A Bayesian Augmented-Learning framework for spectral uncertainty quantification of incomplete records of stochastic processes

Yu Chen[a], Edoardo Patelli*[b], Benjamin Edwards[a], Michael Beer[a,c,d]

[a]*Institute for Risk and Uncertainty, University of Liverpool, Liverpool, UK*
[b]*Department of Civil and Environmental Engineering, University of Strathclyde, Glasgow, UK*
[c]*Institute for Risk and Reliability, Leibniz University Hannover, Hannover, Germany*
[d]*International Joint Research Center for Resilient Infrastructure & International Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University, Shanghai, China*

Abstract

A novel Bayesian Augmented-Learning framework, quantifying the uncertainty of spectral representations of stochastic processes in the presence of missing data, is developed. The approach combines additional information (prior domain knowledge) of the physical processes with real, yet incomplete, observations. Bayesian deep learning models are trained to learn the underlying stochastic process, probabilistically capturing temporal dynamics, from the physics-based pre-simulated data. An ensemble of time domain reconstructions are provided through recurrent computations using the learned Bayesian models. Models are characterized by the posterior distribution of model parameters, whereby uncertainties over learned models, reconstructions and spectral representations are all quantified. In particular, three recurrent neural network architectures, (namely long short-term memory, or LSTM, LSTM-Autoencoder, LSTM-Autoencoder with teacher forcing mechanism), which are implemented in a Bayesian framework through stochastic variational inference, are investigated and compared under many missing data scenarios. An example from stochastic dynamics pertaining to the characterization of earthquake-induced stochastic excitations even when the source load data records are incomplete is used to illustrate the framework. Results highlight the superiority of the proposed approach, which adopts additional information, and the versatility of outputting many forms of results in a probabilistic manner.

Keywords: Missing data, Evolutionary power spectrum, Bayesian Deep Learning, AutoEncoder, Stochastic variational inference

1. Introduction

Stochastic processes are widely adopted in many domains to deal with problems which are random in nature and involve strong nonlinearities, non-stationary processes, and uncertain system parameters [1, 2]. For instance, stochastic dynamics involves analyses of engineering systems subject to random environmental processes, such as earthquake motions or winds, requiring realistic characterization and simulation of these stochastic excitations to ensure robust design [3, 4]. Spectral representations of stochastic processes, notably evolutionary power spectral density function (EPSD), play a central role in the characterization and modelling of these environmental processes, capturing key inherent properties (e.g. nonstationarity) and empowering probabilistic engineering simulations for stochastic dynamic analyses and safety assessment of engineering facilities [5, 6, 7].

However, in practice, the uncertainties of spectral representations due to some data problems are not typically acknowledged: (i) limited data: a large number of data samples are needed for a predefined degree of accuracy in power spectral estimation [8]; (ii) incomplete or non-uniformly sampled data: a ubiquitous problem in virtually any discipline where *in situ* measurements are collected and transferred [8, 9]; (iii) uncertain data: inherent statistical differences that exist across multiple source datasets [10]. Specifically, this study focuses on the incomplete data problem as it is deemed both universal and consequential in various fields, e.g. geophysical [11, 12], meteorological [13], astronomical [14, 15] and seismological [16, 17], in which missing data in measurements is frequently an unavoidable issue, with various reasons causing intermittent failure, including equipment failure (e.g. temporary sensor in harsh

conditions), equipment incompetence (old mechanical instrument for high-velocity vibration, broadband sensors vulnerable to clipping in strong motions), temporary transmission loss for real-time data, plus numerous other reasons including sensor maintenance, usage, data acquisition restrictions and or data-corruption [8, 17, 18]. This problem also turns out to have significant consequences, as incompleteness may lead to false interpretations (e.g. artefacts) if not properly dealt with [16], or render the data unusable, hence breaking existing work pipelines, as re-observe the considered physical/environmental phenomenon is generally not possible.

In many studies and applications, data records contain considerable information towards the understanding of the physical phenomenon/event/scenario (e.g. spectral structure), and towards the development and calibration of empirical models of the physical processes (e.g. spectral models for characterizing seismic stochastic excitations), possibly for a certain condition or scenario, or a specific region. A frequent dilemma is that available observed data are scarce and limited in the first place to justify the formulation and calibration of models accordingly developed [19]. As such, in response to limited data, a method to harness the most of existing data (even if incomplete or nonequidisantly-sampled) is considered essential for a variety of practical applications. Facilitating such uses would then ensure the user to continue with the ensuing analyses, enrich the database, and progressively obtain better informed models, which is of great importance to ensuring robust analyses against data problems in practice.

With varying assumptions, many attempts in tackling incomplete data have been proposed over the years. Various parametric models, assuming certain structures of the underlying process, e.g. AR (autoregressive) or ARMA (autoregressive moving-average) models (see [20, 21, 22] for details), involve fitting a parameterized formulation of the spectral density to the available observation using specific estimators. In [17], a domain specific parameterized formulation, based on the physics-driven earthquake source spectrum of ground motion, is fitted with a maximum likelihood estimator. Sparse methods [23, 24, 25, 26] (e.g. compressive sensing), with additional assumption of the sparsity, for instance, in frequency domain, have been proposed for spectral density estimation where multiple records compatible with a stochastic process are available [9]. Furthermore, a number of nonparametric spectral estimators that worked in an iterative manner to progressively approximate the target spectral density, see e.g. [16, 27, 28] among others. However, there exists an additional barrier that many these spectral estimators are only valid for stationary processes and cannot be straightforwardly extended to nonstationary data [29]. Alternatively, a number of approaches explicitly or implicitly convert the spectral analysis involving missing data into a process of iterative imputation, followed by well-established full-data spectral analysis techniques. Beyond simple interpolation methods (see e.g. [30] for a review), more sophisticated models stand out by utilizing temporal dependency. Notably, neural network models, known for learning complex and nonlinear relations, are seen prospects in learning the underlying process [18, 31, 19] and thereby imputing the missing values [32].

However, despite recent progress, there still exist three main challenges in the spectral analysis of non-stationary process with missing observations: (i) mostly current approaches fail to properly address the uncertainties related to the missing data [33, 8]. Inaccuracies of an imperfect time-history reconstruction will be propagated to spectral estimates. Similarly, for parametric modelling approaches, parameter uncertainties due to the incomplete data are not adequately captured. (ii) many current approaches are developed on the stationary assumption hence inadequate to reflect the spectral nonstationarity; (iii) More importantly, in spite of approaches that handle uncertainties (e.g. Bayesian spectral analyses [34, 35] or interval discrete Fourier transform [36]), most of current approaches are still significantly bounded by a ceiling in performance since they merely base on the very limited information contained in the incomplete data (i.e. available observations).

Therefore, in relation to these challenges, we propose an Augmented-Learning framework that (i) takes advantage of *a-priori* knowledge of the underlying process, enabling to incorporate additional information (physics-based knowledge) into the modelling. (ii) accounts for uncertainty throughout the framework, allowing to provide a host of outputs in a probabilistic manner (e.g. reconstructions, spectral representations, and stochastic-process sample generations). (iii) applicable to nonstationary processes. The present study builds upon a previous work [19] which merely addresses missing data in a stationary setting, whereas this study provides a robust solution to the more general and complex case of arbitrary missing pattern anywhere in a non-stationary setting with significant incompleteness. Such versatility therefore enables its domain-independent feasibility. This paper is structured as follows: in section 2, we begin with a concise review of the theoretical context on which we build our framework. Section 3 then elaborates the main procedures of our proposed framework, followed by a discussion regarding one key component (Bayesian modelling of sequential data) in Section 4. An example application from stochastic dynamics is used to illustrate the framework in section 5, where we present a comprehensive performance comparison of three Bayesian deep learning

models, under a range of missing data scenarios, using quantitative uncertainty metrics.

2. Spectral representations 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 [37]:

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

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 [2] 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 [7, 38, 39]. Inversely, a versatile formula for generating sample realizations compatible with the stochastic process is given by spectral representation method (SRM) [39]:

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

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.

3. Augmented Bayesian learning framework

A large ensemble of complete data samples are often required for stochastic-process spectral density estimation for attaining a predefined adequate degree of accuracy, while we often only have one observed realization in practice [40]. The estimation becomes even more challenging when only partial data is available. Limited information in the partially observed data imposes a ceiling in performance for those accordingly developed methods. To robustly exploit additional information into modelling and break the performance ceiling, a Bayesian Augmented-Learning framework is established. Fig. 1 shows a flowchart of the key procedures of the proposed framework.

We build on the premise that *a priori* knowledge could provide general yet insightful prior expectations of the observation (with variaibility) of the physical process. The *a-priori* information is addressed by generating simulations based on the domain knowledge represented by $\boldsymbol{\theta}_g$.

$$\mathcal{S} = g(\boldsymbol{\theta}_g) \tag{4}$$

specifically, $\boldsymbol{\theta}_g = (\theta_1, \ldots, \theta_n)$ represents a random vector of relevant physical parameters, each component of which stands for a random variable. $g(\cdot)$ represents a generator function, which may just be a model with physics aspects capable of generating stochastic simulations accordingly. Collectively the corresponding probability distribution $p(\boldsymbol{\theta}_g)$ would reflect the variability of the simulations embedded in our prior belief.

Given the data represented by those physics-informed simulations, Bayesian recurrent neural network models $\mathcal{M}$ are trained as probabilistic model representations of the underlying process, whereby the imputation of missing data is conducted as predictions in a recursive manner. Importantly, the epistemic uncertainties of the learned model representations are addressed by putting probability distributions over the model parameters $\boldsymbol{\omega}$ of neural nets, thus giving rise to the posterior distribution $p(\boldsymbol{\omega}|\mathcal{S}, \mathcal{M})$ through the Bayesian inference, as given below:

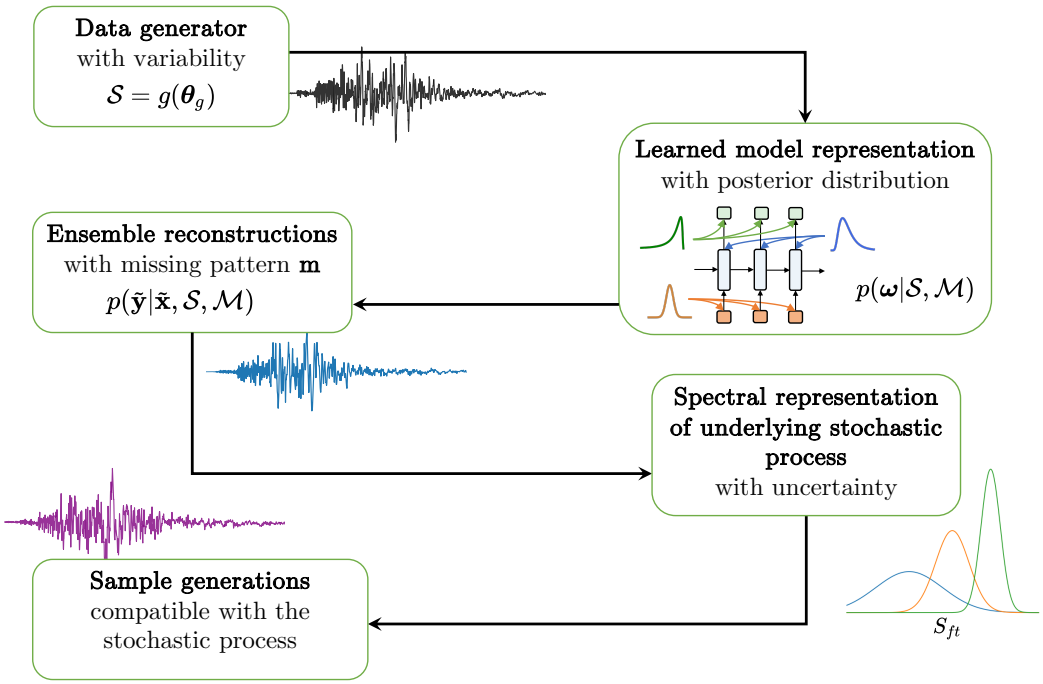


Figure 1: Flowchart of the Augmented Bayesian Learning framework. Main components of the framework include **a**. generating simulations of the physical process with *a-priori* knowledge; **b**. learning model representations of the underlying process with Bayesian recurrent models; **c**. imputing probabilistically the missing values with the learned Bayesian models; **d**. quantifying the uncertainty on spectral representations (e.g. evolutionary spectrum) of the underlying stochastic process **e**. simulating sample realizations of stochastic process as inputs to subsequent problems of random nature (e.g. stochastic dynamics)

$$p(\boldsymbol{\omega}|\mathcal{S}, \mathcal{M}) = \frac{p(\mathcal{S}|\boldsymbol{\omega}, \mathcal{M})p(\boldsymbol{\omega}|\mathcal{M})}{p(\mathcal{S}|\mathcal{M})} \propto p(\mathcal{S}|\boldsymbol{\omega}, \mathcal{M})p(\boldsymbol{\omega}|\mathcal{M}) \tag{5}$$

This marks a key step of the proposed framework, where a probabilistic representation of the underlying processes is learned by recurrent neural network models and further used to reconstruct the incomplete observations. Besides, in this study we also present investigations and comparisons of a few neural network architectures in this regard. With the posterior distribution, an ensemble of recurrent imputations can be obtained by marginalizing out the parameter space, as follows:

$$\mathcal{R} = \int p(\tilde{\mathbf{y}}|\tilde{\mathbf{x}}, \boldsymbol{\omega})p(\boldsymbol{\omega}|\mathcal{S})\mathrm{d}\boldsymbol{\omega} \tag{6}$$

where $\tilde{\mathbf{x}}$ represents the missing samples in a specific recording; $\mathcal{R}$ denotes the reconstructed process, practically through an ensemble of reconstructions, which contain both the recurrent imputations $\tilde{\mathbf{y}}$ and existing observations. Subsequently, uncertainties over spectral representations of the underlying stochastic process (e.g. evolutionary power spectral density) can further be quantified, using any established spectral estimators, non-parametric or parametric, stationary or non-stationary. Importantly, the evolutionary spectral density with respect to a certain time and frequency S_{ft} is represented by a probability distribution, as opposed to a deterministic value.

Closely related to the notion of evolutionary power spectrum is the application of Monte Carlo simulations of compatible sample functions for numerical engineering analyses of stochastic nature [1, 2, 6, 7], for instance conducting stochastic response and reliability assessment for engineering structures subject to stochastic excitations. Corresponding to Eq. (3), our framework maintains the ability to characterise the underlying stochastic process and generate associated sample realizations, even with the source data record is incomplete.

4. Bayesian modelling of sequential data

Bayesian recurrent neural network (BRNN) models are utilized to probabilistically learn the temporal dependency and provide recurrent imputations for the missing data in the measurements. Specifically, three network architectures (namely long short-term memory, or LSTM, LSTM-Autoencoder, LSTM-Autoencoder with teaching forcing mechanism) are investigated. For completeness, we first provide a concise review of the sequential modeling strategies using RNNs and extend these models into Bayesian counterparts to further consider epistemic uncertainty.

4.1. Sequential modelling with LSTM

RNNs are specialized dynamic models that capture temporal patterns in the sequential data (e.g. time series), by maintaining hidden states at each time step [41, 42], see Eq. (7). They feature the recursive structure that consumes the time ordered data one at a time. Its structure is deemed as a deep network once unfolded in time.

$$\mathbf{h}_t = \mathcal{H}(\mathbf{h}_{t-1}, \mathbf{x}_t; \boldsymbol{\omega}) \tag{7}$$

where $\mathbf{h}_t$ and $\mathbf{x}_t$ represent, respectively, the hidden states vector and the input sequence $\mathbf{x}_t \in \mathbb{R}^D$, at time stamp t; $\mathcal{H}$ denotes a hidden layer function, which could represent any sophisticated RNN variant (e.g. long short-term memory, LSTM) parameterized by weights and biases $\boldsymbol{\omega}$. Notably, the LSTM architecture [43], acknowledged for alleviating vanishing or exploding gradients and learning long range temporal dependencies, are found to give state-of-the-art results for a variety of prediction problems of sequential nature [41, 42, 32]. Fig. 2 depicts the diagram of a LSTM unit, which encapsulates the flow of states through three gate functions (namely: forget gate f, input gate i, output gate o) plus a cell update $\tilde{c}$, controlling the flow of information, as shown by following equation [44]:

$$\begin{aligned}
\mathbf{f}_t &= \sigma(\mathbf{W}_{xf}\mathbf{x}_t + \mathbf{W}_{hf}\mathbf{h}_{t-1} + \mathbf{b}_f) \\
\mathbf{i}_t &= \sigma(\mathbf{W}_{xi}\mathbf{x}_t + \mathbf{W}_{hi}\mathbf{h}_{t-1} + \mathbf{b}_i) \\
\mathbf{o}_t &= \sigma(\mathbf{W}_{xo}\mathbf{x}_t + \mathbf{W}_{ho}\mathbf{h}_{t-1} + \mathbf{b}_o) \\
\tilde{\mathbf{c}} &= \tanh(\mathbf{W}_{cx}\mathbf{x}_t + \mathbf{W}_{ch}\mathbf{h}_{t-1} + \mathbf{b}_c) \\
\mathbf{c}_t &= \mathbf{f}_t * \mathbf{c}_{t-1} + \mathbf{i}_t * \tilde{\mathbf{c}}_{t-1}
\end{aligned} \tag{8}$$

where σ stands for the sigmoid function, tanh denotes the hyperbolic function, $*$ is the element-wise multiplication. $\boldsymbol{\omega}$ collectively represent the weight matrices (including biases terms) aforementioned as $\boldsymbol{\omega} = \{\mathbf{W}_{xf}, \mathbf{W}_{hf}, \mathbf{b}_f, \ldots, \mathbf{W}_{\mathbf{y}}, \mathbf{b}_{\mathbf{y}}\}$, which represent the input-to-hidden connections, hidden-to-hidden recurrent connections, and also hidden-to-output connections. With the hidden states at time t sequentially updated as $\mathbf{h}_t = \mathbf{o}_t * \tanh(\mathbf{c}_t)$, the associated prediction is given by [45]:

$$y_t = f_{\mathbf{y}}(\mathbf{h}_t) = \mathbf{h}_t\mathbf{W}_{\mathbf{y}} + \mathbf{b}_{\mathbf{y}} \tag{9}$$

4.2. LSTM-based AutoEncoder

LSTM-based AutoEncoder stands for a specified architecture that concerns a sequence-to-sequence inference problem where a variable length of predictions are desired (often referred as horizon), using an encoder-decoder structure. It consists of an encoder LSTM network that encodes the input data sequence into a context vector $\boldsymbol{\nu}$, whereby a decoder LSTM network is conditioned upon to iteratively generate the output sequence of arbitrary length given the hidden states, as suggested by Eq. (9). Fig. 3 provides a simplified diagram illustrating the encoder-decoder network structure. The context vector, in the latent space, is deemed to have summarized the hidden states sequentially learned from the input data.

From a probabilistic perspective [47] which deems the model as the probabilistic procedure generating the observed data, the autoencoder architecture formulates the conditional distribution of an output sequence given the input sequence, $p(y_1, \ldots, y_H | x_1, \ldots, x_L)$, as given below:

$$p(\mathbf{y}_H|\mathbf{x}_L) = \prod_{t=1}^{H} p(y_t|\boldsymbol{\nu}, y_1, \ldots, y_{t-1}) \tag{10}$$

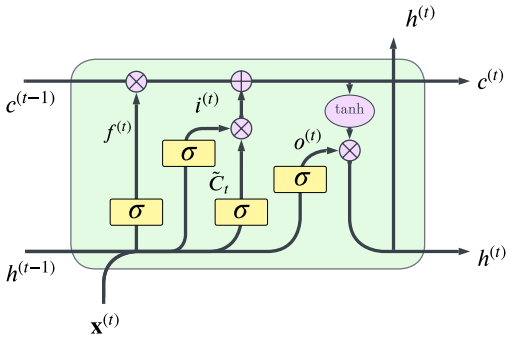


Figure 2: Diagram of a LSTM unit. Three gate functions f, i, o control information passage

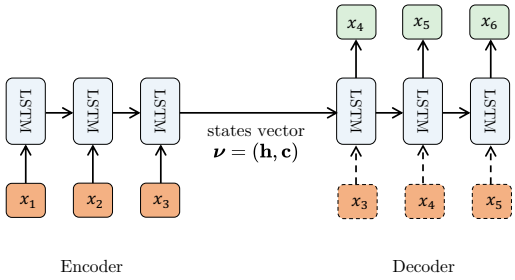


Figure 3: A simplified diagram of the LSTM Autoencoder architecture where a univariate case is illustrated (modified from [46]). The dashed arrow indicate the optional information employed by the teacher forcing mechanism

In the predictive setting of a univariate time series, these sequence pairs (L, H) correspond to a past lagged window as input and future horizon steps as the output, under an autoregressive manner [48]. Note that a univariate case, characterised as $y_t = x_{L+t}$ is shown herein, but note that such architecture is not limited to univariate cases but also excel in modelling input sequences of high dimensions [46] (i.e. $\mathbf{x}_t \in \mathbb{R}^D$ where covariates are available), or even sequences of different domains (such as statistical machine translation [42]). We omit the covariates herein for notation simplicity, but it applies to a multi time-series setting as well when there exist extra features (e.g. relevant physical or geological factors) contributing to the modelling of the physical processes.

During training, the encoder-decoder networks are jointly optimized (e.g. by stochastic gradient descent) to maximize the likelihood of seeing the observed data.

$$\boldsymbol{\omega}^{MLE} = \arg\max_{\boldsymbol{\omega}} \frac{1}{N} \sum_{i=1}^{N} \log p(\mathbf{y}_H^{(i)} | \mathbf{x}_L^{(i)}, \boldsymbol{\omega}) \tag{11}$$

where MLE stands for the maximum likelihood estimation procedure. $\boldsymbol{\omega}$ collectively represents the model parameters and $(\mathbf{x}_L^{(i)}, \mathbf{y}_H^{(i)})$ denotes the window-horizon pairs for the ith data point. Fig. 3 illustrates the architecture of the LSTM Autoencoder for illustration, where the encoder and decoder can be composed of stack of LSTM layers. In particular, Fig. 3 also depicts a variant of the LSTM Autoencoder which characterises the teacher forcing mechanism [42], (i.e. besides conditioning on the context vector, the decoder additionally takes in target sequence but offset by one time-step $\mathbf{y}_H[t-1]$). Effectively, it reinforces the learning of the data generating process by feeding more information (ground-truth information) where available. It should be thus noted the decoding procedures would differ at inference stage due to the lack of this additional information, while the prediction from the last time step will instead be used.

Fundamentally, LSTM-based models are specialized structures trained to learn temporal patterns. Once learned, disregard of the model architecture, it can be exploited for sequence generation [41], time series forecasting [48], stochastic signal simulation [18], and also missing data imputation [32].

4.3. Bayesian Deep Learning

Implicit in the above MLE procedure is the ignorance of model uncertainties. Significant uncertainties may exist on the model configurations that could have generated the data. Especially in the context of limited data, deterministic models, unless properly regularized, are prone to learn too much noise (overfitting) and become overconfident due to the unawareness of model uncertainties [49, 50]. Therefore, in accounting for the model uncertainty (epistemic uncertainty) in neural network models, probability distributions are assigned to model parameters $\boldsymbol{\omega}$ [47]. Particularly, by formulating the uncertainty, Bayesian models achieves a regularizing effect against overfitting [49], which may otherwise be a serious problem in terms of limited and noisy data.

4.3.1. Variational Bayesian Learning

To efficiently approximate the true posterior distribution, under the condition of the huge dimensions of a deep neural network model, stochastic variational inference (see e.g. [51, 52, 53]) 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 [49]:

$$\mathcal{J}(\mathcal{D}, \theta) = \mathrm{KL}[q(\mathbf{w}|\theta) \parallel p(\mathbf{w})] - \mathbb{E}_{q(\mathbf{w}|\theta)} \log p(\mathcal{D}|\mathbf{w}) \tag{12}$$

which stands for the negative lower bound of the evidence term $\log p(\mathcal{D})$, i.e. negative ELBO (see Appendix B for details). The formulation of Eq. (12) 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.

4.3.2. Variational Bayesian inference in RNNs with stochastic regularisation techniques

Evaluation of the stochastic objective and further gradients is challenging and several Monte Carlo estimators are adopted as approximate solutions [53]. 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 (as in Eq. (7)), correspondingly, the negative ELBO in the case of RNN, can be written as [45]:

$$\mathcal{J}_R = \mathbb{E}_{q(\omega)} \log p\Big(\mathbf{y}|f_{\mathbf{y}}^{\omega}(f_{\mathbf{h}}^{\omega}(\mathbf{x}_T, f_{\mathbf{h}}^{\omega}(\dots f_{\mathbf{h}}^{\omega}(\mathbf{x}_1, \mathbf{h}_0)\dots)))\Big) + \mathrm{KL}[q(\boldsymbol{\omega}) \parallel p(\boldsymbol{\omega})] \tag{13}$$

where $\boldsymbol{\omega}$ collectively represents all the parameters in a LSTM model. Corresponding to Eq. (8) these parameters are modelled as random variables. 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 [45]:

$$q(\mathbf{w}_k) = p\mathcal{N}(\mathbf{w}_k; \mathbf{0}, \sigma^2\mathbf{I}) + (1-p)\mathcal{N}(\mathbf{w}_k; \phi_k, \sigma^2\mathbf{I}) \tag{14}$$

where the random weight matrix is factorized over the rows as $\boldsymbol{\omega}_k = g(\phi_k, \boldsymbol{\epsilon}) = \phi_k \cdot \mathrm{diag}(\boldsymbol{\epsilon}_k)$; 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 [52] for details), a further approximation of the stochastic objective function [45]:

$$\mathcal{J}_R \approx -\sum_{l=1}^{N} \log\Big(\mathbf{y}|f_{\mathbf{y}}^{\omega^{(l)}}(f_{\mathbf{h}}^{\omega^{(l)}}(\mathbf{x}_T, f_{\mathbf{h}}^{\omega^{(l)}}(\dots f_{\mathbf{h}}^{\omega^{(l)}}(\mathbf{x}_1, \mathbf{h}_0)\dots)))\Big) + \lambda\|\boldsymbol{\phi}\|_2^2 \tag{15}$$

$$\boldsymbol{\omega}^{(l)} = g(\boldsymbol{\phi}, \boldsymbol{\epsilon}^{(l)}) \text{ with } \boldsymbol{\epsilon}^{(l)} \sim p(\boldsymbol{\epsilon}) \tag{16}$$

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. (13) by L2 regularisation with weight decay λ and variational parameters $\boldsymbol{\phi}$ to be solved, see [50] for more details. In minimizing Eq. (15), 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. Particularly, the above variational Bayesian optimization procedures suggest a large deal of conceptual similarity (but distinct implementation differences) with the dropout mechanism [54], which approximates model averaging of exponentially many different neural nets efficiently.

Correspoding to Eq. (6), substituting the Bernoulli variational distribution for the true posterior then approximates the predictive distribution for each missing point, as given below:

$$\int p(\tilde{\mathbf{y}}|\tilde{\mathbf{x}}, \boldsymbol{\omega}) q(\boldsymbol{\omega}) \mathrm{d}\boldsymbol{\omega} \approx \frac{1}{T}\sum_{t=1}^{T} p(\tilde{\mathbf{y}}|\tilde{\mathbf{x}}, \boldsymbol{\omega}^{(t)}) \tag{17}$$

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.

5. Application example

In this paper we demonstrate the procedures and advantage of the proposed framework with an application example. We extend the analysis of characterizing earthquake-induced stochastic excitations based on records subject to missing data [4] by specifically using a real accelerogram from the European Engineering Strong Motion (ESM) database [55], illustrating how physics-based *a-priori* knowledge can be harnessed to facilitate the estimation (and also uncertainty quantification) of the evolutionary power spectral density (EPSD) of the underlying stochastic process. It's of great importance to stochastic dynamic response and reliability analyses when the associated earthquake scenario is of interests to the seismic hazard/risk assessment of the engineering facility, especially in a data scarce region.

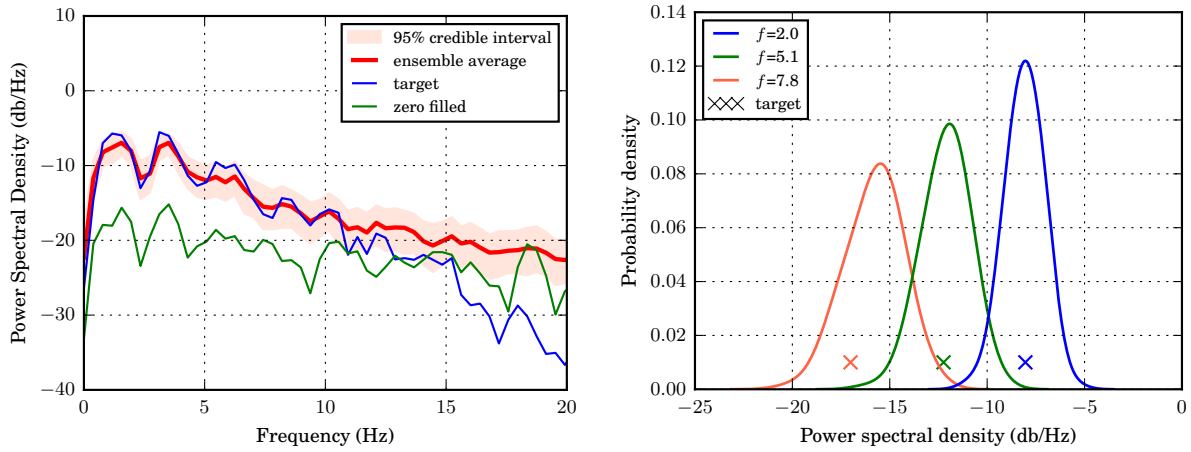
Compared with previous studies (e.g. [18, 9, 4]) which jointly employ multiple incomplete realizations artificially created towards the spectral density estimation of a single process, this study, however, tackles with the empirical recording. It should be noted that when working with empirical recordings, there is generally a single observed seismic recording available as a realization of a stochastic process, plus the true power spectrum of the underlying process being unknown [56]. As such, it increases the difficulty in obtaining accurate spectral representation and motivates the uncertainty to be appropriately accounted for. A seismic record of magnitude $M_w = 6.5$, normal faulting, epicentral distance $R = 18.6km$, recorded at a class A site in Italy is adopted in this analysis, whereby a range of missing scenarios upon this target recording will be created and investigated. The spectral estimates from the otherwise full recording would then serve as the reference or target for comparison. In computing the uncertainty of the stochastic process spectral estimates, this study further considers three Bayesian recurrent model architectures with various windowing settings.

5.1. Data generation based on a-priori information

Physics-based simulations for training the Bayesian neural network models are generated from a nonstationary stochastic ground-motion model [57, 58], which contains *a-priori* seismological knowledge well-calibrated for the region compatible with the target recording. Importantly, these nonstationary simulations are produced through a model formulation that encapsulates physical components (discretized finite-fault comprised of Brune's earthquake point-source model, realistic time envelope function, non-stationarity in frequency, ground-motion variability) [59, 60], using parameters with physical meanings, $\boldsymbol{\theta}_g = [I_a, D_v, F_c, F_b]$, which represent Arias intensity, significant duration, central frequency and frequency bandwidth respectively. These parameters are empirically related to earthquake characteristics, namely $[M_w, R_{epi}, V_{s30}, F_s]$, which respectively represent magnitude, epicentral distance, 30m average shear wave velocity and fault type, via a regressive relationship that entails the contribution of the source, path, and site effects. This model has been validated with the strong motion data of the region of interest to reflect the seismological knowledge of the given region [61].

5.2. Missing data at random locations

For generality, we consider missing data of arbitrary patterns, which are commonly referred to as irregularly-sampled records and widely studied in the literature [14, 15, 13, 11, 12]. Missing data are created following the setting of MCAR (missing completely at random) [62]. We denote the occurrence of missing samples by a binary masking vector $\mathbf{m}_t \in \{0, 1\}$ where $m_t = 1$ represents the missing observation at random time index t, drawn from a uniform distribution [18]. We deem this strategy to have simulated well an unevenly sampling pattern. In fact, as with the increase of missing percentage, missing values are more likely to group together, transitioning into a gapped missing pattern.



(a) Global PSD estimates from the ensemble reconstructions

(b) Probability density functions of PSD with respect to selected frequencies

Figure 4: Uncertainty over estimated power spectral density (PSD) with 70% missing data ($\epsilon_{MP} = 0.7$) based on the Bayesian Autoencoder model with teacher forcing

5.3. Detailed results for one missing scenario by one Bayesian recurrent model

In this section, detailed results are demonstrated for the performance regarding the Bayesian Autoencoder model with teacher-forcing mechanism (abbreviated as BtfAutoencoder), under a serious scenario of 70% randomly missing data (see Fig. 4 - Fig. 8). Note that only one Bayesian recurrent neural network model with respect to one missing scenario is shown herein for conciseness, whereas a comprehensive comparison of various model settings and missing scenarios will follow shortly in Section 5.4.

In this study, 100 physics-based simulations are generated to train the Bayesian recurrent models, which all consist of 4 layers of 128 LSTM units followed by a fully connected layer for prediction purpose. In the case of the two models with Autoencoder architecture, each encoder/decoder model will then be composed of half of these layers, with the final fully-connected layer appended to the decoder model. An ensemble of time domain reconstructions are provided through recursive predictions of the BtfAutoencoder model. Fig. A.15 shows the example incomplete recording with 70% missing data (i.e. missing percentage $\epsilon_{MP} = 70\%$), while Fig. A.16 displays the ensemble of reconstructed time-history (ensemble size as 500). Collectively it can be observed that those imputations closely match the target values from the otherwise complete record and well contained by the 95% credible intervals.

Estimation of the spectral representation of the underlying stochastic excitation plays a central role in stochastic dynamic analyses to accurately capture the system behaviour [2, 3, 7, 63]. Importantly, Fig. 4a displays the uncertainty over the power spectral density estimated from the ensemble reconstructions, with each frequency component corresponding to a probability distribution. These distributions can better be seen in Fig. 4b, which illustratively shows the distribution shape of spectral density values as well as the target PSD (as marked) from the otherwise complete recording at selected frequencies. Despite 70% of data being missing, the ensemble-averaged PSD estimate agrees well with the target PSD from the complete recording, whose spectral peaks follow closely with the ensemble average. Also, the target spectral values across the whole frequency range are well captured by the 95% credible interval bounds, except only for the range higher than $15Hz$. But it should be noted that by plotting in decibel scale, which implicitly reflects results in log scale, the differences in the higher frequency ranges are indeed very small (i.e. in the range of 10^{-3} in linear scale). Therefore, the target PSD is well approximated across whole frequency range by the ensemble average. The cutoff at $20Hz$ is due to the little contribution to the signal power thereafter. By comparison, a baseline approach that simply fills in zeros for missing points suffers a significant power loss, especially in the peak ranges (e.g. 0-$5Hz$).

The stationary (global) PSD estimates are inadequate in reflecting the nonstationary characteristics of seismic motions for giving spectral decomposition in an average sense. These time-varying properties are of particular importance to nonlinear response analysis of engineering structures due to the evolving resonant effect [64]. Therefore,

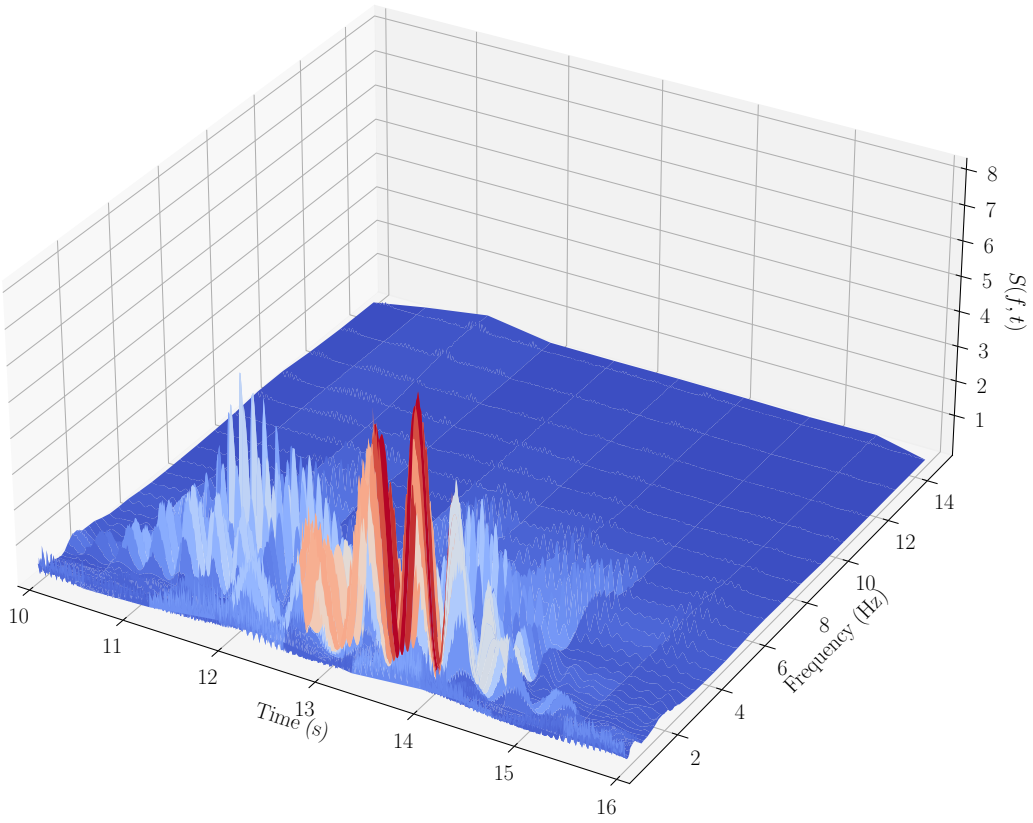


Figure 5: Ensemble-averaged evolutionary spectrum (EPSD) by Morlet wavelet transform with 70% missing data ($\epsilon_{MP} = 0.7$) based on the Bayesian Autoencoder model with teacher forcing

an ensemble of estimates of the evolutionary power spectrum are computed by wavelet transform (Morlet wavelet) [38] in this analysis, with the averaged EPSD shown in Fig. 5. In terms of uncertainty, the probability distribution of EPSD values, $S(f,t)$, at selected time instants and frequency bin are displayed in Fig. 6 for illustration, where 3 representative time instants are selected to show the evolution of spectral estimates. Vertical lines in purple indicate the target spectral values without missing data, which are well captured by the corresponding distribution.

Relying on the Monte Carlo simulation approach [2] powered by the spectral representation method [6] (see Eq. (3)), sample realizations compatible with the underlying stochastic process characterized by the evolutionary spectra are simulated. These synthetic generations could further be employed for stochastic nonlinear dynamic analyses (see e.g. [65, 3, 66, 67]). Seismic excitations are also frequently characterized by response spectra. Fig. 7 shows the pseudo-acceleration response spectra (5% damped) of these sample realizations in light gray, compared to the target response spectrum of the seismic record as highlighted by the thick line in red. It can be seen that, even with a missing rate as high as 70%, the target response spectrum is largely captured by the range of those from our sample realizations, except for longer periods higher than 3 seconds where a bias can be spotted. Notice that such bias is systematically less significant with less missing data, as indicated by Fig. 11 where response spectra associated with missing percentage of 10%, 30%, 50% are shown. Also, it has been found with other stochastic simulation models (see e.g. [64]) in situations where no missing data exist and can be sufficiently mitigated by a high frequency filter. In addition, Fig. 8 shows, side by side, one of the sample realization based on the ensemble-averaged EPSD estimate (at the bottom), along with one of the ensemble reconstructions directly from our BtfAutoencoder model (in the middle), compared with the otherwise full target recording (top).

In summary, it has been shown the applicability of the proposed method in characterizing the stochastic excitations and spectral uncertainty quantification based on incomplete record with 70% of missing data. It allows to provide a

host of probabilistic representations, e.g. reconstructed time-history, evolutionary spectral estimates, response spectra, and additionally Monte Carlo sample simulations of the underlying stochastic process.

5.4. Comprehensive performance comparison of considered models and missing scenarios

Bayesian recurrent models play a central role in learning the temporal dependency and probabilistically represent the data generating process. This study further investigates the performance of spectral estimation and uncertainty quantification regarding the three specialized recurrent architectures. Besides this, a range of missing percentages (ϵ_{MP}) are additionally experimented to encompass various situations with different missing data patterns and missing degrees (up to 70%). Higher ϵ_{MP} will naturally lead to gaps in the data.

Fig. 10 shows the comparison of PSD estimates, by the BtfAutoencoder model, under 3 different missing levels. It can be seen that the ensemble-averaged estimation closely follows the target in all 3 scenarios, especially in identifying spectral peaks. It has previously been noted in Fig. 4 that the discrepancies after $15Hz$ are exponentially exaggerated by the implicitly suggested log scale. These discrepancies are indeed very small (i.e. in the range of 10^{-3} in linear scale). With more missing data, the uncertainties of PSD estimates are accordingly increasing as indicated by the 95% credible interval, and the power loss by the zero-filled method is even stronger, making it an insufficient choice when a significant amount of data missing. In Fig.11, the response spectra of sample generations contain the target response spectrum fairly well, with some amount of variability desired around the target, in order for characterising random excitations for further stochastic response analyses.

In comparing the spectral estimation based on different Bayesian recurrent models, as an example, Fig. 9 shows a comparison of probability distribution between the three Bayesian models, namely Bayesian LSTM, Bayesian AutoEncoder, and Bayesian AutoEncoder with teacher-forcing mechanism, with respect to the scenario of $\epsilon_{MP} = 0.2$ and $f = 1.4Hz$. Importantly, it indicates that uncertainty on the spectral estimate with respect to any frequency or time stamp, under a certain missing scenario, across the Bayesian models of choice, can be accounted for within our framework.

In order to facilitate the quantitative comparison of the performance with respect to missing scenarios and neural network model settings, several measures of uncertainty are designed and reported in both time domain and frequency domain, reflecting the effects on the characterization for both the excitation process and engineering responses. In particular, the spectral dissimilarity is computed by the Wasserstein distance (W_F, see Eq. (18)) between (normalized) power spectral densities [68], reflecting the differences of spectral energy distribution. P_{95} corresponds to an interval coverage probability measure [69] that reflects the percentage of the target values (eg. PSD) being captured by the

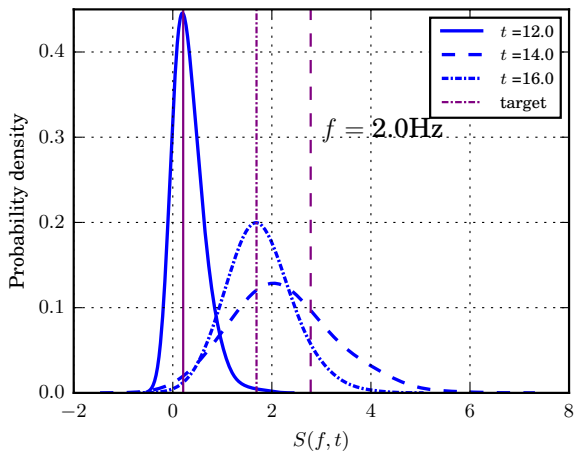


Figure 6: Probability density of estimated evolutionary spectrum (EPSD) shown at selected time instants based on the Bayesian Autoencoder model with teacher forcing ($\epsilon_{MP} = 0.7$). Vertical lines suggest the target spectral value without missing data.

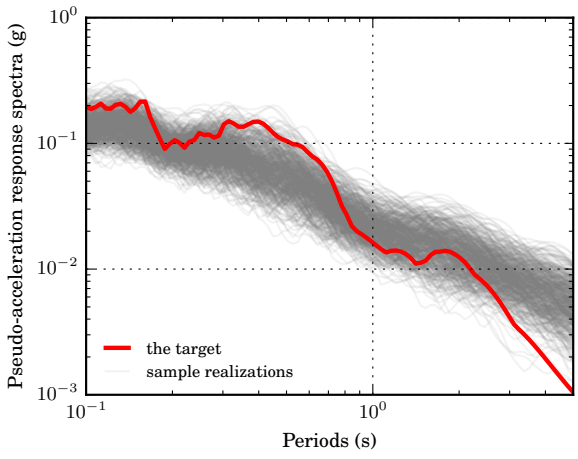


Figure 7: Pseudo-acceleration response spectra (5% damped) of sample realizations of the underlying stochastic process by SRM from the estimated EPSD based on Bayesian Autoencoder model with teacher forcing ($\epsilon_{MP} = 0.7$). Thick line suggests the target response spectrum without missing data.

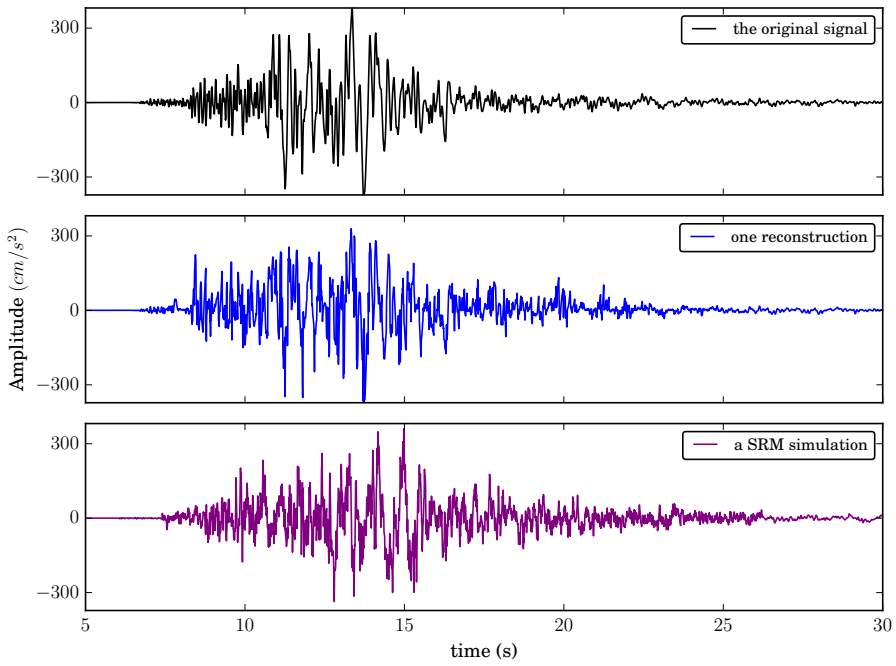


Figure 8: Target recording (top) compared with a direct reconstruction from the Bayesian Autoencoder model with teacher-forcing mechanism (middle) and, as an example of subsequent simulation based on the underlying stochastic process, a sample generation from the ensemble-averaged EPSD using the spectral representation (SRM) method (bottom) $\epsilon_{MP} = 0.7$

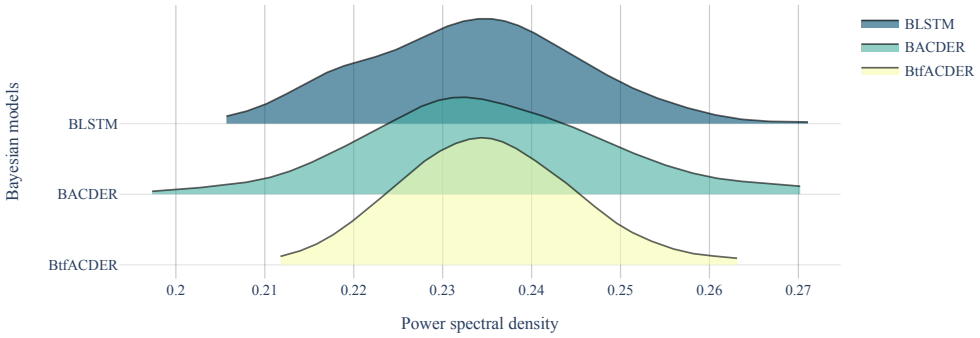
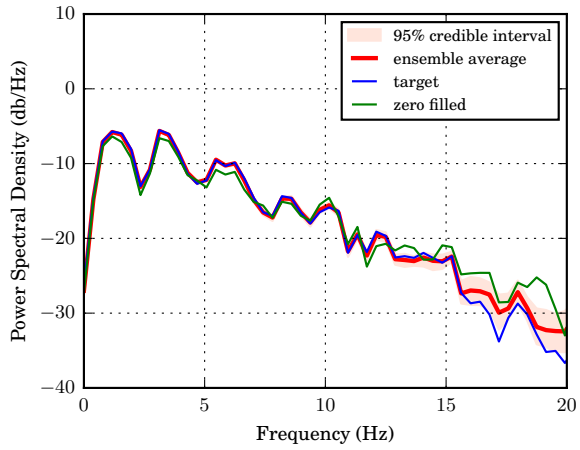


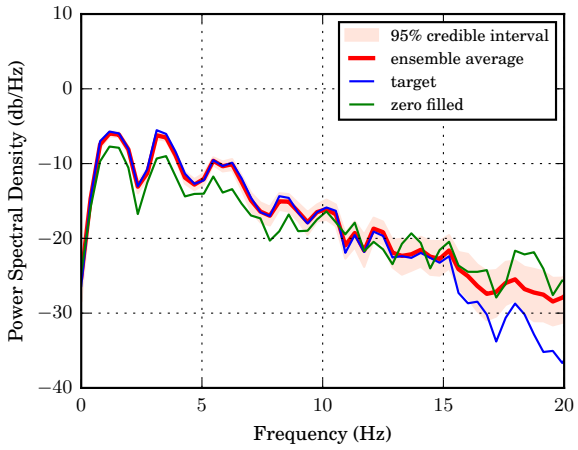
Figure 9: Probability distributions of power spectral density (PSD) for three Bayesian recurrent network models with respect to the scenario f=1.4Hz, $\epsilon_{MP} = 0.2$

estimated 95% credible intervals. In addition, δ_{95} denotes the width between the lower bound y_U and upper bound y_L of the credible intervals, which illustrates the magnitude of uncertainty levels in the estimates. It should be noted that these two measures P_{95} and δ_{95} should be evaluated together as a desired high quality interval will be narrow while capturing a certain portion of data. An extremely wide interval, despite capturing all the ground truth, will instead be of little practical use. Furthermore, e denotes the mean absolute error of time domain reconstructions, which evaluates the accuracy of the imputation. Note that the uncertainty measures, P_{95} and δ_{95}, are reported both in time domain and frequency domain, whereas W_F and e are responsible for denoting the discrepancy in spectral estimates and imputations respectively.

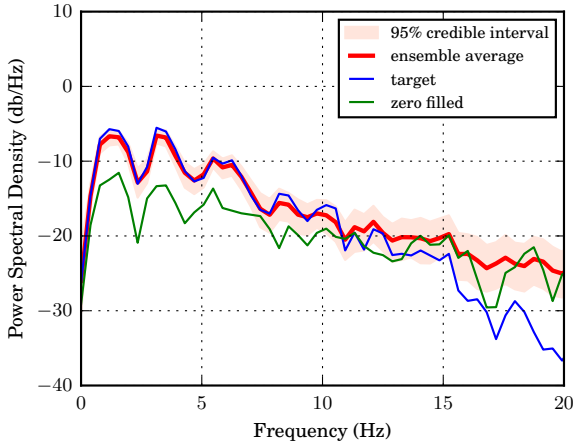
$$W_p(\mu, \nu) = \Big(\int_0^1 |F_\mu^{-1}(q) - F_\nu^{-1}(q)|^p \mathrm{d}q \Big)^{1/p} \tag{18}$$



(a) $\epsilon_{MP} = 10\%$

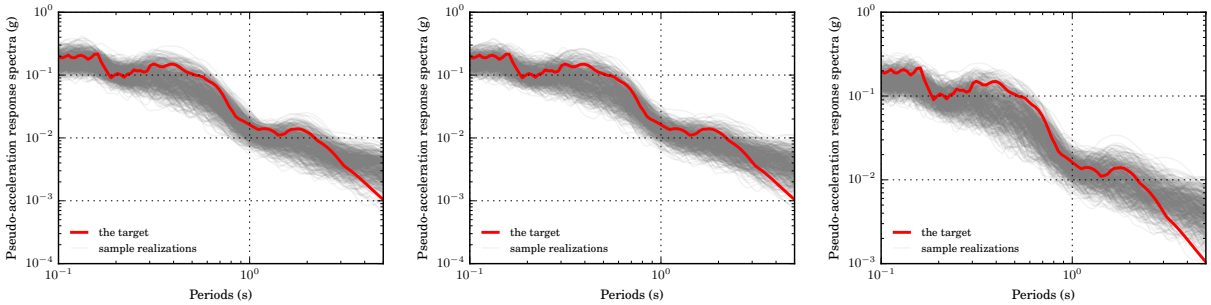


(b) $\epsilon_{MP} = 30\%$



(c) $\epsilon_{MP} = 50\%$

Figure 10: Comparison of estimated PSD under varying missing level based on the Bayesian Autoencoder model with teacher forcing



(a) $\epsilon_{MP} = 10\%$ (b) $\epsilon_{MP} = 30\%$ (c) $\epsilon_{MP} = 50\%$

Figure 11: Comparison of elastic response spectra from sample realizations of the underlying stochastic process based on the Bayesian Autoencoder model with teacher forcing under varying missing level

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 [68].

$$P_{95} = c_f/n_f, \text{ with } c_f = \sum_{i=1}^{n} c_i \tag{19}$$

$$c_i = \begin{cases} 0, y_i \in [y_{U_i}, y_{L_i}] \\ 1, y_i \notin [y_{U_i}, y_{L_i}] \end{cases} \tag{20}$$

$$\delta_{95} = \frac{1}{n} \sum_{i=1}^{n} (y_{U_i} - y_{L_i}) \tag{21}$$

$$e = \frac{1}{n} \sum_{n=1}^{n} |y_i - \hat{y}_i| \tag{22}$$

where c_f is defined by a vector of length n_f (total number of frequency bins), whose element c_i indexes a frequency value captured by the estimated credible interval.

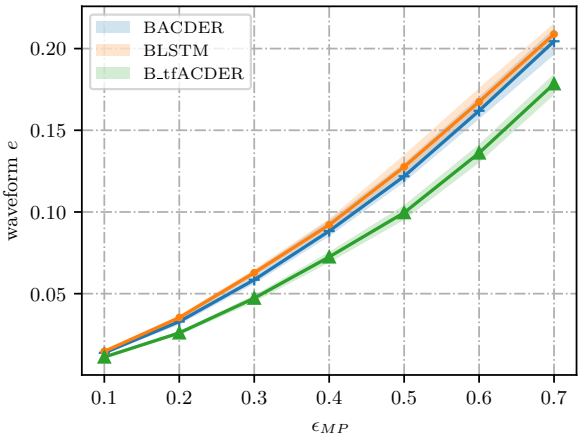
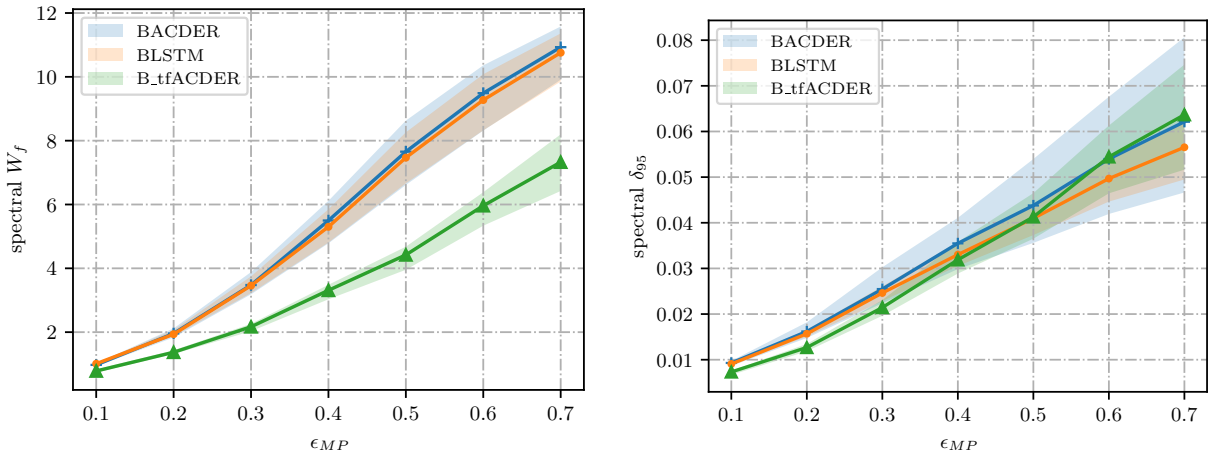


Figure 12: Mean absolute error (e) of the time domain imputations by the Bayesian recurrent neural network models with various windowing choices under a range of missing levels

As a result, Fig. 12 shows the accuracy of time-history reconstructions across the Bayesian recurrent models in the time domain. The shaded region specifically reflects the effects of varying windowing choices considered for each Bayesian model as suggested by Eq. (10). Note that, in working with univariate time series data, the LSTM based models are implemented in an autoregressive fashion. In comparison with an ordinary AR model with linear fixed coefficients, LSTM-based models manifest as dynamic models that can additionally model complex and nonlinear relations. The choices of window pairs, which could be considered as hyperparameters, are specifically tuned and compared in this analysis, as partly shown in Table 1, where the uncertainty evaluations under various missing percentages, by metrics developed in Eq. (18)-Eq. (22), on both spectral density estimates and time domain imputations is listed. Note that abundant windowing settings in terms of (L, H) (in total 250) according to a grid search scheme are tested within the framework, but only a few of them are tabulated in Table 1 due to the limit of space. Their effects are displayed in Fig.12 - Fig.13, where the size of shading suggests the variance of performances between the same model but with varying windowing choices. These shading are shown as 95% credible interval encompassing all the windowing choices considered.

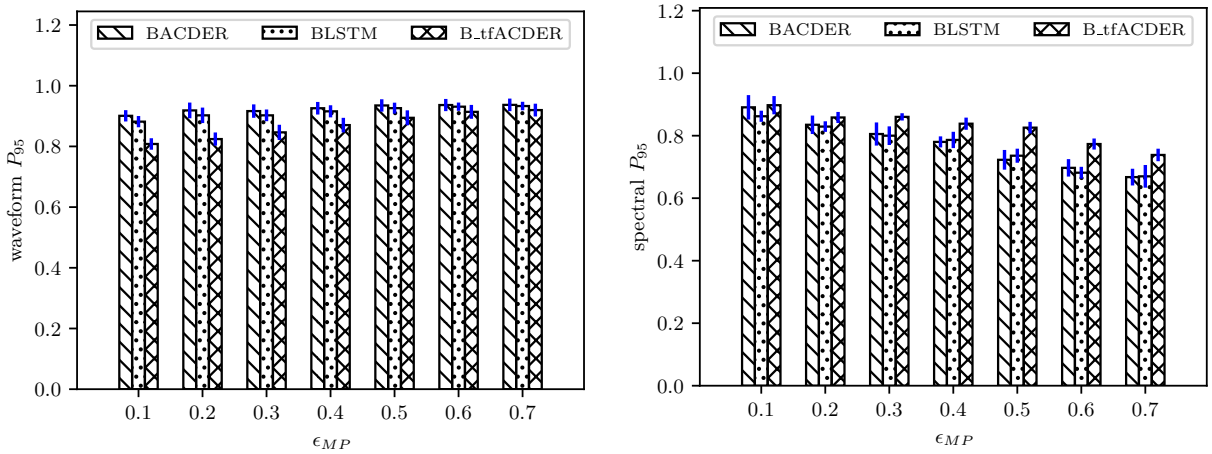
In the frequency domain, Fig. 13a shows the dissimilarity of spectral estimates using the Wasserstein distance, while Fig. 13b shows the degree of spectral uncertainty for the three Bayesian models with respect to each missing percentage. The markers (eg. e in Fig.12, W_f in Fig. 13a, δ_{95} in Fig. 13b) then indicate the mean results in three aspects: (i) the metric is a global measure across the frequency domain or time domain (ii) it is computed as ensemble-average for the ensemble of reconstructions by a certain Bayesian model under a missing scenario; (iii) it is the



(a) Wasserstein distance W_f

(b) Width between lower and upper bounds of credible interval δ_{95}

Figure 13: Spectral power density estimates across a range of missing percentages and three Bayesian recurrent models with varying windowing choices



(a) Prediction interval probability coverage of the time domain reconstructions P^t_{95}

(b) Prediction interval probability coverage of the power spectral density P^f_{95}

Figure 14: Prediction interval probability coverage (P_{95}) results for both recurrent imputations in the time domain and spectral power density estimates in the frequency domain across a range of missing percentages and three Bayesian recurrent models with varying windowing choices

mean of a Bayesian model with abundant windowing choices. In terms of the accuracy of imputation and spectral estimates (see Fig.12 and Fig. 13a), The Bayesian Autoencoder model with teacher forcing outperforms the other two by achieving the lowest discrepancy, while the Bayesian LSTM and Bayesian Autoencoder model have very similar performance. The variability increases as with the missing percentage for all three models, but BtfACDER still has the least variance regarding the windowing settings. In terms of spectral uncertainty (Fig. 13b), all three models have small differences at each missing percentage. But BACDER has the largest the variability.

Fig. 14 displays the prediction interval coverage for both imputations and spectral estimates (PSD). As discussed earlier, when we associate P_{95} with the δ_{95} metric, it is observed that the coverage probability for imputations for all the missing percentages have high coverage probability (over 80%), though at the cost of wider interval bounds for large percentage of missing data. The error bar suggests the variance of varying windowing choices for each Bayesian model. While BtfACDER achieves the highest coverage probability for spectral estimates, it has the lowest coverage probability in imputation compared to the other two models, though the difference is fairly small.

Table 1: Results of uncertainty metrics for a range of missing scenarios with respect to Bayesian models of varying windowing settings. A subset of results are listed here due to the limit of space

		Bayesian LSTM					Bayesian Autoencoder					Bayesian tfAutoencoder				
		L					L, H					L, H				
ϵ_{MP}		5	10	15	20	30	(5, 3)	(10, 3)	(15, 3)	(20, 3)	(30, 3)	(5, 3)	(10, 3)	(15, 3)	(20, 3)	(30, 3)
0.1	e	0.015	0.014	0.015	0.015	0.015	0.014	0.013	0.013	0.013	0.013	0.010	0.011	0.011	0.011	0.011
	P^t_{95}	0.848	0.904	0.888	0.888	0.880	0.888	0.944	0.936	0.912	0.904	0.808	0.824	0.816	0.824	0.840
	W_f	1.010	0.989	1.033	0.984	1.036	0.963	0.949	0.961	0.934	1.003	0.732	0.757	0.804	0.793	0.788
	δ_{95}	0.009	0.009	0.009	0.009	0.009	0.009	0.010	0.010	0.010	0.010	0.008	0.007	0.007	0.008	0.008
	P^f_{95}	0.845	0.893	0.864	0.845	0.864	0.903	0.961	0.932	0.913	0.932	0.913	0.903	0.864	0.922	0.913
0.2	e	0.035	0.034	0.035	0.036	0.037	0.033	0.032	0.031	0.032	0.032	0.025	0.025	0.026	0.026	0.025
	P^t_{95}	0.860	0.936	0.916	0.908	0.892	0.900	0.956	0.948	0.952	0.940	0.844	0.816	0.836	0.848	0.868
	W_f	1.873	1.869	1.892	1.980	2.069	1.832	1.967	1.926	1.977	2.095	1.316	1.376	1.362	1.385	1.377
	δ_{95}	0.015	0.016	0.015	0.016	0.017	0.016	0.018	0.017	0.018	0.018	0.013	0.012	0.013	0.013	0.013
	P^f_{95}	0.806	0.854	0.835	0.816	0.835	0.854	0.864	0.854	0.854	0.854	0.845	0.845	0.845	0.864	0.854
0.3	e	0.063	0.061	0.062	0.063	0.065	0.059	0.057	0.056	0.057	0.056	0.045	0.045	0.046	0.047	0.045
	P^t_{95}	0.88	0.936	0.901	0.907	0.888	0.904	0.944	0.944	0.939	0.944	0.853	0.861	0.856	0.893	0.888
	W_f	3.222	3.35	3.384	3.589	3.748	3.186	3.619	3.723	3.798	3.883	2.144	2.144	2.189	2.253	2.221
	δ_{95}	0.023	0.025	0.024	0.025	0.026	0.023	0.028	0.029	0.028	0.03	0.021	0.021	0.023	0.023	0.023
	P^f_{95}	0.777	0.845	0.825	0.786	0.767	0.816	0.835	0.825	0.835	0.864	0.874	0.854	0.864	0.854	0.854
0.4	e	0.091	0.09	0.091	0.093	0.096	0.088	0.087	0.085	0.088	0.087	0.071	0.07	0.072	0.072	0.069
	P^t_{95}	0.878	0.936	0.918	0.928	0.918	0.914	0.946	0.948	0.952	0.956	0.868	0.888	0.87	0.908	0.918
	W_f	4.807	5.142	5.204	5.505	5.851	4.886	5.836	6.048	6.031	6.115	3.392	3.45	3.476	3.421	3.366
	δ_{95}	0.031	0.036	0.032	0.033	0.034	0.033	0.04	0.041	0.04	0.041	0.032	0.032	0.033	0.034	0.036
	P^f_{95}	0.786	0.835	0.777	0.777	0.757	0.777	0.806	0.777	0.777	0.777	0.835	0.854	0.845	0.874	0.864
0.5	e	0.124	0.124	0.128	0.128	0.135	0.119	0.122	0.121	0.120	0.121	0.096	0.097	0.099	0.099	0.095
	P^t_{95}	0.891	0.941	0.930	0.939	0.930	0.915	0.954	0.960	0.954	0.965	0.877	0.904	0.894	0.931	0.938
	W_f	6.660	7.242	7.381	7.799	8.261	6.676	8.182	8.472	8.430	8.629	4.259	4.675	4.592	4.590	4.646
	δ_{95}	0.037	0.044	0.039	0.041	0.043	0.039	0.050	0.051	0.050	0.054	0.040	0.041	0.043	0.044	0.046
	P^f_{95}	0.748	0.767	0.738	0.728	0.699	0.738	0.757	0.738	0.728	0.718	0.825	0.825	0.825	0.854	0.806
0.6	e	0.163	0.163	0.167	0.169	0.176	0.157	0.164	0.164	0.159	0.162	0.131	0.130	0.136	0.137	0.131
	P^t_{95}	0.911	0.947	0.928	0.945	0.924	0.923	0.957	0.963	0.961	0.964	0.903	0.928	0.932	0.937	0.959
	W_f	8.323	9.126	9.200	9.641	10.082	8.463	10.188	10.360	10.220	10.317	5.935	6.330	6.019	6.191	6.029
	δ_{95}	0.045	0.054	0.049	0.049	0.052	0.047	0.063	0.066	0.061	0.068	0.052	0.054	0.055	0.059	0.061
	P^f_{95}	0.709	0.689	0.680	0.680	0.650	0.689	0.738	0.718	0.709	0.718	0.757	0.738	0.767	0.757	0.777
0.7	e	0.207	0.203	0.208	0.211	0.216	0.203	0.205	0.212	0.203	0.201	0.175	0.174	0.180	0.176	0.172
	P^t_{95}	0.907	0.950	0.938	0.937	0.933	0.918	0.949	0.966	0.957	0.968	0.909	0.927	0.925	0.954	0.954
	W_f	9.853	10.672	10.934	10.966	11.360	10.204	11.542	11.453	11.450	11.188	7.657	8.198	7.597	7.581	7.626
	δ_{95}	0.049	0.062	0.054	0.058	0.059	0.052	0.072	0.079	0.072	0.081	0.060	0.063	0.064	0.071	0.075
	P^f_{95}	0.650	0.728	0.689	0.660	0.621	0.689	0.650	0.660	0.660	0.670	0.757	0.728	0.728	0.728	0.699

ϵ_{MP} missing percentage, e mean absolute error, P^t_{95} interval coverage probability for time domain reconstructions, W_f Wasserstein distance for PSD estimates, δ_{95} credible interval bounds width for PSD estimates, P^f_{95} interval coverage probability for PSD estimates

6. Conclusion

Missing data is a ubiquitous problem in various disciplines, where data observations are crucial for the understanding and model development of the underlying physical process. In this paper, a novel Bayesian Augmented-Learning framework for quantifying the uncertainty in spectral density estimation of stochastic process in the presence of missing data is developed. Many existing spectral estimators accounting for missing data are driven merely from the limited available observations and ignore the uncertainty, hence imposing a ceiling in performance and reliability. This paper, therefore, presents a framework that (i) accounts for uncertainty throughout the framework (ii) takes advantage of prior domain knowledge (iii) applicable to nonstationary processes. It allows to recover the spectral representation of the underlying stochastic process by probabilistically reconstructing the incomplete recording with additional information available (though imperfect) about the underlying physical phenomenon. The proposed method provides a host of characterizations in a probabilistic manner (e.g. reconstructed time-histories, spectral representations, and compatible Monte Carlo sample generations), facilitating the uses in various applications, either spectral-based or waveform-based.

Within the proposed framework, this paper presents a comprehensive performance comparison of three Bayesian

deep recurrent models with various model settings, under a range of missing data scenarios, using quantitative uncertainty metrics. While the results suggest efficacious of all the models even with a significant amount of data missing, the LSTM Autoencoder with teacher forcing provides the most accurate power spectral density estimates. Particularly noteworthy, is the ability of the proposed framework to remain effective even when 70% of data are missing. This robustness under significant incompleteness is largely attributed to the capacity of long-range memory (modelling long-range temporal dependency) and the mechanism of dynamic hidden states benefited from the LSTM architecture. In addition, the combination of teacher-forcing mechanism of the Autoencoder improves information extraction in learning complex temporal relations. By contrast, a classic dense architecture neglects the temporal relation and thus can be unstable (yielding huge spikes) in long-range predictions under a nonstationary setting.

This framework provides a robust solution to the general arbitrary missing data pattern in a non-stationary setting, even under significant incompleteness. Of particular note is the versatility of the framework enabling potential uses in other domains and independence of domain-parameters. While we show a successful example in characterizing stochastic excitation in engineering dynamics, the framework can be adopted in other fields of processes or statistical signals, where some *a priori* knowledge about the underlying process is available, which may typically be in the forms of theories on the governing PDE (partial differential equation), numerical models of complex physical systems, or parameterized stochastic model formulation involving physical variables or parameters.

Importantly, such knowledge provides considerable information regarding the data generating mechanism than merely the remaining incomplete observation. The similar issue of missing observation in the data series of physical processes, and the typical existence of prior studies (i.e. physical models) of relevant physical processes, as well as the versatility of the proposed framework in modelling arbitrary missing data in a nonstationary setting, suggest the generalized feasibility of the proposed framework.

Another noteworthy aspect in accounting for uncertainty within the framework is that we focused on the epistemic uncertainty in learning model representations of the underlying process. But still, the aleatoric uncertainty led by the data noise, which may be more concerning for a less strong motion from long distance, contributes to the uncertainty of reconstruction. However, the estimation of heteroscedastic aleatoric uncertainty in the recurrent prediction during temporal propagation is a nontrivial task, which we will address in the future study.

7. Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

8. Acknowledgements

This work was supported by the EU Horizon 2020 - MSCA Actions project URBASIS [Project no. 813137] Special thanks to Xiangyu Feng for the guidance on the distribution strategy of multi GPU system. The authors are grateful for the supportive suggestions from the anonymous reviewers that have helped improve the paper.

Appendix A. Time domain reconstructed time-history

Appendix B. Stochastic variational inference

Bayesian inference in Deep Learning concerns learning the posterior distribution after seen the data. However, the true posterior is generally intractable due to the complexity of the model (eg. huge dimensions of parameters space). Consider a regression task of learning a model with parameters $\mathbf{w}$ of the conditional distribution $p(\mathbf{y}|\mathbf{x}, \mathbf{w})$ from a dataset $\mathcal{D} : (\mathbf{x}_i, \mathbf{y}_i)_{i=1}^N$, stochastic variational inference finds a variational distribution, parameterised by $\boldsymbol{\theta}$, that minimizes the Kullback-Leibler divergence between the proposed variational distribution and the true posterior:

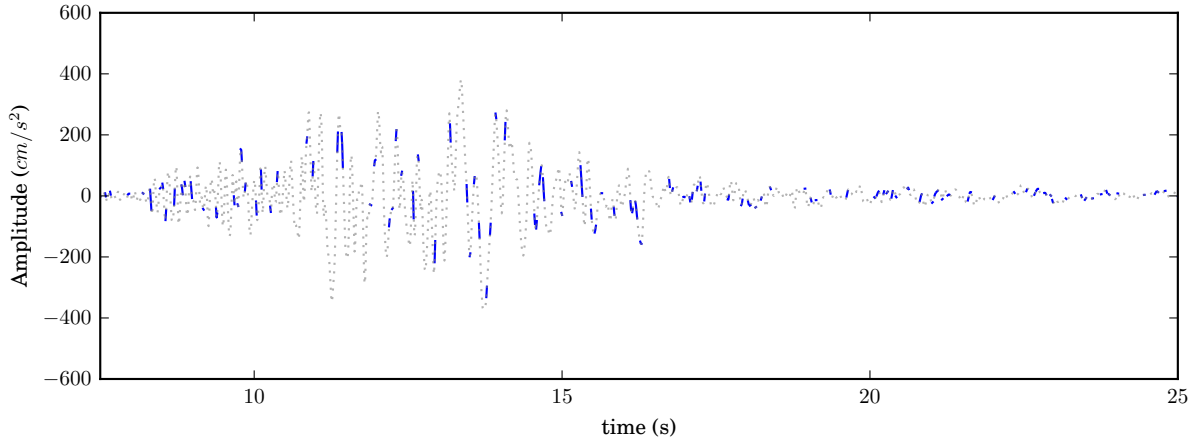


Figure A.15: Example incomplete recording with 70% data randomly missing (ϵ_{MP} = 70%)

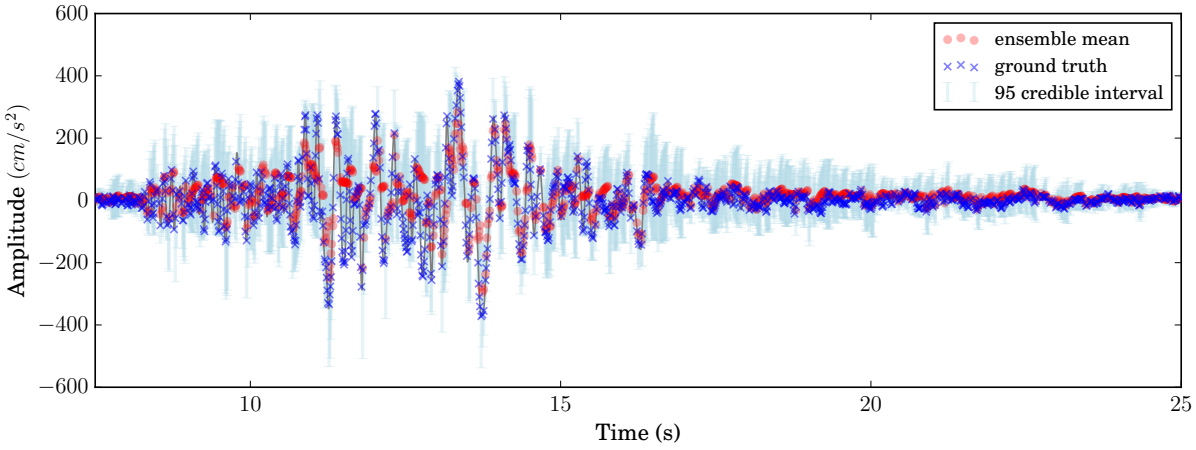


Figure A.16: Uncertainty over the recurrent imputations in the time domain with an ensemble size of 500 based on the Bayesian Autoencoder model with teacher-forcing (ϵ_{MP} = 70%)

$$
\begin{aligned}
D_{KL}[q(\mathbf{w}|\boldsymbol{\theta})||p(\mathbf{w}|\mathcal{D})] &= \mathbb{E}_{q(\mathbf{w}|\boldsymbol{\theta})} \log \frac{q(\mathbf{w}|\boldsymbol{\theta})}{p(\mathcal{D}|\mathbf{w})p(\mathbf{w})} p(\mathcal{D}) \\
&= \mathbb{E}_{q(\mathbf{w}|\boldsymbol{\theta})}[\log q(\mathbf{w}|\boldsymbol{\theta}) - \log p(\mathcal{D}|\mathbf{w}) - \log p(\mathbf{w})] + \log p(\mathcal{D}) \\
&= D_{KL}[q(\mathbf{w}|\boldsymbol{\theta})||p(\mathbf{w})] - \mathbb{E}_{q(\mathbf{w}|\boldsymbol{\theta})} \log p(\mathcal{D}|\mathbf{w}) + \log p(\mathcal{D}) \qquad \text{(B.1)}
\end{aligned}
$$

Rearranging terms could further obtain the evidence lower bound $\mathcal{L}(\mathcal{D}, \boldsymbol{\theta})$ as suggested by Eq. (B.2). Importantly, as the marginal log likelihood $\log p(\mathcal{D})$ is constant with respect to θ, maximizing the ELBO will equivalently minimize the original KL divergence.

$$
\begin{aligned}
\mathcal{L}(\mathcal{D}, \boldsymbol{\theta}) &= \mathbb{E}_{q(\mathbf{w}|\theta)} \log p(\mathcal{D}|\mathbf{w}) - \mathrm{KL}[q(\mathbf{w}|\boldsymbol{\theta}) \,||\, p(\mathbf{w})] \\
\mathcal{L}(\mathcal{D}, \boldsymbol{\theta}) &= \log p(\mathcal{D}) - \mathrm{KL}[q(\mathbf{w}|\boldsymbol{\theta})||p(\mathbf{w}|\mathcal{D})] \leqslant \log p(\mathcal{D}) \qquad \text{(B.2)}
\end{aligned}
$$

Appendix C. List of symbols

ω	Angular frequency
$A(\omega, t)$	Time and frequency dependent modulating function
$Z(\omega)$	Orthogonal process
$S(\omega, t)$	Evolutionary power spectral density function
$\mathcal{S}$	*A priori* data
$\boldsymbol{\theta_g}$	Physical parameters characterised with *a priori* knowledge
$\mathcal{M}$	Bayesian recurrent neural network models
$\mathcal{R}$	Ensemble of reconstructions
$\boldsymbol{\omega}$	Weights and bias of a neural network model
$\tilde{\mathbf{x}}$	Missing data
$\tilde{\mathbf{y}}$	Recurrent imputations by the Bayesian recurrent models
$\mathbf{h}_t$	Hidden states at time t
$\mathcal{H}$	Hidden layer function of a recurrent network
f	Forget gate
o	Output gate
i	Input gate
$\tilde{\mathbf{c}}$	Cell update
$\boldsymbol{\nu}$	latent vector from encoder
$\boldsymbol{\epsilon}$	Bernoulli sample vector
I_a	Arias intensity
D_v	Significant duration
F_c	Central frequency
F_b	Frequency bandwidth
M_w	Earthquake magnitude
R_{epi}	Epicentral distance
V_{s30}	Shear wave velocity
F_s	Fault type
$\mathbf{m}$	Masking vector for missing data
L	Lagged window size
H	Horizon size
e	Mean absolute error
P_{95}	Interval coverage probability measure
W_f	Wasserstein Fourier distance
δ_{95}	Width between the upper bound and lower bound
F_{-1}	Inverse cumulative distribution
$\mathcal{J}$	Loss objective
$\mathbf{y}_L$	Lower bound of the credible interval
$\mathbf{y}_U$	Upper bound of the credible interval

References

[1] M. Shinozuka, G. Deodatis, Simulation of stochastic processes by spectral representation (1991).

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

[3] A. D. Kiureghian, K. Fujimura, Nonlinear stochastic dynamic analysis for performance-based earthquake engineering, Earthquake Engineering & Structural Dynamics 38 (5) (2009) 719–738.

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

[5] S.-S. P. Lai, Statistical characterization of strong ground motions using power spectral density function, Bulletin of the Seismological Society of America 72 (1) (1982) 259–274.

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

[7] P. Spanos, I. Kougioumtzoglou, Harmonic wavelets based statistical linearization for response evolutionary power spectrum determination, Probabilistic Engineering Mechanics 27 (1) (2012) 57–68.
[8] Y. Zhang, L. Comerford, I. A. Kougioumtzoglou, E. Patelli, M. Beer, Uncertainty quantification of power spectrum and spectral moments estimates subject to missing data, ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 3 (4) (2017) 04017020.
[9] L. Comerford, I. A. Kougioumtzoglou, M. Beer, Compressive sensing based stochastic process power spectrum estimation subject to missing data, Probabilistic Engineering Mechanics 44 (2016) 66–76.
[10] M. Behrendt, M. Bittner, L. Comerford, M. Beer, J. Chen, Relaxed power spectrum estimation from multiple data records utilising subjective probabilities, Mechanical Systems and Signal Processing 165 (2022) 108346.
[11] L. Smith-Boughner, C. Constable, Spectral estimation for geophysical time-series with inconvenient gaps, Geophysical Journal International 190 (3) (2012) 1404–1422.
[12] J. P. Musial, M. M. Verstraete, N. Gobron, Comparing the effectiveness of recent algorithms to fill and smooth incomplete and noisy time series, Atmospheric chemistry and physics 11 (15) (2011) 7905–7923.
[13] A. Shtiliyanova, G. Bellocchi, D. Borras, U. Eza, R. Martin, P. Carrère, Kriging-based approach to predict missing air temperature data, Computers and Electronics in Agriculture 142 (2017) 440–449.
[14] D. Kondrashov, R. Denton, Y. Shprits, H. Singer, Reconstruction of gaps in the past history of solar wind parameters, Geophysical Research Letters 41 (8) (2014) 2702–2707.
[15] T. Stahn, L. Gizon, Fourier analysis of gapped time series: Improved estimates of solar and stellar oscillation parameters, in: Helioseismology, Asteroseismology, and MHD Connections, Springer, 2008, pp. 31–52.
[16] S. Baisch, G. H. Bokelmann, Spectral analysis with incomplete time series: an example from seismology, Computers & Geosciences 25 (7) (1999) 739–750.
[17] S. Maranò, B. Edwards, G. Ferrari, D. Fäh, Fitting earthquake spectra: colored noise and incomplete data, Bulletin of the Seismological Society of America 107 (1) (2017) 276–291.
[18] L. Comerford, I. A. Kougioumtzoglou, M. Beer, An artificial neural network approach for stochastic process power spectrum estimation subject to missing data, Structural Safety 52 (2015) 150–160.
[19] Y. Chen, E. Patelli, B. Edwards, M. Beer, A physics-informed bayesian framework for characterizing ground motion process in the presence of missing data, Earthquake Engineering & Structural Dynamics 52 (7) (2023) 2179–2195. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/eqe.3877, doi:https://doi.org/10.1002/eqe.3877.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/eqe.3877
[20] P. M. Broersen, R. Bos, Time-series analysis if data are randomly missing, IEEE transactions on instrumentation and measurement 55 (1) (2006) 79–84.
[21] R. Martin, Autoregression and irregular sampling: Spectral estimation, Signal processing 77 (2) (1999) 139–157.
[22] R. Bos, S. De Waele, P. M. Broersen, Autoregressive spectral estimation by application of the burg algorithm to irregularly sampled data, IEEE Transactions on Instrumentation and Measurement 51 (6) (2002) 1289–1294.
[23] Y. Yang, S. Nagarajaiah, Harnessing data structure for recovery of randomly missing structural vibration responses time history: Sparse representation versus low-rank structure, Mechanical Systems and Signal Processing 74 (2016) 165–182.
[24] Y. Zhang, L. Comerford, I. A. Kougioumtzoglou, M. Beer, Lp-norm minimization for stochastic process power spectrum estimation subject to incomplete data, Mechanical Systems and Signal Processing 101 (2018) 361–376.
[25] G. D. Pasparakis, K. R. dos Santos, I. A. Kougioumtzoglou, M. Beer, Wind data extrapolation and stochastic field statistics estimation via compressive sampling and low rank matrix recovery methods, Mechanical Systems and Signal Processing 162 (2022) 107975.
[26] Y. Wang, T. Zhao, K.-K. Phoon, Direct simulation of random field samples from sparsely measured geotechnical data with consideration of uncertainty in interpretation, Canadian Geotechnical Journal 55 (6) (2018) 862–880.
[27] Y. Wang, P. Stoica, J. Li, T. L. Marzetta, Nonparametric spectral analysis with missing data via the em algorithm, Digital signal processing 15 (2) (2005) 191–206.
[28] P. Stoica, J. Li, J. Ling, Y. Cheng, Missing data recovery via a nonparametric iterative adaptive approach, in: 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2009, pp. 3369–3372.
[29] M. I. Knight, M. A. Nunes, G. P. Nason, Spectral estimation for locally stationary time series with missing observations, Statistics and Computing 22 (4) (2012) 877–895.
[30] M. Lepot, J.-B. Aubin, F. H. Clemens, Interpolation in time series: An introductive overview of existing methods, their performance criteria and uncertainty assessment, Water 9 (10) (2017) 796.
[31] M. Beer, P. D. Spanos, A neural network approach for simulating stationary stochastic processes, Structural engineering and mechanics: An international journal 32 (1) (2009) 71–94.
[32] Z. Che, S. Purushotham, K. Cho, D. Sontag, Y. Liu, Recurrent neural networks for multivariate time series with missing values, Scientific reports 8 (1) (2018) 1–12.
[33] L. Comerford, I. A. Kougioumtzoglou, M. Beer, On quantifying the uncertainty of stochastic process power spectrum estimates subject to missing data, International Journal of Sustainable Materials and Structural Systems 2 (1-2) (2015) 185–206.
[34] F. Tobar, Bayesian nonparametric spectral estimation, in: S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Vol. 31, Curran Associates, Inc., 2018.
URL https://proceedings.neurips.cc/paper/2018/file/abd1c782880cc59759f4112fda0b8f98-Paper.pdf
[35] 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.
[36] M. Behrendt, M. de Angelis, L. Comerford, Y. Zhang, M. Beer, Projecting interval uncertainty through the discrete fourier transform: An application to time signals with poor precision, Mechanical Systems and Signal Processing 172 (2022) 108920.
[37] M. Priestley, Power spectral analysis of non-stationary random processes, Journal of Sound and Vibration 6 (1) (1967) 86–97.
[38] P. D. Spanos, G. Failla, Evolutionary spectra estimation using wavelets, Journal of Engineering Mechanics 130 (8) (2004) 952–960.

[39] J. Liang, S. R. Chaudhuri, M. Shinozuka, Simulation of nonstationary stochastic processes by spectral representation, Journal of Engineering Mechanics 133 (6) (2007) 616–627.
[40] D. E. Newland, An introduction to random vibrations and spectral analysis, Longman Publishing Group, 1984.
[41] A. Graves, Generating sequences with recurrent neural networks, arXiv preprint arXiv:1308.0850 (2013).
[42] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using rnn encoder-decoder for statistical machine translation, arXiv preprint arXiv:1406.1078 (2014).
[43] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (8) (1997) 1735–1780.
[44] A. Graves, A.-r. Mohamed, G. Hinton, Speech recognition with deep recurrent neural networks, in: 2013 IEEE international conference on acoustics, speech and signal processing, Ieee, 2013, pp. 6645–6649.
[45] Y. Gal, Z. Ghahramani, A theoretically grounded application of dropout in recurrent neural networks, Advances in neural information processing systems 29 (2016).
[46] N. Srivastava, E. Mansimov, R. Salakhudinov, Unsupervised learning of video representations using lstms, in: F. Bach, D. Blei (Eds.), Proceedings of the 32nd International Conference on Machine Learning, Vol. 37 of Proceedings of Machine Learning Research, PMLR, Lille, France, 2015, pp. 843–852.
URL https://proceedings.mlr.press/v37/srivastava15.html
[47] K. P. Murphy, Machine Learning: A Probabilistic Perspective, MIT press, 2012.
[48] D. Salinas, V. Flunkert, J. Gasthaus, T. Januschowski, Deepar: Probabilistic forecasting with autoregressive recurrent networks, International Journal of Forecasting 36 (3) (2020) 1181–1191.
[49] C. Blundell, J. Cornebise, K. Kavukcuoglu, D. Wierstra, Weight uncertainty in neural network, in: International conference on machine learning, PMLR, 2015, pp. 1613–1622.
[50] Y. Gal, Z. Ghahramani, Dropout as a bayesian approximation: Representing model uncertainty in deep learning, in: international conference on machine learning, PMLR, 2016, pp. 1050–1059.
[51] A. Graves, Practical variational inference for neural networks, Advances in neural information processing systems 24 (2011).
[52] D. P. Kingma, M. Welling, Auto-encoding variational bayes, arXiv preprint arXiv:1312.6114 (2013).
[53] D. M. Blei, A. Kucukelbir, J. D. McAuliffe, Variational Inference: A Review for Statisticians, Journal of the American Statistical Association 112 (518) (2017) 859–877. arXiv:1601.00670, doi:10.1080/01621459.2017.1285773.
[54] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, The journal of machine learning research 15 (1) (2014) 1929–1958.
[55] L. Luzi, G. Lanzano, C. Felicetta, M. D'Amico, E. Russo, S. Sgobba, F. Pacor, O. W. G. 5, Engineering strong motion database (esm)(version 2.0), Istituto Nazionale di Geofisica e Vulcanologia (INGV). DOI: https://doi. org/10.13127/ESM 2 (2020).
[56] R. Narayana Iyengar, K. Sundara Raja Iyengar, A nonstationary random process model for earthquake accelerograms, Bulletin of the Seismological Society of America 59 (3) (1969) 1163–1188.
[57] G. Pousse, L. F. Bonilla, F. Cotton, L. Margerin, Nonstationary stochastic simulation of strong ground motion time histories including natural variability: Application to the k-net japanese database, Bulletin of the Seismological Society of America 96 (6) (2006) 2103–2117.
[58] F. Sabetta, A. Pugliese, G. Fiorentino, G. Lanzano, L. Luzi, Simulation of non-stationary stochastic ground motions based on recent italian earthquakes, Bulletin of Earthquake Engineering 19 (9) (2021) 3287–3315.
[59] D. M. Boore, Comparing stochastic point-source and finite-source ground-motion simulations: Smsim and exsim, Bulletin of the Seismological Society of America 99 (6) (2009) 3202–3216.
[60] B. Edwards, B. Zurek, E. Van Dedem, P. Stafford, S. Oates, J. Van Elk, B. DeMartin, J. Bommer, Simulations for the development of a ground motion model for induced seismicity in the groningen gas field, the netherlands, Bulletin of Earthquake Engineering 17 (2019) 4441–4456.
[61] J. Sunny, M. De Angelis, B. Edwards, Ranking and selection of earthquake ground-motion models using the stochastic area metric, Seismological Research Letters 93 (2A) (2022) 787–797.
[62] R. J. Little, D. B. Rubin, Statistical analysis with missing data, Vol. 793, John Wiley & Sons, 2019.
[63] I. A. Kougioumtzoglou, P. D. Spanos, An approximate approach for nonlinear system response determination under evolutionary stochastic excitation, Current science (2009) 1203–1211.
[64] S. Rezaeian, A. Der Kiureghian, A stochastic ground motion model with separable temporal and spectral nonstationarities, Earthquake Engineering & Structural Dynamics 37 (13) (2008) 1565–1584.
[65] F. Jalayer, J. Beck, Effects of two alternative representations of ground-motion uncertainty on probabilistic seismic demand assessment of structures, Earthquake engineering & structural dynamics 37 (1) (2008) 61–79.
[66] S. Rezaeian, N. Luco, Example applications of a stochastic ground motion simulation methodology in structural engineering, in: 15th World Conf. Earthquake Engineering,(WCEE), 2012.
[67] C. Vlachos, K. G. Papakonstantinou, G. Deodatis, Structural applications of a predictive stochastic ground motion model: Assessment and use, ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 4 (2) (2018) 04018006.
[68] E. Cazelles, A. Robert, F. Tobar, The wasserstein-fourier distance for stationary time series, IEEE Transactions on Signal Processing 69 (2020) 709–721.
[69] T. Pearce, A. Brintrup, M. Zaki, 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.