*Proceedings of the 32nd European Safety and Reliability Conference (ESREL 2022)*
*Edited by* Maria Chiara Leva, Edoardo Patelli, Luca Podofillini, and Simon Wilson
 *Published by* Research Publishing, Singapore.
doi: 10.3850/978-981-18-5183-4_S02-04-389-cd

# Uncertainty quantification over spectral density estimation for strong motion process with missing data

Yu Chen
*Institute for Risk and Uncertainty, University of Liverpool, UK. E-mail: yu.chen2@liverpool.ac.uk*

Edoardo Patelli
*Department of Civil and Environmental Engineering, University of Strathclyde, UK. E-mail: edoardo.patelli@strath.ac.uk*

Ben Edwards
*School of Environmental Sciences, University of Liverpool, UK. E-mail: edwardsb@liverpool.ac.uk*

Michael Beer
*Institute for Risk and Reliability, Leibniz Univ. of Hannover, Germany. E-mail: beer@irz.uni-hannover.de*

Jaleena Sunny
*School of Environmental Sciences, University of Liverpool, UK. E-mail: jaleena.sunny@liverpool.ac.uk*

In this paper, the challenge of quantifying the uncertainty in the estimation of power spectral density (stationary and nonstationary) of ground motion processes subject to missing data is addressed. Specifically, to exploit additional information besides the incomplete recording, simulated ground motions are generated by a stochastic finite-fault model, with its region-specific parameters (source, attenuation, and site parameters) modeled as probability distributions. Then a Bayesian neural network is constructed to probabilistically learn the temporal patterns from such uncertain time-series data. Epistemic uncertainties on the model parameters of the Bayesian neural network model are learnt via variational inference. Thanks to the probabilistic merit of the Bayesian neural network, an ensemble of reconstructed realizations can be obtained, which leads to a probabilistic power spectrum, with each frequency component represented by a probability distribution. This framework is of great importance to researches such as stochastic structural dynamics, where accurate stochastic representations are needed for characterizing engineering excitation processes but faced with incomplete ground motion recordings.

*Keywords*: stochastic processes, evolutionary spectral density, ground motion, variational inference, uncertainty quantification

## 1. Introduction

The analysis of structures under random dynamic excitations, such as ground motions, requires realistic characterization of the stochastic excitations. Power spectral density, especially evolutionary power spectral density (EPS), provides a useful representation of the ground motion processes (Liu, 1970; Shinozuka and Deodatis, 1988; Shields and Deodatis, 2013). However, spectral estimation becomes challenging when only limited and partial recordings are available. Missing data exist in both historical and modern earthquake time histories due to intermittent instrumentation failure (for instance old mechanical instruments in historical strong motions or cheap unreliable temporary instruments that leads to clipping around the peak motion) (Maranò et al., 2017; Zhang et al., 2016).

A number of signal reconstruction methods that fill in the missing values in the time domain are proposed (e.g. see Kondrashov and Ghil (2006); Kondrashov et al. (2014); Comerford et al. (2015a)). An obvious advantage is that classical spectral methods working on equidistant data, such as periodogram, can still be employed. However, due to the convolutional nature of Fourier transform, inaccuracies of the imperfect recon-

struction will be propagated to spectral estimates (Tobar, 2018). Most of current methods fail to account for the uncertainties related to the missing data (Comerford et al., 2015b; Zhang et al., 2017).

To address this challenge, a Bayesian neural network (BNN) based method is developed herein to compute the uncertainty regarding the estimation of power spectral density (stationary and nonstationary) of ground motion processes subject to missing data. Firstly, to exploit additional information besides the incomplete recording, simulated strong motions are generated by a stochastic finite-fault model, with its region-specific parameters (source, attenuation, and site parameters) modeled as probability distributions. Then a Bayesian neural network is constructed to probabilistically learn the temporal patterns from such uncertain time-series data. More specifically, epistemic uncertainties on the model parameters of the Bayesian neural network model are learnt via variational inference. Thanks to the probabilistic merit of the Bayesian neural network, an ensemble of reconstructed realizations can be obtained, which leads to a probabilistic power spectrum, with each frequency component represented by a probability distribution.

## 2. Stochastic finite-fault model

Building on a stochastic model of spectral amplitudes of ground motions, finite fault modeling is useful for simulating ground motions for a large earthquake by simulation of many small earthquake as subfaults that constitute an extended fault plane (Boore, 2003; Atkinson and Boore, 2006). At the core it is a seismological model (see Eq. (1)) of ground motion's amplitude spectrum with source, path and site parameters, which encapsulate the physics of the earthquake process and wave propagation.

$$A(f) = \frac{CM_0}{1+(f/f_0)^2} Z(R) \exp[-\pi f R/Q(f)\beta] \exp\left(-\pi f \kappa_0\right) \tag{1}$$

where $M_0$ is the seismic moment and $f_0$ is the corner frequency, whose dependence on $M_0$ is given by $f_0 = 4.9 \times 10^6 \beta(\Delta\sigma/M_0)^{1/3}$. $\Delta\sigma$ is referred to as stress drop, $\beta$ represents the shear wave velocity in the vicinity of the source. The constant $C$ is given by: $C = R_{\Theta\Phi} V F/(4\pi\rho_s\beta^3 R_0)$, where $R_{\Theta\Phi}$ is the radiation pattern; $V$ represents the partition of total shear-wave energy into horizontal components; $F$ accounts for the free-surface effect; $R_0$ is the a reference distance and $\rho$ is the density in the vicinity of the source. $Z(R)$ is the geometrical spreading function defined by a piecewise series of straight lines. The quality factor $Q(f)$ is an inverse measure of anelastic attenuation. The term $\exp(-\pi f\kappa_0)$ represents a high-cut filter that accounts for the attenuation of the high-frequency motions.

## 3. A Bayesian approach for spectral density estimation with missing data

### 3.1. *Autoregressive modeling scheme*

It's established that finite time series can be well approximated by autoregressive AR($p$) models. An artificial neural network model could be considered as a dynamic autoregressive model that predicts the next value $y_t$ with a window of past lagged values ($[y_{t-1}, \ldots, y_{t-d}]$), as given by

$$y_t = f(\boldsymbol{x}_{t-1}; \mathbf{w}), \text{with } \boldsymbol{x}_{t-1} = [y_{t-1}, \ldots, y_{t-d}] \tag{2}$$

In the context of missing data, such autoregressive strategy facilitates the imputation of the missing sample at each instant, given the past window, so that a real-time reconstruction can be achieved. Particularly, as opposed a linear combination of fixed coefficients in a classic AR($p$) model, a neural network model instead is known for the ability to learn complex nonlinear feature interactions in a time series.

Implicit in such strategy is the uncertainties in learning the underlying generating process, and also in doing the iterative imputation. Generally, limited amount of data has restricted machine learning models from effectively learning the true underlying data generating process. Significant uncertainties exist on the model configurations that may have explained the limited data. Consequently, such uncertainties further compromise

the generalization power of learned models as predictions from uncertain/unrepresentative models can still be unreliable and over confident. Therefore, a Bayesian neural networks (BNN) is constructed to account for the model uncertainties, especially in a context of limited data.

### 3.2. *Bayesian neural networks and Variational Inference*

A Bayesian neural network is equivalent to an ensemble of an infinite number of neural networks. A predictive distribution can be made for each possible configuration of the weights, weighted according to the posterior distribution, to make a prediction about the missing value, as shown below in Eq. 3:

$$p(y_t|\boldsymbol{x}_{t-1},\mathcal{D}) = \int p(\mathbf{w}|\mathcal{D})p(y_t|\boldsymbol{x}_{t-1},\mathbf{w})d\mathbf{w} = \mathbb{E}_{p(\mathbf{w}|\mathcal{D})}[p(y_t|\boldsymbol{x}_{t-1},\mathbf{w})] \quad (3)$$

where $y_t$ and $\boldsymbol{x}_{t-1}$ represents the prediction and the lagged window pair in the autoregressive scheme (Eq. 2); $\mathbf{w}$ are the weights and biases of the neural network model and $\mathcal{D}$ represents the training data. As exact Bayesian inference to the posterior $p(\mathbf{w}|\mathcal{D})$ is intractable and MCMC based methods are bounded by the huge dimensions of the neural network, alternatively, variational inference turned to approximate the true posterior by finding a variational distribution on the weights $q(\mathbf{w}|\theta)$, parameterized by $\theta$, that minimizes the Kullback-Leibler (KL) divergence between $q(\mathbf{w}|\theta)$ and the true posterior $p(\mathbf{w}|\mathcal{D})$:

$$\begin{aligned}\mathcal{F}(\mathcal{D},\theta) &= \mathrm{KL}[q(\mathbf{w}|\theta) \parallel p(\mathbf{w}|\mathcal{D})] \quad (4)\\ &= \int q(\mathbf{w}|\theta)\log\frac{q(\mathbf{w}|\theta)}{p(\mathbf{w})p(\mathcal{D}|\mathbf{w})}d\mathbf{w}\\ &= \mathrm{KL}[q(\mathbf{w}|\theta) \parallel p(\mathbf{w})] - \mathbb{E}_{q(\mathbf{w}|\theta)}\log p(\mathcal{D}|\mathbf{w})\end{aligned}$$

In minimizing the loss function (Eq. 4), a naive attempt to directly take derivatives with respect to (w.r.t) $\theta$ involves an integral over $\mathbf{w}$, which is computationally intractable. A strategy of using Monte Carlo sampling to evaluate expectations, is implemented for further approximation (Blundell et al., 2015).

$$\mathcal{F}(\mathcal{D},\theta) \approx \sum_{i=1}^{n}\log q(\mathbf{w}^{(i)})|\theta) - \log p(\mathbf{w}^{(i)}) - \log p(\mathcal{D}|\mathbf{w}^{(i)}) \quad (5)$$

Further, assume a diagonal Gaussian distribution as the proposed variational posterior $q(\mathbf{w}|\theta)$, parameterized by $\theta = (\mu, \sigma)$, where $\mu$ and $\sigma$ are vector of mean and standard deviation of the probability distribution of weights. This will therefore double the parameters of a neural network model. Subsequently, a sample of posterior weights can be obtained to deal with the back propagation, from parameter-free noises via a transformation: $\mathbf{w} = \mu + \sigma \odot \epsilon$, known by its name as reparameterization trick (Kingma and Welling, 2013), where $\epsilon \in \mathcal{N}(0, I)$ and $\odot$ represents pointwise multiplication. After such transformation, classical gradient-based optimization algorithms (eg. stochastic gradient descent) can still be used for updating $\mu$ and $\sigma$, similarly as updating weights in the classical way during training.

## 4. Numerical experiments

In this section one real accelerogram recording from the ESM (Engineering Strong Motion) database is used to demonstrate the performance of the proposed approach. Note that one can generally have only one observed seismic recording as a realization of a stochastic process, therefore the spectral estimations from such full recording would then serve as the reference for comparison.

In this paper, given a ground motion time-history record, power spectral density (PSD) estimates are derived by Welch method (stationary case) (Welch, 1967), and the evolutionary power spectrum (EPS) are estimated from short time Fourier transform (Liang et al., 2007).

Based on the meta data information from such recording, source, attenuation, and site parameters to Eq. (1) are known and then simulations can be generated. Variabilities of such parameters are considered in Table 1, while other deterministic parameters are given in Table 2.

Missing data are created at random locations, drawn from a uniform distribution of the time index in the recording, given by: (Comerford et al.,

Table 1. Statistical parameters of the stochastic finite fault model

| Parameter | Mean | Standard deviation | Distribution type |
|---|---|---|---|
| Log stress | 1.70 | 0.31 | Gaussian |
| Kappa | $15^a$ | $70^b$ | Uniform |
| Depth | 39 | 155 | Gaussian |

*Source*: Region specific parameters are referenced from (Bindi et al., 2011); $^{a,b}$ represent the minimum and maximum bound in the case of a uniform distribution

Table 2. Source and Path parameters of the Stochastic finite fault model

| Parameter | Description | Value |
|---|---|---|
| $\rho_s$ | density of the medium | 2.7 |
| $\beta$ | shear wave velocity | 3.2 |
| $V$ | horizontal partition | $1/\sqrt{2}$ |
| $R_{\Theta\Phi}$ | radiation pattern | 0.55 |
| $F$ | free-surface factor | 2 |
| $R_0$ | reference distance | 10 |
| $Z(R)$ | geometrical spreading | $b^1 = -1.35$; $b^2 = -0.58$; $b^3 = -1.53$; |
| $Q$ | quality factor | $Q = 250.4f^{0.29}$ |

*Source*: Region specific parameters are referenced from (Bindi et al., 2011)

2015a).

$$f_0(t) = \begin{cases} f(t) \ , \ r_a(t) \geqslant m \\ 0 \ , \ r_a(t) < m \end{cases} \tag{6}$$

where $f_0(t)$ is the recording with missing data, from the original complete recording $f(t)$; $r_a$ is a vector of equally spaced numbers from 0 to 1 arranged in random order and $m$ is the fraction of missing data. An example of incomplete recording with 40% missing data is tested in this analysis, which can be seen in Fig. 1, where the blue bars at the bottom indicates the locations of the missing values. Without explicitly stating otherwise, the results shown in the following sections are based on this example of incomplete recording.

With the BNN trained from 30 simulations, we apply such model to predict the missing values

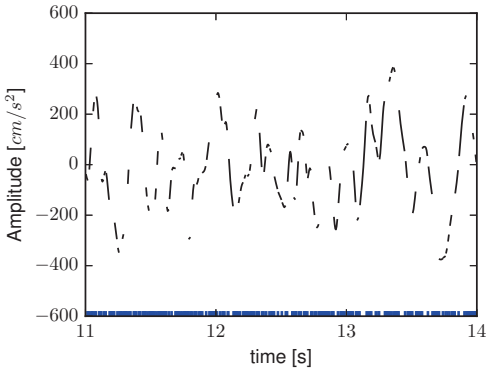


Fig. 1. The recording with 40% missing data at random locations

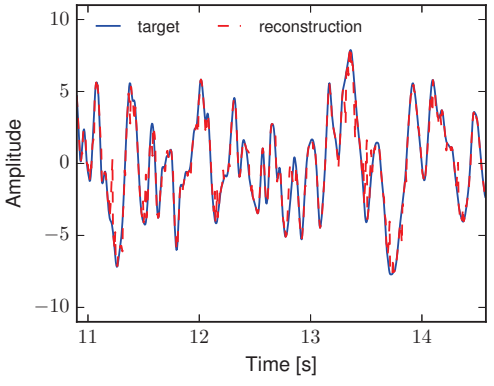


Fig. 2. One reconstructed time-history from the ensemble (Amplitude is normalized)

of the incomplete recording (by Eq.( 3)). An ensemble of 500 reconstructions is made, based on which the uncertainties over the estimated power spectrum can be seen in Fig. 3. In particular, Fig.2 shows one reconstructed time-history from the ensemble, which matches well with the waveform of the original recording. Despite a significant portion of data missing (40%), the ensemble mean PSD agrees well with the target PSD from a complete recording. Generally target PSD values across the whole frequency range are well captured in the 95% credible interval bounds.

Moreover, it can be seen that in lower frequency ranges, the ground truth PSD values mostly lie near the upper bound of the 95% intervals, while at higher frequency ranges (eg. $>$ 12Hz), the

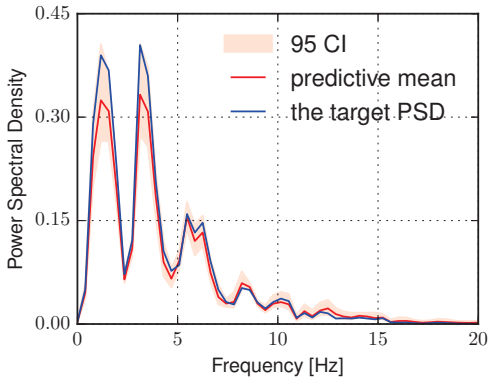


Fig. 3. Power spectral density of the ensemble reconstructions

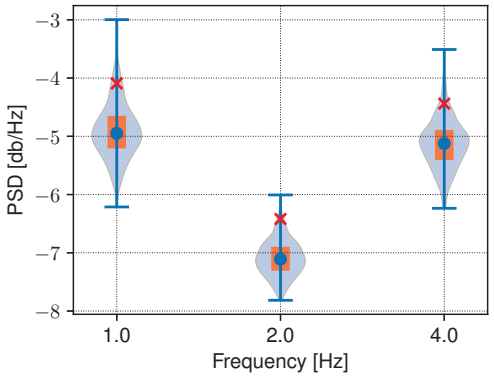


Fig. 4. The distribution of spectral density values with respect to frequencies

ground truth PSD values are closer to the lower bound. It suggested different behaviours when the BNN model learns the temporal patterns.

In more details, Fig. 4 illustratively shows the distribution shape of spectral density values with respect to frequencies, as well as statistics regarding the ensemble PSD estimations are also shown in Fig. 4. The box within represents the regular box plot where quantiles such as 25%, median and 75% are shown. The blue circle represents the median value while the red cross represents the ground truth, i.e., the PSD value from the full recording.

The stationary PSD estimates provide merely the average spectral distribution, without time information. But engineering interests exist in obtaining a time-varying spectral representation due to the "moving resonance" of nonlinear structures. As such, estimates of the mean EPS of the ensemble are shown in Fig. 5. More importantly, the distributions of spectral density values, $S(f,t)$, at various time instants and frequency bins are displayed in Fig. 6, where 4 representative combinations of time instants and frequency bins are selected to show the uncertainty of spectral estimates.

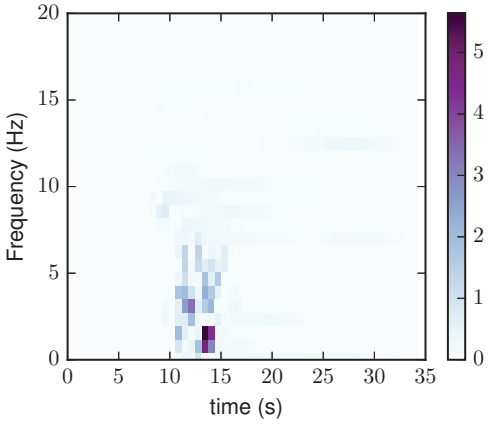


Fig. 5. Mean EPS of an ensemble of reconstructions

Response spectra (pseudo-acceleration, 5% damped) of the 500 reconstructions are shown in gray (see Fig. 7), which closely match the reference response spectra from the complete recording shown in red. As a comparison, Fig. 8 shows the response spectra in the case of only 10% data are randomly removed. It suggests that such reconstructions can be potentially used for a scenario-based structural analyses (eg. time history analysis) when the otherwise complete recording has desired magnitude or distance. This is of great engineering importance due to the scarcity of recorded motions for many earthquake scenarios and site locations. A reliable and response-spectrum compatible reconstruction can enrich the database of strong motions.

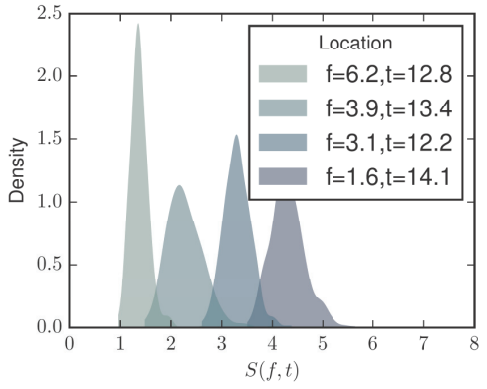


Fig. 6. Uncertainty over estimates of evolutionary power spectra

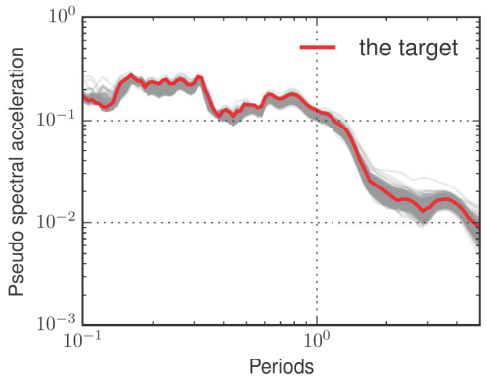


Fig. 7. Pseudo spectral acceleration (5% damped, missing portion 40%)

## 5. Conclusion

A new method has been proposed in this study to estimate the power spectrum (stationary and nonstationary) of the ground motion process based on a real incomplete recording. In particular, this method takes advantage of additional information besides the incomplete recording, for instance the meta data information regarding the earthquake and site characteristics. Such prior knowledge are incorporated in the probabilistic distributions of region-specific parameters (source, attenuation, and site parameters) of a stochastic finite fault model, and its resulting simulations. A Bayesian neural network model is constructed to probabilistically learn the temporal patterns from such aleatoric simulations. Epistemic uncertainties regarding the the temporal patterns are therefore reflected in the posterior distributions of the model parameters of the BNN, learnt by variational inference.

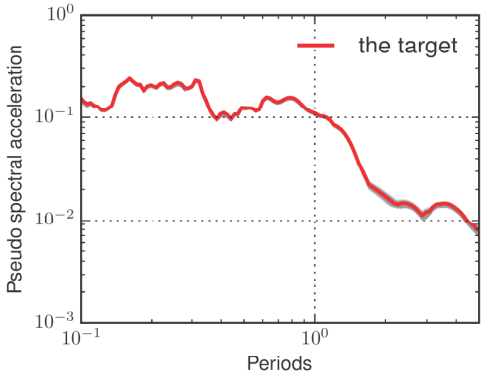


Fig. 8. Pseudo spectral acceleration (5% damped, missing portion 10%)

Results show that even with 40% data missing, the proposed method can provide imputed waveforms as well as spectral estimations that agree well with those of the complete recording. More importantly, uncertainties brought by the missing data have been accounted for in the PSD/EPS/response-spectrum estimates of the ground motion process, based on the ensemble reconstructions. Narrow bounds of credible intervals are seen on such spectral estimations. These response-spectrum compatible reconstructions can be of great engineering importance to enrich the strong motion database.

A thorough validation of the proposed method regarding various missing data types and ground motion scenarios will come shortly.

### Acknowledgement

This work was supported by the EU Horizon 2020 - MSCA Actions project URBASIS [Project no. 813137];

## References

Atkinson, G. M. and D. M. Boore (2006). Earthquake ground-motion prediction equations for eastern north america. *Bulletin of the seismological society of America 96*(6), 2181–2205.

Bindi, D., F. Pacor, L. Luzi, R. Puglia, M. Massa, G. Ameri, and R. Paolucci (2011). Ground motion prediction equations derived from the italian strong motion database. *Bulletin of Earthquake Engineering 9*(6), 1899–1920.

Blundell, C., J. Cornebise, K. Kavukcuoglu, and D. Wierstra (2015). Weight uncertainty in neural network. In *International conference on machine learning*, pp. 1613–1622. PMLR.

Boore, D. M. (2003). Simulation of ground motion using the stochastic method. *Pure and applied geophysics 160*(3), 635–676.

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

Comerford, L., I. A. Kougioumtzoglou, and M. Beer (2015b). On quantifying the uncertainty of stochastic process power spectrum estimates subject to missing data. *International Journal of Sustainable Materials and Structural Systems 2*(1-2), 185–206.

Kingma, D. P. and M. Welling (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Kondrashov, D., R. Denton, Y. Shprits, and H. Singer (2014). Reconstruction of gaps in the past history of solar wind parameters. *Geophysical Research Letters 41*(8), 2702–2707.

Kondrashov, D. and M. Ghil (2006). Spatio-temporal filling of missing points in geophysical data sets. *Nonlinear Processes in Geophysics 13*(2), 151–159.

Liang, J., S. R. Chaudhuri, and M. Shinozuka (2007). Simulation of nonstationary stochastic processes by spectral representation. *Journal of Engineering Mechanics 133*(6), 616–627.

Liu, S. (1970). Synthesis of stochastic representations of ground motions. *The Bell System Technical Journal 49*(4), 521–541.

Maranò, S., B. Edwards, G. Ferrari, and D. Fäh (2017). Fitting earthquake spectra: colored noise and incomplete data. *Bulletin of the Seismological Society of America 107*(1), 276–291.

Shields, M. and G. Deodatis (2013). Estimation of evolutionary spectra for simulation of non-stationary and non-gaussian stochastic processes. *Computers & Structures 126*, 149–163.

Shinozuka, M. and G. Deodatis (1988). Stochastic process models for earthquake ground motion. *Probabilistic engineering mechanics 3*(3), 114–123.

Tobar, F. (2018). Bayesian nonparametric spectral estimation. *Advances in Neural Information Processing Systems 31*.

Welch, P. (1967). 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*(2), 70–73.

Zhang, J., J. Hao, X. Zhao, S. Wang, L. Zhao, W. Wang, and Z. Yao (2016). Restoration of clipped seismic waveforms using projection onto convex sets method. *Scientific reports 6*(1), 1–10.

Zhang, Y., L. Comerford, I. A. Kougioumtzoglou, E. Patelli, and M. Beer (2017). 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*(4), 04017020.