthquake itself, sensor threshold limitations and measurement errors. Additionally the sensor could capture the data incorrectly, e.g. extreme features or other causes such as sensor maintenance, bandwidth limitation or data acquisition restrictions. For this reason, the real data records must be represented in an appropriate manner and the uncertainties mitigated as much as possible (Comerford et al. (2017)).
Since stochastic dynamics have been studied very efficiently in recent decades, different models have been developed. One of these 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 (Powell and Crandall (1958), Lin and Cai (1995), Li and Chen (2009)). In earthquake engineering, for example, the use of the PSD dates back to Housner (1947) or Kanai (1957). At least in the linear case considered in this work, a relationship between the system response can be derived with an elegant relationship between the power spectra of the input data and that of the output data (Chen et al. (2013)).
To develop a load model, that utilises such a random excitation process, the more real data records are available the better, since the numerical results are statistically more accurate for a large amount of data. In addition, the underlying physics should be understood well. Because both cases are often not satisfied, other approaches must be found to develop a load model which represents the data in the best possible way (Comerford et al. (2017)).
One of these approaches is the use of other average statistics rather than the mean value of an ensemble of real data records. In this work the median and weighted median approach are utilised.
In some cases, the power spectra estimated from the real data records may exhibit high spectral variance. Then it might be useful to estimate more than one average power spectra and apply them individually to the system.
This work is structured as following: An explanation how the environmental processes data are available and how a power spectrum is estimated from a stochastic process is presented in section 2. In section 3 is described how the median and weighted median spectrum can be utilised in order to use different average power spectra. In section 4 the problem of real data records with high spectral variance is discussed. Additionally, the estimation of more than one average power spectrum for the application to systems is presented. The final conclusion is given in section 5.

2. STOCHASTIC PROCESS REPRESENTATION AND POWER SPECTRUM ESTIMATION

This section describes the stochastic process and how it can be transformed into a power spectrum density.
Given a real-valued stationary process $X(t)$. For each of these processes exist a corresponding orthogonal process $Z(\omega)$. Thus, $X(t)$ can than be written in the form

$$X(t) = \int_0^\infty e^{i\omega t} dZ(\omega), \tag{1}$$

where $Z(\omega)$ has the following properties:

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

In equation 2, $S_X(\omega)$ describes the two-sided power spectrum of the stationary process $X(t)$, see Comer-

ford et al. (2016). To generate a stationary stochastic process, the following model is considered in this work (Shinozuka and Deodatis (1991)):

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

where

$$\begin{aligned} \omega_n &= n\Delta\omega, \quad n = 0,1,2,\ldots,N-1 \\ \Delta\omega &= \frac{\omega_u}{N} \end{aligned} \quad (4)$$

with $N \to \infty$ and Φ_n as uniformly distributed random phase angles in the range $0 \leq \Phi_n < 2\pi$. For the power spectrum density $S_X(\omega)$

$$S_X(\omega) = \frac{1}{4}\sigma^2 b^3 \omega^2 e^{-b|\omega|}, \quad -\infty < \omega < \infty \quad (5)$$

is used. 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 (Shinozuka and Deodatis (1988)).
An example of a generated stochastic process is depicted in figure 1. To transform a stochastic process

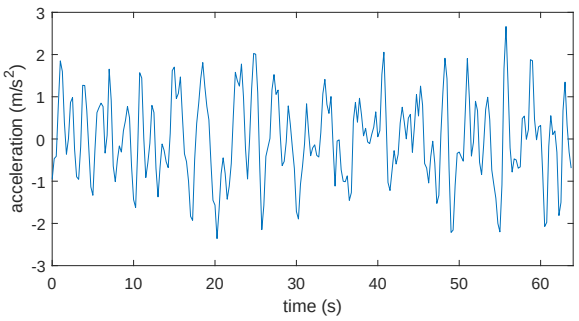

Figure 1: Generated stochastic process

from time domain to frequency domain, the discrete Fourier transform is applied. A frequently used estimator of the power spectrum is the periodogram (Newland (2012)) which can be understood as the squared absolute value of the discrete Fourier transform of the time signal $x(t)$:

$$S_X(\omega_k) = \lim_{T\to\infty} \frac{2\Delta T}{T}\left|\sum_{t=0}^{T-1} x_t e^{-2\pi i k t/T}\right|^2 \quad (6)$$

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 figure 1 is transformed to the power spectrum, which is depicted in figure 2.

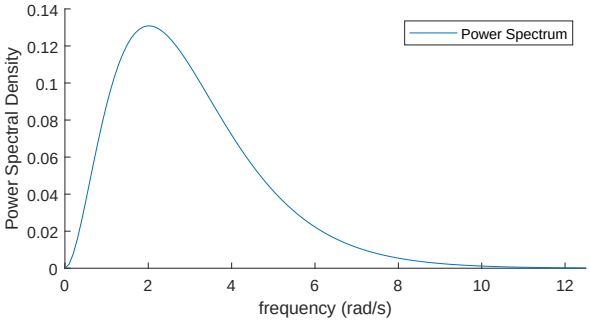

Figure 2: Estimated power spectral density

3. CALCULATION OF AN AVERAGE POWER SPECTRUM

This section considers the calculation of an average spectrum of an ensemble of real data records. In addition to the previously known method of the mean spectrum, the calculation of the median and the weighted median spectrum, respectively, with its advantages is shown. Furthermore, the case is considered that multiple spectral models are developed from a single ensemble of real data records when the ensemble exhibit high spectral variance.

3.1. Mean and Median

To obtain a more reliable response statistics from an ensemble of power spectra for a non-ergodic process, it might be useful to calculate the median instead of the mean. For the linear case considered in this work, the assumption holds that the higher the power of the input spectrum around the natural frequencies of the system, the higher the probability that the system will fail. Especially in the case where the ensemble of power spectra consists of only a few power spectra that are significantly below the mean, the median should be calculated. This results in a higher power compared to the mean, especially for the peak frequencies. Spectral outliers with low power are thus less taken into account. Overall, the median spectrum then results in a higher power than the mean spectrum. This is depicted in figure 3.

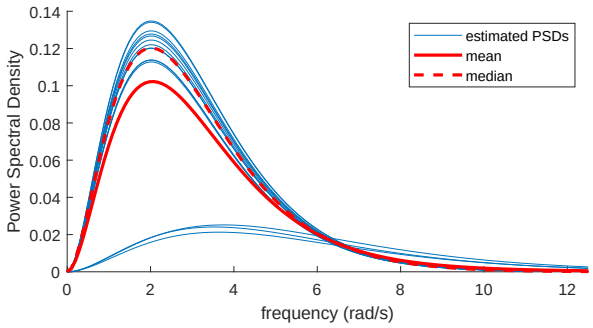

Figure 3: Mean and median spectra with outliers with low power

For the case that only a few power spectra are above the mean value, the mean value should still be used, see figure 4. In this case, the mean leads to a higher power around the natural frequencies of the system.

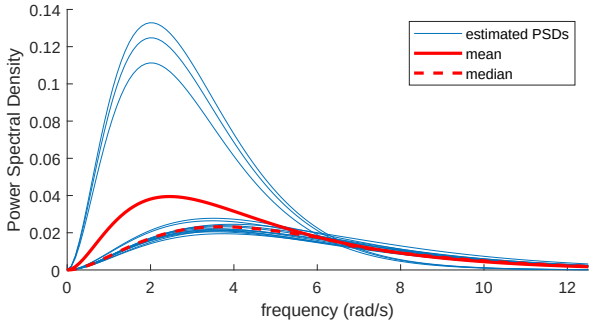

Figure 4: Mean and median spectra with outliers with high power

Particularly, in absence of additional knowledge about the problem or the underlying physics, it is useful to calculate mean and median spectrum and choose the worst case of them, i.e. the one with the higher power around the natural frequencies of the systems. If further knowledge is available, this should always be combined in deciding which spectrum to more crucial.

3.2. Weighted Median

As an extension to the calculation of the median spectrum, a weighted median spectrum can be utilised. The weights are calculated using the inverse spectral distance of the individual power spectra to the median spectrum. The inverse spectral distance is used because nearby power spectra should get a higher weight than far away ones. The discrete Itakura-Saito distance (McAulay (1984), El-Jaroudi and Makhoul (1991)) is used to calculate the spectral distances

$$D_{IS}(P_1,P_2) = \frac{1}{N}\sum_{n=1}^{N}\left[\frac{P_1(\omega_n)}{P_2(\omega_n)} - \log\left(\frac{P_1(\omega_n)}{P_2(\omega_n)}\right) - 1\right] \tag{7}$$

where P_1 and P_2 are the two considered power spectra, ω_n are the frequency points and N is the total number of frequency points. Since this is a non-symmetric distance measure with respect to the arguments, the distances in both directions are calculated and from this the mean value is formed:

$$w_{ij} = \frac{1}{2}\left(D_{IS}(P_i,P_j) + D_{IS}(P_j,P_i)\right) \tag{8}$$

These weights are, as initially mentioned, inverted and normalized. The weighted power spectra are used to calculate the weighted median spectrum. This will shift the median spectrum in the direction of the closest samples, showing higher estimated power compared to the unbiased median, see Figure 5.

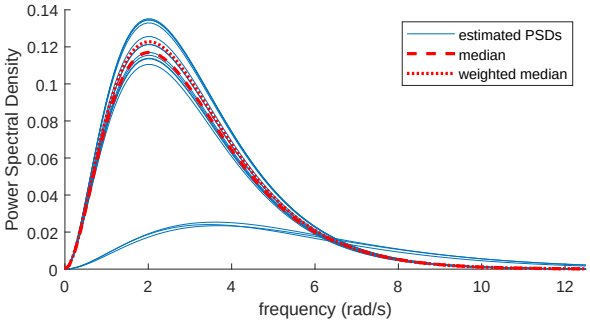

Figure 5: Weighted Median calculated with the Itakura-Saito distance

4. IDENTIFICATION OF MULTIPLE POWER SPECTRA

For power spectrum ensembles that exhibit high spectral variance, two or more excitation spectrum models may be produced in an effort to better represent the spectral range of the process. First, the ensemble of power spectra must be classified

into spectral groups to generate the different spectral models. For the classification, the correlation between the individual power spectra is calculated and the Pearson correlation coefficient is used

$$\rho(P_1, P_2) = \frac{1}{N-1} \sum_{i=0}^{N-1} \left(\frac{P_{1_i} - \mu_{P_1}}{\sigma_{P_1}} \right) \left(\frac{P_{2_i} - \mu_{P_2}}{\sigma_{P_2}} \right) \tag{9}$$

where P_1 and P_2 are the two considered power spectra, μ_{P_1} and σ_{P_1} are mean and standard deviation of P_1, respectively, and μ_{P_2} and σ_{P_2} are mean and standard deviation of P_2 and N is the total number of frequency points.
The correlation of the power spectra is classified into the spectral groups using the k-means algorithm, for which the number of groups must be defined manually beforehand. An example is shown in figure 6, where the ensemble of power spectra is classified into three groups. From these spectral

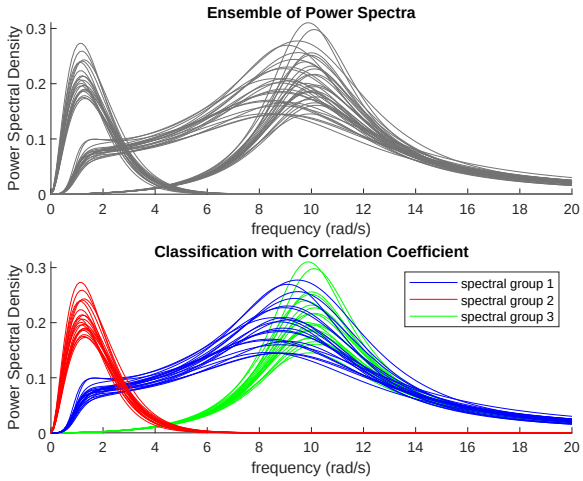

Figure 6: Identification of spectral groups

groups the aforementioned methods (see section 3) can be utilised to calculate an average power spectrum like the mean, median or weighted median for the application to the system. As an example, figure 7 shows the calculated mean power spectra of the ensemble after the classification into groups. To demonstrate, that these mean power spectra represent the ensemble much better than a single mean power spectrum for the whole ensemble, this one is also depicted.

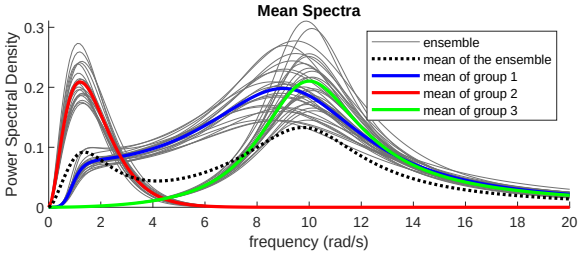

Figure 7: Mean power spectra after classification

5. CONCLUSION

The calculation of the median or the weighted median spectrum, respectively, provide further averaging for the utilisation of ensembles of real data records. The choice of whether to use the mean or one of the two proposed median spectra however will be case dependent and should be combined with any other knowledge and reasoning related to the problem. Any information about the system may lead to a better choice of the average spectrum. In absence of additional information it is useful to calculate all of the average power spectra and choose the worst case, i.e. the one with the higher power especially in the range of the natural frequencies of the system. The natural frequencies are of course dependent on the application. For more complex applications it might be difficult to identify the worst case power spectrum from the model without running simulations. Therefore it might be necessary to suggest that a simulation for all models is carried out then to design according to the worst output case. This means that all of the models can be considered and argued that in many cases the mean would provide a better model, as well as the median or weighted median, respectively.
For the ensembles which exhibit a high spectral variance, more than one average spectrum can be calculated. The consideration of multiple spectral models will yield in a more accurate overall response statistics than a single model. Therefore, for each spectral model derived from the ensemble, different simulations must be performed, from which the overall response can be calculated. As this can lead to a significant increase in computation time, especially for large and complex model analyses, each time should be carefully considered whether

two or more power spectra represent the response statistics more accurate than a single power spectrum.

6. ACKNOWLEDGMENT

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'.

7. REFERENCES

Chen, J., Sun, W., Li, J., and Xu, J. (2013). "Stochastic harmonic function representation of stochastic processes." *Journal of Applied Mechanics*, 80(1), 011001.

Comerford, L., Jensen, H., Mayorga, F., Beer, M., and Kougioumtzoglou, I. (2017). "Compressive sensing with an adaptive wavelet basis for structural system response and reliability analysis under missing data." *Computers & Structures*, 182, 26–40.

Comerford, L., Kougioumtzoglou, I. A., and Beer, M. (2015). "An artificial neural network approach for stochastic process power spectrum estimation subject to missing data." *Structural Safety*, 52, 150–160.

Comerford, L., Kougioumtzoglou, I. A., and Beer, M. (2016). "Compressive sensing based stochastic process power spectrum estimation subject to missing data." *Probabilistic Engineering Mechanics*, 44, 66–76.

El-Jaroudi, A. and Makhoul, J. (1991). "Discrete all-pole modeling." *IEEE Transactions on signal processing*, 39(2), 411–423.

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

Kanai, K. (1957). "Semi-empirical formula for the seismic characteristics of the ground." *Bulletin of the earthquake research institute*, 35, 309–325.

Li, J. and Chen, J. (2009). *Stochastic dynamics of structures*. John Wiley & Sons.

Lin, Y.-K. and Cai, G.-Q. (1995). *Probabilistic structural dynamics: advanced theory and applications*. McGraw-Hill New York.

McAulay, R. (1984). "Maximum likelihood spectral estimation and its application to narrow-band speech coding." *IEEE transactions on acoustics, speech, and signal processing*, 32(2), 243–251.

Newland, D. E. (2012). *An introduction to random vibrations, spectral & wavelet analysis*. Courier Corporation.

Powell, A. and Crandall, S. (1958). *Random Vibration*. The Technology Press of the Massachusetts Institute of Technology, Cambridge.

Shinozuka, M. and Deodatis, G. (1988). "Response variability of stochastic finite element systems." *Journal of Engineering Mechanics*, 114(3), 499–519.

Shinozuka, M. and Deodatis, G. (1991). "Simulation of stochastic processes by spectral representation." *Applied Mechanics Reviews*, 44(4), 191–204.