UNCECOMP 2023
5th ECCOMAS Thematic Conference on
Uncertainty Quantification in Computational Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou (eds.)
Athens, Greece, 12—14 June 2023

SPECTRAL DENSITY ESTIMATION OF STOCHASTIC PROCESSES UNDER MISSING DATA AND UNCERTAINTY QUANTIFICATION WITH BAYESIAN DEEP LEARNING

Yu Chen[1], Edoardo Patelli[2], Benjamin Edwards[1], Michael Beer[3]

[1]Institute for Risk and Uncertainty, University of Liverpool
Brownlow Hill, Liverpool, UK
e-mail: {yuchen2, edwardsb}@liverpool.ac.uk

[2] Department of Civil and Environmental Engineering, University of Strathclyde
16 Richmond St, Glasgow G1 1XQ
e-mail: edoardo.patelli@strath.ac.uk

[3] Institute for Risk and Reliability, Leibniz Univ. of Hannover
Welfengarten 1, 30167 Hannover, Germany
e-mail: beer@irz.uni-hannover.de

Keywords: Missing data, Stochastic Variational inference, Uncertainty Quantification, Evolutionary power spectral density, LSTM.

Abstract. *Stochastic processes are widely adopted in many domains to deal with problems which are stochastic in nature and involve strong nonlinearity, nonstationarity and uncertain system parameters. However, the uncertainties of spectral representation of the underlying stochastic processes have not been adequately acknowledged due to the data problems in practice, for instance, missing data. Therefore, this paper proposes a novel method for uncertainty quantification of spectral representation in the presence of missing data using Bayesian deep learning models. A range of missing levels are tested. An example in stochastic dynamics is employed for illustration.*

This is a peer-reviewed, accepted author manuscript of the following article: Chen, Y., Patelli, E., Edwards, B., & Beer, M. (2023). Spectral density estimation of stochastic processes under missing data and uncertainty quantification with Bayesian deep learning. 1-10. 5th International Conference on Uncertainty Quantification in Computational Science and Engineering, Athens, Greece.

1 Introduction

Spectral density estimation involves describing how the power of a signal (or time series) is distributed over frequency, often enabling inferences and model development towards the underlying physical processes. However, real world observations are often imperfect in the sense that they are limited, noisy, unevenly-spaced or missing [1, 12, 5]. In this sense, one of the main challenges in spectral density estimation is to investigate the propagation of uncertainty, from imperfect observations to the subsequent spectral representation. In particular, nonuniformly sampled data (commonly referred to as missing data problem) is a frequent issue in practice which concerns various fields, see [1] for a review.

Standard spectral analysis methods (e.g. Fourier transform based) typically require data to be equidistantly spaced. But in practice it is often difficult for incomplete measurements to be reevaluated, especially for geophysical or environmental data series. As such, it is almost impossible to reconstruct the missing samples not measured with certainty. Nevertheless, with assumptions and prior knowledge of the underlying process, there are many efforts dedicated to the spectral analysis in the presence of missing data.

Parametric models give rise to a parametric formulation of the spectrum by imposing certain structures on the underlying stochastic process. However, these assumptions can be inefficient and sometimes unwarranted, for example the ignorance of nonstationarity for many spectral estimators based on the stationary assumption [12]. Semi-parametric methods, sometimes known as sparse methods, are also proposed. In [7] a compressive sensing approach with additional assumption of the sparsity in frequency domain, have been proposed for spectral density estimation where multiple records compatible with a stochastic process are available. However, in many practical applications we only have access to a single realization, and sometimes that realization is corrupted by missing data. Alternatively, a variety of methods explicitly or implicitly transform the spectral analysis with missing data into iterative imputation of missing samples [18], followed by established full-data spectral analysis (either non-parametric estimators or parametric models). Notably, neural network models have shown prospects in learning the temporal dependency in data of sequential nature, such as time series [15].

However, despite these recent progresses, it should still be noted that a deterministic approach with certain assumptions is inadequate to account for the inherent uncertainty associated with the missing data. Besides, another challenge is to recover the nonstationary characteristics of time series corrupted by missing gaps, which adds additional complexity in reflecting the evolution of spectral density for those missing periods.

This study addresses the spectral uncertainty quantification of stochastic processes under missing data using Bayesian LSTM (long short memory model).

2 Spectral uncertainty quantification

2.1 Spectral representation of stochastic processes

In this section, a brief review of the theory of the spectral representation of stochastic processes (stationary and non-stationary) is outlined, providing a basis for the proposed framework. In particular, focus is on power spectral estimation and simulation of the corresponding processes. A general non-stationary random process, with respect to a family of oscillatory functions, can be represented in the form [14]:

$$X_t = \int_{-\infty}^{\infty} A(\omega, t) e^{i\omega t} \mathrm{d}Z(\omega) \tag{1}$$

where $\phi_t(\omega) = A(\omega, t)e^{i\omega t}$ represent the oscillatory functions, of which $A(\omega, t)$ suggests a slowly varying and frequency-dependent modulating function and $Z(\omega)$ is an orthogonal process; $\{X_t\}$ is termed as oscillatory processes whose (two-sided) evolutionary power spectral density is further given as:

$$S(\omega, t) = |A(\omega, t)|^2 S(\omega) \tag{2}$$

where $S(\omega)$ represents the power spectral density function in the case of a stationary process with a family of complex exponentials, i.e., $\phi_t(\omega) = e^{i\omega t}$. The semi-stationary property due to the slowly-changing spectra premise facilitates the practical estimation of the evolutionary spectra given a realization record via non-stationary time-frequency methods, e.g. wavelet transforms [16, 17, 13].

Inversely, a versatile formula for generating sample realizations compatible with the stochastic process is given by spectral representation method (SRM) [13]:

$$x^{(i)}(t) = \sqrt{2} \sum_{n=0}^{N-1} \sqrt{2S(\omega_n, t)\Delta\omega} \cos(\omega_n t + \Phi_n^{(i)}) \tag{3}$$

where $x^{(i)}(t)$ is a sample simulation, $\Phi^{(i)}$ is the set of independent random phase angles, distributed uniformly over the interval $[0, 2\pi]$, for the ith sample realizations; N and $\Delta\omega$ relate to the discretization of the frequency domain.

2.2 Variational Bayesian inference in Recurrent neural networks

In accounting for the epistemic uncertainty, probability distributions are introduced upon the model parameters of the neural network model that is employed to learn the temporal dynamics of the underlying process. A resulting issue is the huge dimensionality of a deep learning model. To efficiently approximate the true posterior distribution of these many model parameters, under the framework Bayesian infernce, stochastic variational inference (see e.g. [10, 11, 2]) involves in optimizing an approximate to the intractable true posterior. It optimizes the parameters of a proposed variational distribution $q(\mathbf{w}|\theta)$ so that the Kullback-Leibler (KL) divergence between the approximate distribution and the true posterior after seeing data $\mathcal{D}$ is minimised: $\theta^* = \arg\min_\theta \mathrm{KL}[q(\mathbf{w}|\theta) \parallel p(\mathbf{w}|\mathcal{D})]$. It thus leads to the minimization of a general stochastic objective function for neural network models in the Bayesian supervised learning setting [3]:

$$\mathcal{J}(\mathcal{D}, \theta) = \mathrm{KL}[q(\mathbf{w}|\theta) \parallel p(\mathbf{w})] - \mathbb{E}_{q(\mathbf{w}|\theta)} \log p(\mathcal{D}|\mathbf{w}) \tag{4}$$

which stands for the negative lower bound of the evidence term $\log p(\mathcal{D})$, i.e. negative ELBO. The formulation of Eq. (4) is interpreted as a tradeoff between the two composing terms: the variational distribution needs to both explain the observed data well, while be close to the prior.

Evaluation of the stochastic objective and further gradients is challenging and several Monte Carlo estimators are adopted as approximate solutions [2]. Additional difficulty comes with the complexity of the architectures of deep learning models (e.g. LSTM in this analysis) than the regular fully-connected networks. With the recurrent network architecture, correspondingly, the negative ELBO in the case of RNN, can be written as [8]:

$$\mathcal{J}_R = \mathbb{E}_{q(\boldsymbol{\omega})} \log p\Big(\mathbf{y}|f_{\mathbf{y}}^{\boldsymbol{\omega}}(f_{\mathbf{h}}^{\boldsymbol{\omega}}(\mathbf{x}_T, f_{\mathbf{h}}^{\boldsymbol{\omega}}(\dots f_{\mathbf{h}}^{\boldsymbol{\omega}}(\mathbf{x}_1, \mathbf{h}_0)\dots)))\Big) + \mathrm{KL}[q(\boldsymbol{\omega}) \parallel p(\boldsymbol{\omega})] \tag{5}$$

where $\boldsymbol{\omega}$ collectively represents all the parameters in a LSTM model. Specifically, a Bernoulli variational distribution for each matrix row $\boldsymbol{\omega}_k$ is proposed on the basis of a mixture of Gaussians with small variance σ^2 [8]:

$$q(\mathbf{w}_k) = p\mathcal{N}(\mathbf{w}_k; \mathbf{0}, \sigma^2\mathbf{I}) + (1-p)\mathcal{N}(\mathbf{w}_k; \phi_k, \sigma^2\mathbf{I}) \tag{6}$$

where the random weight matrix is factorized over the rows as $\boldsymbol{\omega}_k = g(\phi_k, \boldsymbol{\epsilon}) = \phi_k \cdot \text{diag}(\boldsymbol{\epsilon}_k)$; ϕ_k represent the variational parameters; diag means the diagonal matrix operation. Following the idea of Monte Carlo estimator to approximate expectation and reparameterization to remove the dependence of $q(\cdot)$ in the integral (see a Gaussian case in [11] for details), a further approximation of the stochastic objective function [8]:

$$\mathcal{J}_R \approx -\sum_{l=1}^{N} \log\left(\mathbf{y}|f_{\mathbf{y}}^{\boldsymbol{\omega}^{(l)}}(f_{\mathbf{h}}^{\boldsymbol{\omega}^{(l)}}(\mathbf{x}_T, f_{\mathbf{h}}^{\boldsymbol{\omega}^{(l)}}(\dots f_{\mathbf{h}}^{\boldsymbol{\omega}^{(l)}}(\mathbf{x}_1, \mathbf{h}_0)\dots)))\right) + \lambda||\boldsymbol{\phi}||_2^2 \tag{7}$$

$$\boldsymbol{\omega}^{(l)} = g(\boldsymbol{\phi}, \boldsymbol{\epsilon}^{(\boldsymbol{l})}) \text{ with } \boldsymbol{\epsilon}^{(\boldsymbol{l})} \sim p(\boldsymbol{\epsilon}) \tag{8}$$

where $p(\boldsymbol{\epsilon})$ denotes a Bernoulli distribution with parameter p given in advance as hyperparameters; $\lambda||\boldsymbol{\phi}||_2^2$ suggests a further approximation of the second term in Eq. (5) by L2 regularisation with weight decay λ and variational parameters $\boldsymbol{\phi}$ to be solved, see [9] for more details. In minimizing Eq. (7), for efficiency a new realization $\boldsymbol{\omega}^{(l)}$ is sampled for each input $\mathbf{x}_i$ data point. In particular, note that the weight sharing mechanism in RNN requires the same weight realizations being used at each time step, suggesting the same (but random) masking given by the Bernoulli distribution is passed throughout time steps.

Substituting the Bernoulli variational distribution for the true posterior then approximates the predictive distribution for each missing point, as given below:

$$\int p(\tilde{\mathbf{y}}|\tilde{\mathbf{x}}, \boldsymbol{\omega}) q(\boldsymbol{\omega}) \mathrm{d}\boldsymbol{\omega} \approx \frac{1}{T}\sum_{t=1}^{T} p(\tilde{\mathbf{y}}|\tilde{\mathbf{x}}, \boldsymbol{\omega}^{(t)}) \tag{9}$$

where $\tilde{\mathbf{x}}$ represents the missing samples and $\tilde{\mathbf{y}}$ the recurrent imputations. It yields a predictive distribution for each missing time point. Effectively, it amounts to implement T stochastic forward passes $\{\boldsymbol{\omega}^{(t)}\}_{t=1}^{T} \sim q(\boldsymbol{\omega})$, obtained from T realizations of the variational Bernoulli distribution parameterized by the parameter p, through the network model and average the results. Iteratively sampling from the model's predictive distribution at each step, coupled with the accordingly updated hidden states, produces an ensemble of reconstructions.

3 Example application

In this study, we investigate the challenge that, when characterizing earthquake-induced stochastic excitations, the source load record is corrupted with missing gaps [6]. Specifically, instead of conducting the spectral estimation towards the same stochastic process with multiple artificial records (which is generally not the case in practice), only one real seismic record is employed to demonstrate the performance of the proposed method in spectral estimation and uncertainty quantification. A seismic record of magnitude $M = 6.5$, normal faulting, epicentral distance $R = 18.6km$, recorded at a class A site in Italy is adopted from the ESM database.

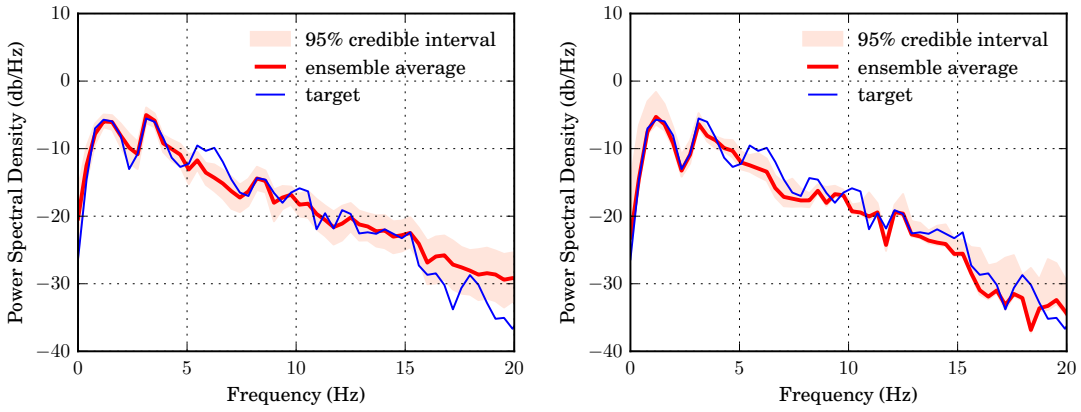
Many existing spectral estimators aim to perform spectral estimation from the available samples. However, the very limited information contained in the available samples, especially when there is a high proportion of incomplete data, imposes a performance ceiling for these

approaches and restricts the potential for robust uses in practice. As such, to build an informed neural network model for the imputation task, we equip the network model with prior knowledge by training the model from physics-based simulations. Given the incomplete record from an environmental process, there exists a great deal of information besides the available samples, for instance, in this case, the meta data (e.g magnitude, epicentral distance, 30m shear wave velocity), local site effects and regional seismicity. In this study, a well-calibrated finite fault stochastic model that encapsulates various factors affecting ground motions (e.g. source, path, and site effects) into a parametric formulation of the Fourier amplitude spectrum, is adopted:

$$A(f;\boldsymbol{\Theta}) = E(f, M; \boldsymbol{\Theta_e}) P(f, R; \boldsymbol{\Theta_p}) G(f; \boldsymbol{\Theta_s}) \tag{10}$$

where M is magnitude, R epicentral distance, the vector $\boldsymbol{\Theta} = (\boldsymbol{\Theta_e}, \boldsymbol{\Theta_p}, \boldsymbol{\Theta_s})$ represent the region specific parameters reflecting the source, path, and site effects from a physics-based perspective (see [4, 19] for a detailed explanation on each component of the parameterized spectrum). These region specific parameters are taken from the seismological analysis of the corresponding region. As a result, 100 simulations are created to initialize the Bayesian LSTM model. For comparison purpose, another model with classic fully-connected layers is also constructed. A range of percentages (ϵ_{MP}) of missing data are considered, in particular, 10% to 70% of the data are removed over 10 gaps at random locations.

Fig. 1 show the power spectral density estimation from an ensemble of 500 reconstructed time histories based on the non-parametric Welch's method. It can be seen that both methods have approximated the target PSD fairly well with the ensemble average and the credible intervals have well contained the target from the otherwise complete recording.



(a) Bayesian LSTM model (b) Bayesian dense neural network model

Figure 1: Welch-based estimates on power spectral density and associated credible intervals under $\epsilon_{MP} = 30\%$

To quantitatively reflect the dissimilarity between PSD estimates under different missing scenarios and Bayesian models, the Wasserstein Fourier distance [5], which computes the Wasserstein distance between (normalized power spectral density), is adopted, given below:

$$W_p(\mu, \nu) = \Big(\int_0^1 |F_\mu^{-1}(q) - F_\nu^{-1}(q)|^p \mathrm{d}q \Big)^{1/p} \tag{11}$$

where F^{-1} denote the inverse cumulative distribution (also known as quantile function of q) of two probability measures of interest μ, ν, as in the normalized power spectral density [5].

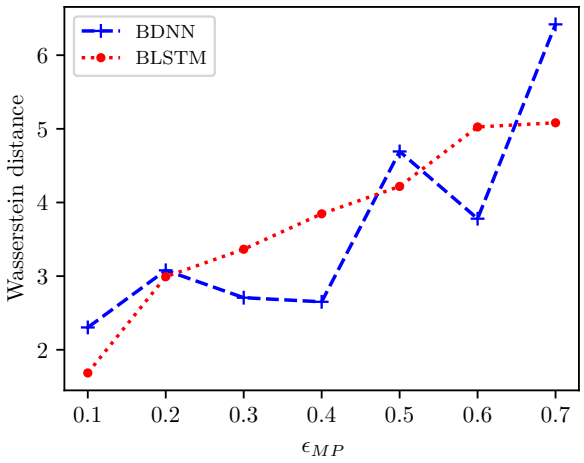


Figure 2: Wasserstein metric of spectral dissimilarity of PSD estimates for the two Bayesian mode

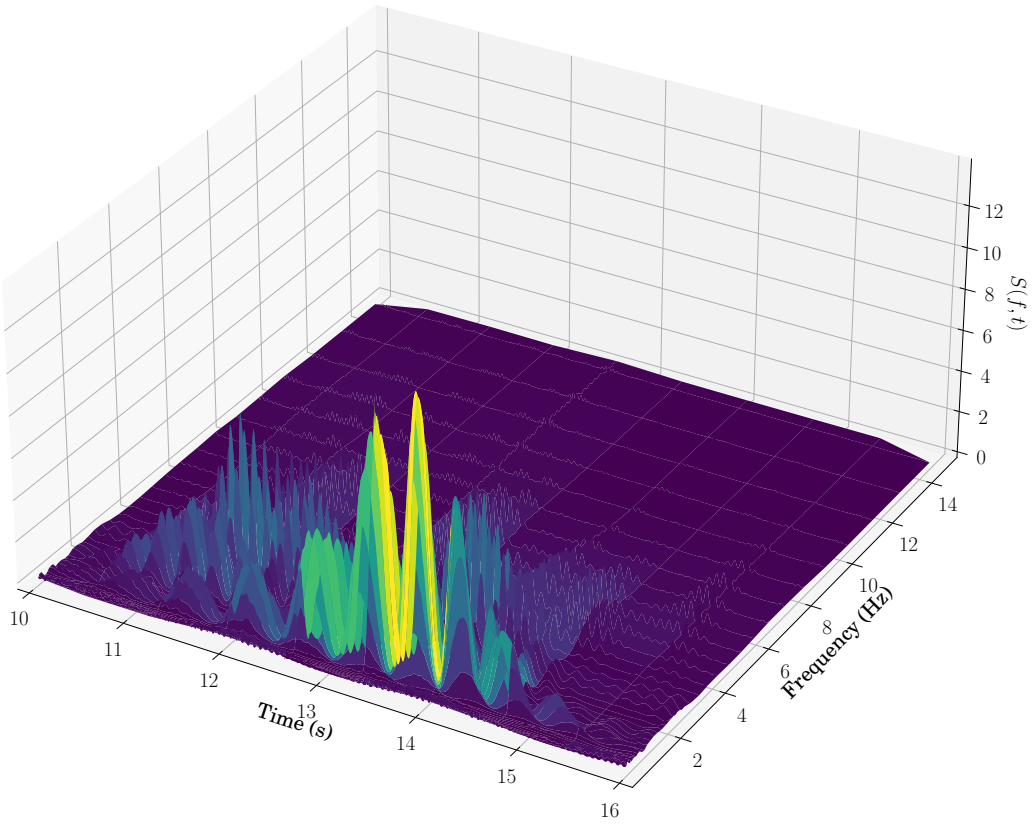


Figure 3: Morlet wavelet based evolutionary power spectra (EPS) without missing data

As highlighted before, many physical processes are in fact nonstationary. A realistic spectral representation that have captured these characteristics (e.g. temporal nonstationarity and

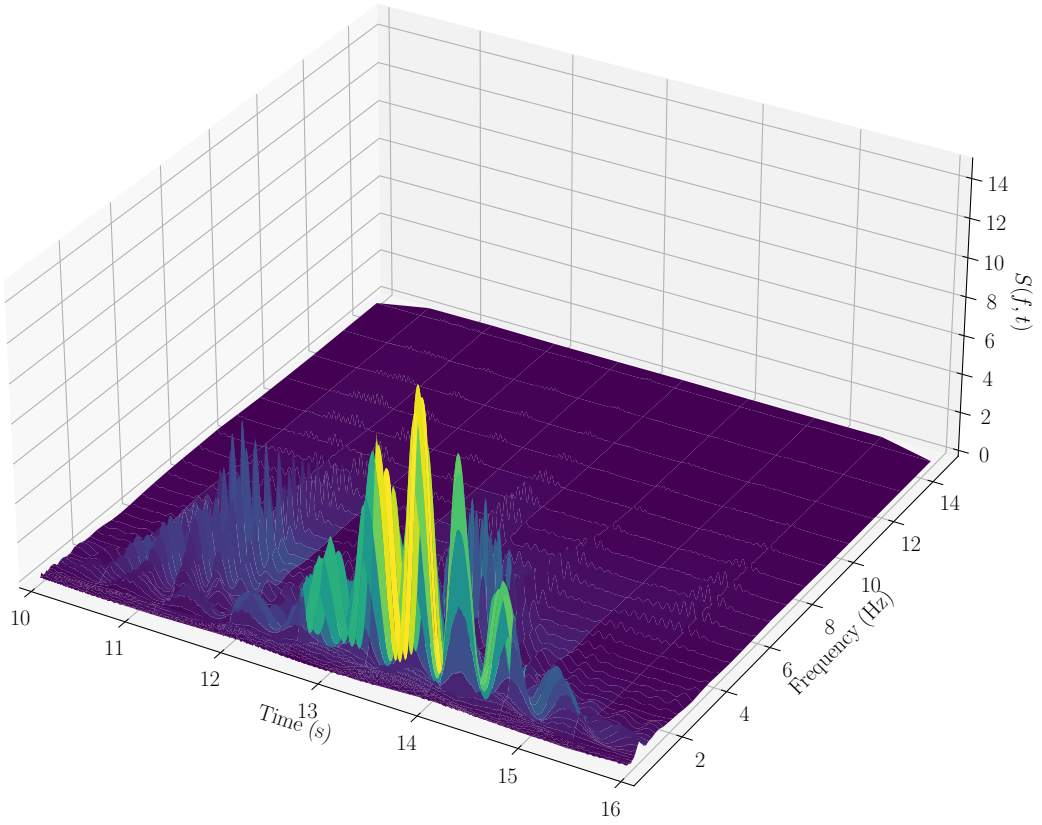


Figure 4: Ensemble-avaeraged Morlet wavelet based EPS estimate by the Bayesian dense neural network model ($\epsilon_{MP} = 40\%$)

spectral nonstationarity) is vital in describing the response behavior of engineering structures subject to these physical processes. A continuous gap will make it difficult to estimate the evolutionary spectral density S_{ft} at those missing time stamps. Specifically, Morlet wavelet is adopted (see [17] for details) to estimate the EPS (evolutioanry power spectrum) from the ensemble reconstructions.

Fig. 5 displays the target EPS without missing data. Fig. 4 shows the ensemble-averaged EPS estimate for Bayesian DNN model in the case of 40% missing data. It should be noted that while the BDNN model is capable of producing reasonable PSD estimates up to 70%, it yields evident discrepancy on the EPS estimates since 50%. These discrepancies can be illustratively seen in Fig. 6b, where a suite of sample realizations compatible with the underlying stochastic process, represented by the estimated EPS, are generated by Eq. 3 using the spectral representation method.

By comparison, Fig. 5 suggests that the Bayesian LSTM model is effective in the spectral density estimation even with 70% missing data, though with a noticeable loss of power. A direct comparison of these two models, as shown in Fig. 6. The quality of EPS estimates are reflected by the subsequent sample realizations. It can be seen that Bayesian LSTM model managed to create effective sample realizations even with 70% missing data. Under the effect such significant amount of incompleteness, a bias can be found regarding spectral periods over 3 seconds. In terms of Bayesain DNN model, significant discrepancies are found with both EPS estimate and associated sample realizations.

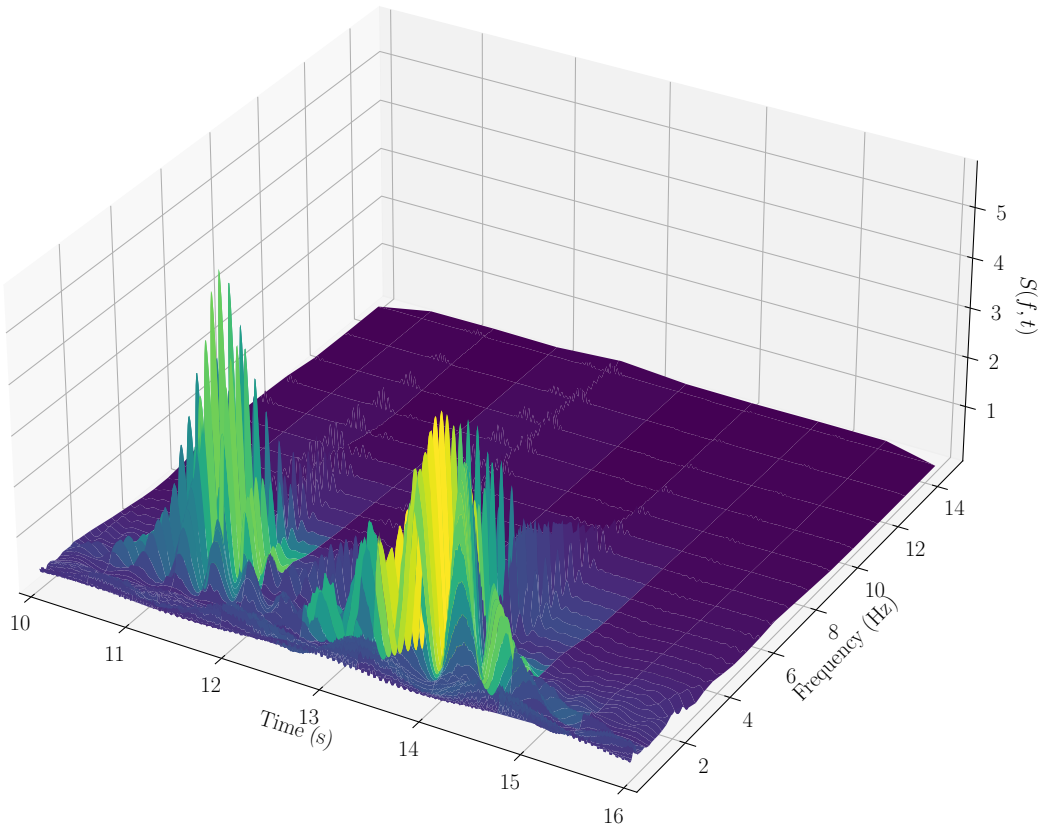
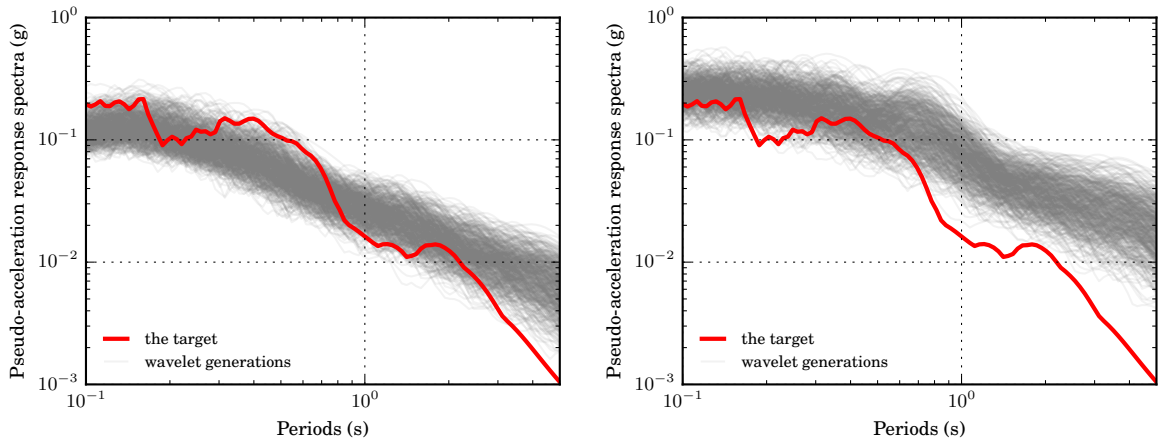


Figure 5: Ensemble-averaged Morlet wavelet based EPS estimate by the Bayesian LSTM model ($\epsilon_{MP} = 70\%$)



(a) Bayesian LSTM model $\epsilon_{MP} = 70\%$ (b) Bayesian DNN model $\epsilon_{MP} = 50\%$

Figure 6: Pseudo-acceleration response spectra of sample realizations generated by Eq. (3) under different missing percentages

4 Conclusion

In this paper, the challenge of quantifying the uncertainty in stochastic process spectral estimates based on an incomplete realization with missing gaps is addressed. In particular, a Bayesian LSTM model is proposed to account for the uncertainty in the time domain reconstructions and subsequently the uncertainty in the subsequent spectral estimates. A range of missing levels are tested. The results suggest that LSTM model outperforms the classic fully connected model in the estimation of evolutionary spectra. Of particular note is that the proposed Bayesian model is effective even with a missing percentage as high as 70%.

5 Code

The Python code for the implementation of the spectral representation method and also the estimation of evolutionary spectrum with wavelet transform is provided by the authors at `https://github.com/leslieDLcy/StoSpecRep`. The code for the implementation of the stochastic finite fault model is provided at `https://github.com/leslieDLcy/stoexsim.git`.

6 Acknowledgement

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

References

[1] Prabhu Babu and Petre Stoica. "Spectral analysis of nonuniformly sampled data–a review". In: *Digital Signal Processing* 20.2 (2010), pp. 359–378.

[2] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. "Variational Inference: A Review for Statisticians". In: *Journal of the American Statistical Association* 112.518 (2017), pp. 859–877. ISSN: 1537274X. DOI: 10.1080/01621459.2017.1285773. arXiv: 1601.00670.

[3] Charles Blundell et al. "Weight uncertainty in neural network". In: *International conference on machine learning*. PMLR. 2015, pp. 1613–1622.

[4] David M Boore. "Simulation of ground motion using the stochastic method". In: *Pure and applied geophysics* 160 (2003), pp. 635–676.

[5] Elsa Cazelles, Arnaud Robert, and Felipe Tobar. "The Wasserstein-Fourier distance for stationary time series". In: *IEEE Transactions on Signal Processing* 69 (2020), pp. 709–721.

[6] L Comerford et al. "Compressive sensing with an adaptive wavelet basis for structural system response and reliability analysis under missing data". In: *Computers & Structures* 182 (2017), pp. 26–40.

[7] Liam Comerford, Ioannis A Kougioumtzoglou, and Michael Beer. "Compressive sensing based stochastic process power spectrum estimation subject to missing data". In: *Probabilistic Engineering Mechanics* 44 (2016), pp. 66–76.

[8] Yarin Gal and Zoubin Ghahramani. "A theoretically grounded application of dropout in recurrent neural networks". In: *Advances in neural information processing systems* 29 (2016).

[9] Yarin Gal and Zoubin Ghahramani. "Dropout as a bayesian approximation: Representing model uncertainty in deep learning". In: *international conference on machine learning*. PMLR. 2016, pp. 1050–1059.

[10] Alex Graves. "Practical variational inference for neural networks". In: *Advances in neural information processing systems* 24 (2011).

[11] Diederik P Kingma and Max Welling. "Auto-encoding variational bayes". In: *arXiv preprint arXiv:1312.6114* (2013).

[12] Marina I Knight, Matthew A Nunes, and Guy P Nason. "Spectral estimation for locally stationary time series with missing observations". In: *Statistics and Computing* 22 (2012), pp. 877–895.

[13] Jianwen Liang, Samit Ray Chaudhuri, and Masanobu Shinozuka. "Simulation of non-stationary stochastic processes by spectral representation". In: *Journal of Engineering Mechanics* 133.6 (2007), pp. 616–627.

[14] MB Priestley. "Power spectral analysis of non-stationary random processes". In: *Journal of Sound and Vibration* 6.1 (1967), pp. 86–97.

[15] David Salinas et al. "DeepAR: Probabilistic forecasting with autoregressive recurrent networks". In: *International Journal of Forecasting* 36.3 (2020), pp. 1181–1191.

[16] PD Spanos and IA Kougioumtzoglou. "Harmonic wavelets based statistical linearization for response evolutionary power spectrum determination". In: *Probabilistic Engineering Mechanics* 27.1 (2012), pp. 57–68.

[17] Pol D Spanos and Giuseppe Failla. "Evolutionary spectra estimation using wavelets". In: *Journal of Engineering Mechanics* 130.8 (2004), pp. 952–960.

[18] Petre Stoica, Erik G Larsson, and Jian Li. "Adaptive filter-bank approach to restoration and spectral analysis of gapped data". In: *The Astronomical Journal* 120.4 (2000), p. 2163.

[19] Christopher Vetter and Alexandros A Taflanidis. "Comparison of alternative stochastic ground motion models for seismic risk characterization". In: *Soil Dynamics and Earthquake Engineering* 58 (2014), pp. 48–65.