Proc. of the 8th Intl. Symp. on Reliability Engineering and Risk Management (ISRERM 2022)
Edited by Michael Beer, Enrico Zio, Kok-Kwang Phoon, and Bilal M. Ayyub
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-5184-1_MS-06-179-cd

Uncertainty Quantification Over Spectral Estimation of Stochastic Processes Subject to Gapped Missing Data Using Variational Bayesian Inference

Yu Chen*[1], Edoardo Patelli[2], Michael Beer[3], and Ben Edwards[1]

[1]*Institute for Risk and Uncertainty, University of Liverpool, Liverpool, UK*
[2]*Department of Civil and Environmental Engineering, University of Strathclyde, Glasgow, UK*
[3]*Institute for Risk and Reliability, Leibniz Univ. of Hannover, Hannover, Germany*

Abstract

In this work we quantify the uncertainty over Power Spectral Density estimation of stochastic processes based on realizations with gapped missing data. For the purpose of imputation, a fully-connected neural network architecture that works in an autoregressive manner is firstly constructed to probabilistically capture the temporal patterns in the time series data. Particularly, under the Bayesian scheme, uncertainties with respect to the parameters of the neural network model (i.e. weights) are introduced by multivariate Gaussian distribution. During training, the posteriors are learnt through variational inference approach. As a result, the missing gaps can be recursively imputed via our neural network in each realization, and thanks to the probabilistic merit of Bayesian inference, an ensemble of reconstructed realizations can then be obtained. Further, by resorting to a Fourier-based spectral estimation method, a probabilistic power spectrum could be derived, with each frequency component represented by a probability distribution.

Keywords: Variational Bayesian inference, Bayesian neural network, missing data, stochastic process, spectral estimation

1. Introduction

Power spectral density (PSD) function plays an important role in the representation of stochastic processes, which empowers statistical analyses of time series data of vast disciplines in science and engineering. However, spectral estimation requires a significant amount of complete data samples to attain a predefined adequate degree of accuracy, and thus can be challenging when only partial observations are available [1]–[3]. Missing data situation is ubiquitous in the recording of many physical processes, either due to the malfunction of measurement equipment such as clipping in seismic recordings [4] or the natural irregularities in many astronomical observations [5].

There are a bunch of signal reconstruction methods that fill in the missing values in the time domain (e.g. see [6]–[8]). An obvious advantage is that classical spectral methods that work on equidistant data, such as periodogram, can still be employed. However, due to the convolutional nature of Fourier transform, inaccuracies of the imperfect reconstruction will be propagated to spectral estimates. Most of current methods fail to account for the uncertainties related to the missing data [1]. Similarly, for parametric models that assume a certain structure of the underlying stochastic processes, for example an autoregressive model (see [9], [10]), the parameter uncertainties due to the incomplete data aren't well captured.

Alternatively, Bayesian spectral estimation methods naturally deal with uncertainty. For example, in [3] Tobar proposed a nonparametric Bayesian generative model for non-uniformly-sampled data. In [11] Christmas proposed a Bayesian spectral estimation method by assuming Student-t distributed noise.

However, most of these methods are dedicated to missing data in a scattered manner, as termed as non-uniformly-sampled data or irregularly-sampled or unevenly-spaced data in literature. Insufficient attention are focused on the missing gap situation, where a large continuous chunk of data are missing. It's been noted in some researches that longer gaps lead to harder imputations [8]. While simple methods such as linear interpolation may still provide reasonable predictions in the unevenly-spaced situation, they may lead to more serious distortion to the original signal when the gap is relatively long [12]. Normally iterative methods are employed in filling in an gap [6], [13]. This leads to a stronger need to account for the uncertainty since iterative imputation may propagate the errors from earlier imputations. In this paper, we proposed a strategy based on Bayesian neural networks to account for the uncertainty during the recursive reconstruction strategy of gapped missing data.

*E-Mail: yuchen2@liverpool.ac.uk

2. Methodologies

2.1. Spectral representation for stochastic processes

Consider a real-values stationary process, $X(t)$, there exists a corresponding complex orthogonal process $Z(\omega)$ such that $X(t)$ can be represented in the form [14]:

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

where $dZ(\omega)$ is an orthogonal process having the following two properties:

$$\mathbb{E}(|\mathrm{d}Z^2(\omega)|) = 4S_X(\omega)\mathrm{d}\omega \tag{2}$$

$$\mathbb{E}(\mathrm{d}Z(\omega)) = 0 \tag{3}$$

In Eq. 3, $S_X(\omega)$ is the two-sided power spectrum of the process $X(t)$. In addition, a versatile formula for generating realizations compatible with the stationary process of Eq.1 is given by [15]:

$$f(t) = 2\sum_{n=0}^{N-1} \sqrt{S_f(\omega)\Delta(\omega)}\cos(\omega_n t + \phi_n) \tag{4}$$

where ϕ_n is the independent random phase angle distributed uniformly over the interval $[0, 2\pi]$; N and $\Delta(\omega)$ relate to the discretization of the frequency domain.

2.2. The converge 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

$$y_t = f(\boldsymbol{x}_{t-1}; \mathbf{w}), \text{with } \boldsymbol{x}_{t-1} = [y_{t-1}, \ldots, y_{t-d}] \quad (5)$$

As opposed to the fixed coefficients in a classic AR(p) model, a neural network model instead learn a time-varying pattern depending on the input lagged window. Moreover, it's known for the ability to learn complex nonlinear feature interactions in a time series.

The 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. Specifically, when there is a gap in the middle of a signal, a converge strategy, where we impute the missing locations present in between, from both directions, can be adopted. In this sense, the reconstruction is made by iterative predictions.

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 in that 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 missing data.

2.3. Bayesian neural network and variational inference

Uncertainties on the weights of a neural network model are modelled by probability distributions and learnt in a Bayesian scheme, given the observed training data. In this case the training data are the two part of observations around the missing gap and thus two BNN models are constructed to implement the forward-backward iterative prediction strategy.

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. 6:

$$\begin{aligned} p(y_t|\boldsymbol{x}_{t-1}, \mathcal{D}) &= \int p(\mathbf{w}|\mathcal{D}) p(y_t|\boldsymbol{x}_{t-1}, \mathbf{w}) d\mathbf{w} \\ &= \mathbb{E}_{p(\mathbf{w}|\mathcal{D})}[p(y_t|\boldsymbol{x}_{t-1}, \mathbf{w})] \end{aligned} \quad (6)$$

where y_t and $\boldsymbol{x}_{t-1}$ represents the prediction and the lagged window pair in the autoregressive scheme (Eq. 5); $\mathbf{w}$ are the weights and biases of the neural network model and $\mathcal{D}$ represents the training data. As exact Bayesian inference

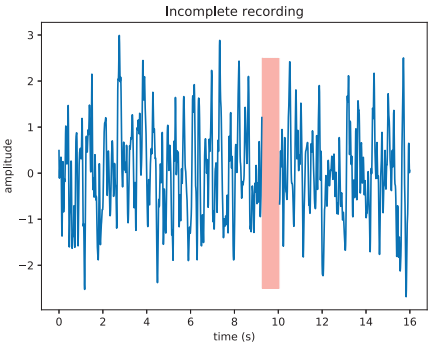

Figure 1. A Kanai Tajimi realization with 10% gap indicated by the red vertical bar.

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, variaitional inference turned to approximate the true posterior by finding a variational distribution on the weights $q(\mathbf{w}|\theta)$, parameterized by θ, that minimizes the Kullback-Leibler (KL) divergence between $q(\mathbf{w}|\theta)$ and the true posterior $p(\mathbf{w}|\mathcal{D})$:

$$\begin{aligned} \mathcal{F}(\mathcal{D}, \theta) &= \mathrm{KL}[q(\mathbf{w}|\theta) \parallel p(\mathbf{w}|\mathcal{D})] \\ &= \int q(\mathbf{w}|\theta) \log \frac{q(\mathbf{w}|\theta)}{p(\mathbf{w}) p(\mathcal{D}|\mathbf{w})} d\mathbf{w} \\ &= \mathrm{KL}[q(\mathbf{w}|\theta) \parallel p(\mathbf{w})] - \mathbb{E}_{q(\mathbf{w}|\theta)} \log p(\mathcal{D}|\mathbf{w}) \end{aligned} \quad (7)$$

In minimizing the loss function (Eq. 7), a naive attempt to directly take derivatives with respect to (w.r.t) θ involves an integral over $\mathbf{w}$, which is computationally intractable. A strategy of using Monte Carlo sampling to evaluate expectations, are implemented for further approximation [16].

$$\mathcal{F}(\mathcal{D}, \theta) \approx \sum_{i=1}^{n} \log q(\mathbf{w}^{(i)})|\theta) - \log p(\mathbf{w}^{(i)}) - \log p(\mathcal{D}|\mathbf{w}^{(i)}) \quad (8)$$

Suppose a diagonal Gaussian distribution as the variational posterior $q(\mathbf{w}|\theta)$, parameterized by $\theta = (\mu, \sigma)$, where μ and σ are vector of mean and standard deviation of the probability distribution of weights, thus doubling the parameters of a neural network model. Furthermore, by obtaining a sample of posterior weights from parameter-free noises via a transformation: $\mathbf{w} = \mu + \sigma \odot \epsilon$, known by its name as reparameterization trick [17], where $\epsilon \in \mathcal{N}(0, I)$ and $\odot$ represents pointwise multiplication. Classical gradients-based optimization progresses can be used for updating μ and σ, similarly as updating weights in the classical way during training.

3. Numerical experiments

In this work, to investigate the uncertainties related to the missing data, two sources of randomness are considered:

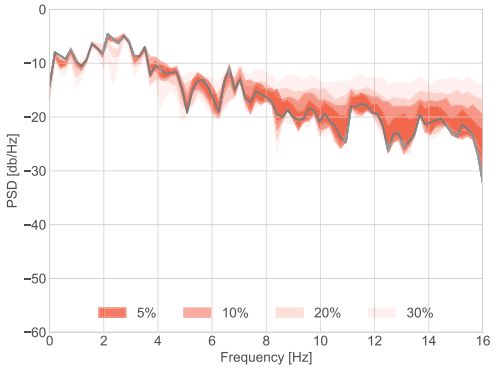

Figure 2. Uncertainty levels of different gap sizes.

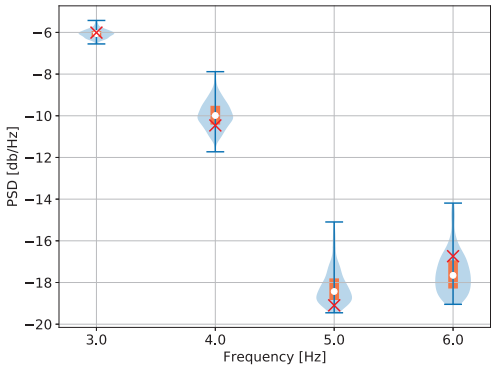

Figure 3. Distribution of the spectral density value on each frequency with 10% missing data.

gap size and gap location. Correspondingly, three scenarios with different missing gap configurations are conducted. A Kanai Tajimi (KT) model, which was widely used in stochastic dynamics of structures to represent a stationary ground motion process, is adopted herein to give the PSD function for the underlying stochastic process:

$$G(\omega) = \frac{1 + 4\xi_g^2(\omega/\omega_g)^2}{1 - (\omega/\omega_g)^2 + 4\xi_g^2(\omega/\omega_g)^2} G_0 \tag{9}$$

where G_0 is the constant power spectral density of the input white noise process; ω_g and ξ_g can be interpreted as the characteristic frequency and damping ratio of the ground. Then spectrum compatible realizations are generated via the Spectral Representation method [15] (see Eq. 4).

For experiment purposes, the parameters values as [2] are used herein: $G_0 = 0.01$, $\omega_g = 5\pi$ rad/s, ξ_g =0.63. One single missing data gap at a random location is created in a KT compatible simulation. As an example, Fig. 1 shows a realization of the KT model with a gap of 10% data artificially created. Two BNN models are constructed by training on the two corresponding sides of observations. Recursive imputation by a forward-backward converge scheme are employed for the reconstruction of the gap. Due to the probabilistic nature of the BNN models, an ensemble of reconstructions are obtained as suggested in [18]. It is expressed mathematically by Eq. 6. Then, the nonparametric Welch method [19] is used to compute the PSD estimate on each reconstruction of the ensemble. Welch method is chosen for an acknowledged improved performance on spectral leakage, bias and variability than a classical periodogram.

3.1. Uncertainty on gap size

Conceivably, the size of the gap affects the energy loss in the observed signal. While zero filling in an unevenly-spaced signal may be generally acceptable when only lacking a few data points, a continuous series of zeros lead to prominent energy loss.

Fig. 2 depicts the 95% credible intervals (95% CI) of the PSD estimates with respect to frequencies, in a decibel scale, with a gap size of different sizes. Representatively, four gap sizes (from 5% up to 30%) are considered. A gap of each size is independently created on the same realization at the same location, leaving the size of a gap the one variable that matters. As indicated by the interval range, it's apparent that uncertainty levels are increasing with larger gaps. This is intuitive as the longer the gap is, the more imputations, as opposed to real observations, are inputted for spectral estimation, leading to higher uncertainties. The increasing uncertainties with longer gaps are seen at higher frequencies ranges, approximately from 7 to 15 hz. Also, it can be seen in such higher frequency ranges, the ground truth are mostly located near the bottom of the 95% intervals. It suggested that the trained BNN is having more difficulties estimating the spectral density values for higher frequency components.

To quantitatively account for the uncertainties under different configurations of missing gap, several uncertainty metrics are designed and tabulated, as shown in Table 1. Inspire by [20], a Prediction Interval Coverage Probability (*PICP*) is defined as:

$$PICP = \frac{c}{n} \tag{10}$$

where c represents the total number of frequencies whose PSD estimated value is captured by the 95% credible interval. If note the predicted lower and upper bound as y_L and y_U, then c can be defined by an index variable k_i of length n that represents if every frequency value is captured by the estimated credible interval:

$$c = \sum_{i=1}^{n} k_i \tag{11}$$

$$k_i = \begin{cases} 1 & y_{Li} \le y_i \le y_{Ui} \\ 0 & \text{else} \end{cases} \tag{12}$$

In addition, *AREA* represents the area between the lower and upper bounds and *MAE* is the mean absolute error of the PSD estimates across the frequency range.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} y_i - \hat{y}_i \tag{13}$$

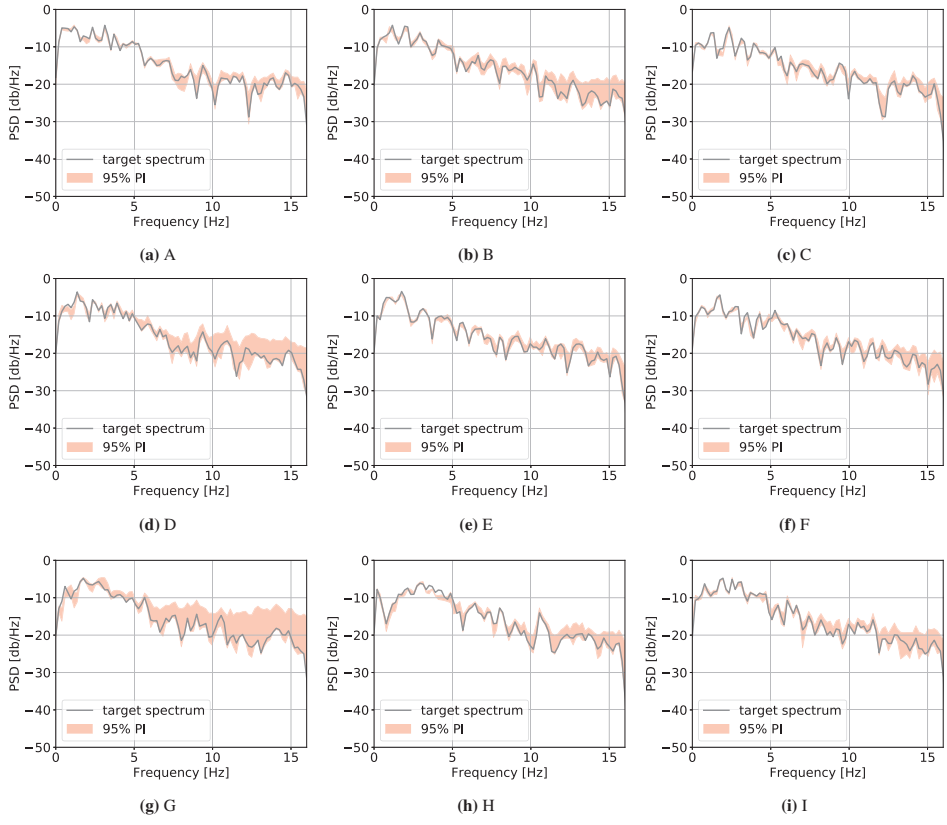

Figure 4. Multiple different realizations with 5% missing gap at random locations

Table 1. Uncertainty metrics with different missing data configurations

gap size	*PICP* [1]	*AREA* [2]	*MAE* [3]
5%	76.92%	17.31	0.39
10%	80.77%	24.18	0.59
20%	69.23%	33.24	1.13
30%	73.08%	53.04	1.72

[1] prediction interval coverage probability of spectral estimates; [2] mean absolute error of spectral estimates;
[3] the area of interval bounds

With the metric *AREA* illustrates the magnitude of uncertainty levels, and *MAE* evaluates the accuracy of the mean estimation, *PICP* has reported the percentage of ground truth PSD values being captured by the credible intervals across the frequency domain.

In more details, we can see the distribution of the spectral density values with respect to each frequency bin, as shown in Fig. 3 using a violin plot. The violin plot contains a box inside, which is the same as a regular box plot where quantiles such as 25%, median and 75% are shown. The white circle represents the median value while the red cross represents the ground truth, i.e., the PSD value from the full realization. In addition, the shape of the distribution of the PSD with respect to each frequency value are also shown via kernel density estimation. As shown in Fig. 3, with 10% missing data, the estimated 95% credible interval well captured the the ground truth.

A thorough tabulation regarding the performance of our proposed method under different gap sizes are listed in Table 1. Quantitatively it can be seen that with longer gaps comes with higher uncertainties, and lower accuracy. Indeed, Fig. 2 has illustratively shown such two effects.

3.2. Uncertainty on gap location

Consider that the gap may occur at any random arbitrary location in the signal, a gap with the size as 5% missing data is created at 20 times, each at a different random location in one complete realization. Uncertainty metrics are listed in Table 2. It can be seen that while the uncertainty level and coverage levels are similar between different locations, there are some deviation in terms of *MAE* errors. Such difference may be resulted from the discrepant length of existing training data observations separated by the gap. Except for the

Table 2. Uncertainties on the location of the gap

gap location	*PICP*	*AREA*	*MAE*
A	0.63	18.83	0.69
B	0.79	18.37	0.43
C	0.69	18.22	0.69
D	0.63	17.76	0.67
E	0.69	17.08	0.65
F	0.63	18.01	0.67
G	0.85	17.05	0.50
H	0.87	17.76	0.38
I	0.65	16.51	0.73
J	0.71	15.85	0.55
K	0.75	16.89	0.43
L	0.58	17.05	0.72
M	0.63	16.13	0.52
N	0.60	18.11	0.68
O	0.69	20.52	0.68
P	0.67	16.39	0.66
Q	0.81	16.32	0.43
R	0.79	17.80	0.40
S	0.73	17.94	0.50
T	0.77	17.65	0.49

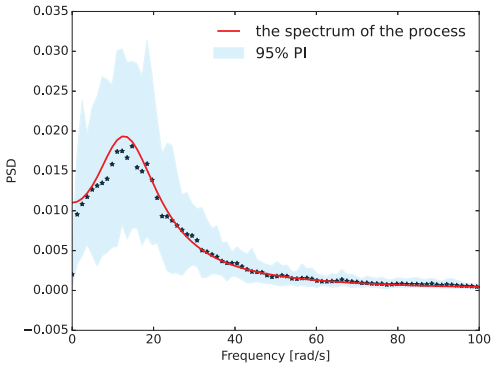

Figure 5. Estimated PSD for 30 realizations - 5% missing gap at random locations

gap is located at exactly in the middle, these two data parts for training two models are of different lengths. One would expect a model with more training data may result in better results. It suggested that due to the location of gap, an weighted strategy may need to be introduced to account for the discrepancies of the two learned models.

3.3. Uncertainty of a hybrid situation

To consider the randomness of the realization and the gap location, we have thereby generated 30 realizations from the KT model in Eq. 9. Such experiment setting is similar as the that in [8], with the only difference as the pattern of missing data: a continuous gap is considered herein whereas scattered missing points are used in their case. Of which, 9 realizations are shown in Fig. 4 as an attempt to illustrate the variance of reconstructed performance based on different realizations. The estimated power spectral density from a single realizations is noisy while an average over an ensemble of realizations stand as a better approximation to the underlying smoothing PSDF, as shown in Fig. 5.

4. Conclusion

This paper presents an investigation of the uncertainties as to the spectral density estimation of incomplete time series with missing gaps. As a reconstruction based method, a Bayesian neural network (BNN) model that works in an autoregressive manner is used to probabilistically learn the temporal patterns in the existing observations and with which iterative imputation are made as reconstructions. Parameter uncertainties of the BNN are represented in probability distributions during training and predictive uncertainties of the iterative reconstructions can thus be computed. For the missing gap, the randomness from the gap size, location as well as the simulated realization from the underlying process are considered. The results suggested generally our interval estimations have well included the ground truth even for a missing gap up to 30%. Longer gaps lead to higher uncertainties and larger errors, especially for relatively higher frequencies. For the BNN based reconstructions, the location of the gap will affect the observations to be trained on and therefore affect the PSD estimates, especially the accuracy. More thorough experiments regarding the bias and variance of the proposed BNN reconstruction approach will be done shortly.

Code

The Python code for the implementation of a Kanai Tajimi model and the Spectral Representation Method can be found at `https://github.com/leslieDLcy/KTnSRM`.

Acknowledgment

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

References

[1] L. Comerford, I. A. Kougioumtzoglou, and M. Beer, "On quantifying the uncertainty of stochastic process power spectrum estimates subject to missing data," *International Journal of Sustainable Materials and Structural Systems*, vol. 2, no. 1-2, pp. 185–206, 2015.

[2] Y. Zhang, L. Comerford, I. A. Kougioumtzoglou, E. Patelli, and 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*, vol. 3, no. 4, p. 04 017 020, 2017.

[3] F. Tobar, "Bayesian nonparametric spectral estimation," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[4] S. Maranò, B. Edwards, G. Ferrari, and D. Fäh, "Fitting earthquake spectra: Colored noise and incomplete data," *Bulletin of the Seismological Society of America*, vol. 107, no. 1, pp. 276–291, 2017.

[5] J. T. VanderPlas, "Understanding the lomb–scargle periodogram," *The Astrophysical Journal Supplement Series*, vol. 236, no. 1, p. 16, 2018.

[6] D. Kondrashov and M. Ghil, "Spatio-temporal filling of missing points in geophysical data sets," *Nonlinear Processes in Geophysics*, vol. 13, no. 2, pp. 151–159, 2006.

[7] D. Kondrashov, R. Denton, Y. Shprits, and H. Singer, "Reconstruction of gaps in the past history of solar wind parameters," *Geophysical Research Letters*, vol. 41, no. 8, pp. 2702–2707, 2014.

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

[9] R. Bos, S. De Waele, and P. M. Broersen, "Autoregressive spectral estimation by application of the burg algorithm to irregularly sampled data," *IEEE Transactions on Instrumentation and Measurement*, vol. 51, no. 6, pp. 1289–1294, 2002.

[10] P. M. Broersen, S. De Waele, and R. Bos, "Autoregressive spectral analysis when observations are missing," *Automatica*, vol. 40, no. 9, pp. 1495–1504, 2004.

[11] J. Christmas, "The effect of missing data on robust bayesian spectral analysis," in *2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, IEEE, 2013, pp. 1–6.

[12] C. Munteanu, C. Negrea, M. Echim, and K. Mursula, "Effect of data gaps: Comparison of different spectral analysis methods," in *Annales Geophysicae*, Copernicus GmbH, vol. 34, 2016, pp. 437–449.

[13] P. Stoica, E. G. Larsson, and J. Li, "Adaptive filter-bank approach to restoration and spectral analysis of gapped data," *The Astronomical Journal*, vol. 120, no. 4, p. 2163, 2000.

[14] M. Priestley, "Power spectral analysis of non-stationary random processes," *Journal of Sound and Vibration*, vol. 6, no. 1, pp. 86–97, 1967.

[15] M. Shinozuka and G. Deodatis, "Simulation of stochastic processes by spectral representation," 1991.

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

[17] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[18] U. B. Oparaji, E. Patelli, S. Rong-Jiun, M. Bankhead, and J. Austin, "Robust Artificial Neural Network for Reliability and Sensitivity Analysis of Complex Non-linear Systems," *Neural Networks*, vol. 96, pp. 80–90, Dec. 2017. DOI: 10.1016/j.neunet.2017.09.003.

[19] S. Miller and D. Childers, *Probability and random processes: With applications to signal processing and communications*. Academic Press, 2012.

[20] T. Pearce, A. Brintrup, M. Zaki, and A. Neely, "High-quality prediction intervals for deep learning: A distribution-free, ensembled approach," in *International conference on machine learning*, PMLR, 2018, pp. 4075–4084.