Chapter 14
A Generative Modeling Approach for the Translation of Operational Variables to Short-term Vibrations

Arthur Hatstatt and Konstantinos E. Tatsis

Abstract The vibration response assessment of engineering systems using the statistical features of vibration signals is a masked problem, which is limited by the sampling resolution of these statistics. One of the problems falling in this category is the vibration response assessment of wind turbine systems using SCADA data. Despite the fact that the resolution of the latter can be improved by slightly increasing the sampling rate of SCADA systems, which can deliver up to 1-min statistics, a connection is missing between SCADA data and short-term vibrations. Within this context, this paper proposes the use of Gaussian Process Latent Force (GPLF) models in a generative sense for the conversion of one-minute SCADA data to short-term vibration data. To this end, GPLF models are used for the modeling of the generalized modal responses of wind turbines, which are obtained on the basis of sparse acceleration measurements. The parameters of the GPLF models are conditioned on the operational variables of the system, which are available from SCADA data, and the identified model is used to generate short-term vibration signals with the aim of preserving their spectral and statistical properties. The proposed method is tested on a simulated case study and the performance is assessed in terms of the agreement between the generated time series and the ones measured from the simulated model.

Keywords Time series modeling · Gaussian process regression · Structural health monitoring · Latent force model · Generative modeling

Introduction

Vibration based methods are one of the most effective monitoring techniques to assess the structural integrity of engineering systems. Traditionally, vibration-based monitoring methods have relied on installing vibration sensors on critical components in order to capture high-resolution data that allows engineers to evaluate the dynamic behavior of the system under various conditions, offering insights into potential failures or anomalies in the operation. Among others, these methods allow to effectively capture damage accumulation and predict the remaining useful life of systems subjected to cyclic loading. However, there are oftentimes practical limitations to widespread adoption of vibration sensors in certain engineering systems due to installation and maintenance of sensors, which may increase costs and operational downtime. Within this context, various alternatives have been examined in the literature in order to leverage the availability of data from existing systems.

The state-of-the-art Structural Health Monitoring (SHM) methods for engineering systems have seen significant advancements through the use of data-driven approaches. These have been developed and employed either in combination with physics-based models [1–3] and reduced-order system representations [4, 5] for the identification and tracking of system changes or as completely data-driven methods, such as Convolutional Neural Networks (CNNs) [6, 7] for the detection faults using multi-source data or Long Short-Term Memory (LSTM) [8, 9] networks for the identification of damage in time series. Despite these advancements, challenges remain in real-world implementation, particularly in acquiring large, high-quality datasets and accounting for environmental factors [10].

Arthur Hatstatt
Chair of Risk, Safety and Uncertainty Quantification, ETH Zürich, Zürich, Switzerland
e-mail: michael.mcgurk@strath.ac.uk

Konstantinos E. Tatsis
Swiss Data Science Center (SDSC), ETH Zürich, Zürich, Switzerland
e-mail: j.yuan@soton.ac.uk

Brian Damiano et al. (eds.), *Structural Health Monitoring & Machine Learning, Vol. 12*, Conference Proceedings of the Society for Experimental Mechanics Series, https://doi.org/10.13052/97887-438-0157-3_14

This contribution is motivated by wind turbine structures, which are oftentimes difficult to instrument and the availability of Supervisory Control and Data Acquisition (SCADA) data has not been leveraged for vibration-based assessment. As such, the methodology proposed in this paper aims at bridging the gap between statistical features of vibration signals, such as the ones offered by SCADA systems, for the generation of vibration signals that can stochastically reproduce the response experienced by the system under various conditions of operation. To this end, a physics-informed Gaussian Process Latent Force (GPLF) [11, 12] is developed for the modeling of the modal response signals, which can be used to compose the entire system response. The model is constructed by leveraging the knowledge of the system structural properties and the features characterising the vibration response signals, such as the energy due to operational harmonics. The aim of the proposed surrogate approach is to served as a generative model that can produce time histories of the vibration response, such that their statistical and spectral information matches the one of the original signals.

This paper is organized into three main sections. The first section outlines the methodology, focusing on the Gaussian Process Latent Force Model as the surrogate model, along with its formulation. The second section presents the application case, detailing the framework used to study the vibrational data of a wind turbine and the analytical approach taken. Finally, the third section showcases the results, offering a comparative analysis of the predictive capabilities of the surrogate model against the ground truth data obtained from simulations.

Methodology

In this section, the formulation of Gaussian Process Latent Force models is described for the representation of vibration response signals obtained from structural or mechanical systems. To this end, the response of the system is decomposed into the modal contributions, whose spectral content is easier to characterise and subsequently represented by a temporal Gaussian Process, whose dynamics are driven by a set of latent forces.

Structural dynamics

Consider an n-degree-of-freedom linear time-invariant dynamic system whose vibration response is governed by the following continuous-time differential equation

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{C}\dot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u}(t) = \mathbf{S}_{\mathrm{p}}\mathbf{p}(t) \tag{1}$$

where $\mathbf{u} \in \mathbb{R}^n$ is the vector of displacements, $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{K} \in \mathbb{R}^{n\times n}$ denote the mass, stiffness and damping matrices respectively, $\mathbf{p}(t) \in \mathbb{R}^{n_{\mathrm{p}}}$ is the vector of external forcing terms and $\mathbf{S}_{\mathrm{p}} \in \mathbb{R}^{n\times n_{\mathrm{p}}}$ is a selection matrix that assigns the forcing terms to their corresponding locations. In engineering systems, the dynamics are typically governed by a few vibration modes and as such, the response can be expressed as follows

$$\mathbf{u}(t) = \mathbf{\Phi}\mathbf{q}(t) \tag{2}$$

where $\mathbf{\Phi} \in \mathbb{R}^{n\times n_{\mathrm{m}}}$ is a matrix whose columns contain the mode shapes and $\mathbf{q}(t) \in \mathbb{R}^{n_{\mathrm{m}}}$ denotes the vector of generalized modal coordinates. In view of this expression and under the condition of proportional damping, the initial equation of motion, described by Eq. (1), can be decomposed into m second order differential equations. Therefore, the dynamics of the i-th generalized coordinate $q_i(t)$ are governed by the following equation

$$\ddot{q}_i(t) + 2\xi_i\omega_i\dot{q}_i(t) + \omega_i^2 q_i(t) = f_i(t) \tag{3}$$

which applies for $i = 1, 2, \ldots, n_{\mathrm{m}}$. Moreover, ω_i denotes the natural frequency of the i-th mode, ξ_i the corresponding damping ratio and $f_i(t)$ is the projection of the external loads to the modal space, which is obtained by the following expression

$$f_i(t) = \boldsymbol{\phi}_i^{\mathrm{T}}\mathbf{S}_{\mathrm{p}}\mathbf{p}(t) \tag{4}$$

in which $\boldsymbol{\phi}_i$ denotes the i-th column of the mode shape matrix.

Modal expansion

Under the availability of vibration measurements from a number of sensing points, which are herein denoted by $\mathbf{y}(t) \in \mathbb{R}^{n_{\mathrm{s}}}$, the projection postulated by Eq. (2) can be written selectively for all sensing channels as follows

$$\mathbf{y}(t) = \sum_{i=1}^{n_{\mathrm{m}}} \boldsymbol{\phi}_{\mathrm{s},i}\, q_i(t) \tag{5}$$

where $\boldsymbol{\phi}_{\mathrm{s},i} \in \mathbb{R}^{n_\mathrm{s} \times 1}$ denotes the mode shape values of the i-th vibration mode at the locations where measurements are available.

Given that the number of measurements n_s is equal to or greater than the number of vibration modes n_m that are taken into account for the representation of the system dynamics, the time history of the generalized modal coordinates $q_i(t)$ for $i = 1, 2, \ldots, n_\mathrm{m}$ can be retrieved by solving the inverse problem

$$\mathbf{q}(t) = \boldsymbol{\Phi}_\mathrm{s}^{\dagger} \, \mathbf{y}(t) \tag{6}$$

where $\boldsymbol{\Phi}_\mathrm{s} = \begin{bmatrix} \boldsymbol{\phi}_{\mathrm{s},1} & \ldots & \boldsymbol{\phi}_{\mathrm{s},m} \end{bmatrix} \in \mathbb{R}^{n_\mathrm{s} \times n_\mathrm{m}}$ is the matrix of mode shapes at the sensing points and $\boldsymbol{\Phi}_\mathrm{s}^{\dagger}$ denotes the generalized inverse of $\boldsymbol{\Phi}_\mathrm{s}$, which is in this work taken equal to the Moore-Penrose pseudo-inverse $\boldsymbol{\Phi}_\mathrm{s}^{\dagger} = (\boldsymbol{\Phi}_\mathrm{s}^\mathrm{T} \boldsymbol{\Phi}_\mathrm{s})^{-1} \boldsymbol{\Phi}_\mathrm{s}^\mathrm{T}$. It should be noted that Eq. (5) implies the availability of displacement measurements however, a similar expression would apply when acceleration, velocity or strain signals are measured. In this contribution, it will be assumed that the vibration response measurements are obtained from acceleration sensors and therefore Eq. (6) delivers the time history of modal accelerations, which are to be modeled by means of GPLF models.

Gaussian process regression

A temporal Gaussian Process (GP) $h(t) \sim GP(0, \kappa(\tau))$ with stationary covariance function $\kappa(t, t') = \kappa(\tau)$, where $\tau = t - t'$, can be obtained as the output of a linear and constant in parameters differential equation, whose dynamics are driven by white noise $w(t)$ with spectral density $S_w(\omega) = \sigma_\omega$. This results in the following signal representation

$$\frac{d^m h(t)}{dt^m} + a_{m-1} \frac{d^{m-1} h(t)}{dt^{m-1}} + \ldots + a_0 h(t) = w(t) \tag{7}$$

where $a_0, \ldots, a_{m-1}$ are known parameters. The equivalent of Eq. (7) in a compact state-space form is obtained as follows

$$\dot{\mathbf{z}}(t) = \mathbf{F}_\mathrm{c} \, \mathbf{z}(t) + \mathbf{L}_\mathrm{c} w(t) \tag{8a}$$

$$h(t) = \mathbf{H}_\mathrm{c} \, \mathbf{z}(t) + v(t) \tag{8b}$$

where $\mathbf{z} = \begin{bmatrix} h(t) \; \frac{dh(t)}{dt} \; \ldots \; \frac{d^{m-1}h(t)}{dt^{m-1}} \end{bmatrix}^\mathrm{T} \in \mathbb{R}^m$ denotes the state vector and $v(t) \in \mathbb{R}$ is the measurement noise, which is assumed to be white Gaussian with variance σ_v. The state matrices are functions of the differential equation parameters and are given by the following expressions

$$\mathbf{F}_\mathrm{c} = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \\ -a_0 & -a_1 & -a_2 & \ldots & -a_{m-1} \end{bmatrix}, \quad \mathbf{L}_\mathrm{c} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \tag{9}$$

while $\mathbf{H}_\mathrm{c} \in \mathbb{R}^{1 \times m}$ is a boolean matrix selecting the signal $h(t)$ from the state vector. The continuous-time equations can be transferred to the discrete time domain using a time step $\Delta t = 1/f_\mathrm{s}$, where f_s denotes the sampling frequency, and converted to the following algebraic equations

$$\mathbf{z}_{k+1} = \mathbf{F}_\mathrm{d} \mathbf{z}_k + \mathbf{L}_\mathrm{d} w_k \tag{10a}$$

$$h_{k+1} = \mathbf{H}_\mathrm{d} \mathbf{z}_k + v_k \tag{10b}$$

where the form of the matrices depends on the assumption regarding the variation of the input. Namely, for a zero-order hold assumption, which implies a constant value of the input between two successive steps, the discrete-time state matrices, which are denoted by the subscript $\square_\mathrm{d}$ are $\mathbf{F}_\mathrm{d} = e^{\mathbf{F}_\mathrm{c} \Delta t}$ and $\mathbf{L}_\mathrm{d} = (\mathbf{F}_\mathrm{d} - \mathbf{I}) \mathbf{F}_\mathrm{c}^{-1} \mathbf{L}_\mathrm{c}$, while the output matrix $\mathbf{H}_\mathrm{d}$ is identical to $\mathbf{H}_\mathrm{c}$. The spectral density of the signal $h(t)$ can be obtained analytically from the state-space matrices and is given by following equation

$$\begin{aligned} S_\mathrm{h}(\omega) &= H(i\omega) \sigma_\mathrm{w} H(-i\omega) \\ &= \mathbf{H}_\mathrm{c} (\mathbf{F}_\mathrm{c} + i\omega \mathbf{I})^{-1} \mathbf{L}_\mathrm{c} \, \sigma_w \, \mathbf{L}_\mathrm{c}^\mathrm{T} (\mathbf{F}_\mathrm{c} + i\omega \mathbf{I})^{-\mathrm{T}} \mathbf{H}_\mathrm{c}^\mathrm{T} \end{aligned} \tag{11}$$

where $H(i\omega)$ is the transfer function and i denotes the imaginary unit. It should be noted that when the measurement noise $v(t)$ is additionally considered, the expression of the spectral density of $h(t)$ described above should be slightly adjusted. Due to the fact that the noise term has only an additive effect to the signal $h(t)$, the spectral density of the latter should be added to the one of the measurement noise.

Due to the relation between the spectral density and the covariance function, which under stationary conditions are connected through the Fourier transform, the spectral characteristics of the signal $h(t)$ can be inversely defined by specifying the covariance kernel of the signal, which results in a specific spectral content $S_{\mathrm{h}}(\omega)$. The latter in turn can be converted to a state-space representation, according to the factorization postulated by Eq. (11). It is shown in [11] that such a conversion is possible, when the spectral density of the signal is amenable to a rational form.

Covariance functions

Various classes of covariance kernels exist, each of which can attribute different features to the modeled signal. This work is based on the use of Matern and periodic kernels, whose spectral density has an analytical description and their corresponding state-space matrices are presented below.

Matern kernel

The family of Matern kernels are used to describe stationary covariance functions for different types of random processes, while having only a few tunable hyperparameters. One of the most typically employed kernels belonging to this family is the one for $\nu = 3/2$, whose parameters are l and α. The state-space matrices in this case are obtained as

$$\mathbf{F}_{\mathrm{c}} = \begin{bmatrix} 0 & 1 \\ \lambda^2 & -2\lambda \end{bmatrix}, \quad \mathbf{L}_{\mathrm{c}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{H}_{\mathrm{c}} = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

with $\lambda = \frac{\sqrt{3}}{l}$ and $\sigma_w = 12\sqrt{3}\frac{\alpha^2}{l^3}$. Accordingly, the Matern kernel with $\nu = 5/2$, in which case it is again parametrized with respect to l and α, corresponds to the following state-space matrices

$$\mathbf{F}_{\mathrm{c}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -\lambda^3 & -3\lambda^2 & -3\lambda \end{bmatrix}, \quad \mathbf{L}_{\mathrm{c}} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{H}_{\mathrm{c}} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

where $\lambda = \frac{\sqrt{5}}{l}$ and $\sigma_w = 400\sqrt{5}\frac{\alpha^2}{3l^5}$.

Periodic kernel

The periodic kernel enables the modeling of signals which repeat themselves exactly, such as the harmonic signals induced by rotating machines. This type of kernel is entirely deterministic, which implies that the response in time domain does not depend on the noise term. The state-space matrices of a signal whose covariance function is described by a periodic kernel are

$$\mathbf{F}_{\mathrm{c}} = \begin{bmatrix} 0 & -\omega \\ \omega & 0 \end{bmatrix}, \quad \mathbf{L}_{\mathrm{c}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{H}_{\mathrm{c}} = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

where ω denotes the eigenfrequency, while $\sigma_w = 0$.

Harmonic oscillator

An extension of the periodic kernel in terms of spectral features is offered by the kernel that represents a harmonic oscillator. This represented by the following matrices

$$\mathbf{F}_{\mathrm{c}} = \begin{bmatrix} 0 & 1 \\ -\omega^2 & -2\xi\omega \end{bmatrix}, \quad \mathbf{L}_{\mathrm{c}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{H}_{\mathrm{c}} = \begin{bmatrix} -\omega^2 & -2\xi\omega \end{bmatrix}$$

where ω denotes the eigenfrequency and ξ is the corresponding damping ratio. The dynamics of such a model are driven by a white noise signal with spectrum σ_w.

Gaussian process latent force model

In this section, the GP model is extended in order to accommodate a more expressive structure, which is capable of representing the spectral and statistical characteristics of vibration response signals. Within this context, the aim is to construct a GP model, whose dynamics are driven by a set of latent forces, each of which is modeled as a GP and delivers certain features to the target signal $y(t)$. This is achieved by combining the contributions of latent forces which are characterised by the covariance functions presented in the previous Subsection. The model formulation is described below, where the knowledge of the system dynamics can be incorporated into a GP model that follows the covariance function of a harmonic oscillator. This GP is used to model the dynamics of the decomposed modal signals and is augmented with a number of latent variables $\mathbf{z}_k(t)$, for $k = 1, 2, \ldots, K$. The state-space representation of the augmented system in continuous-time is obtained by defining the state vector $\mathbf{z}^{\mathrm{a}}(t) = \begin{bmatrix} u(t) & \dot{u}(t) & \mathbf{z}_1^{\mathrm{T}}(t) & \ldots & \mathbf{z}_K^{\mathrm{T}}(t) \end{bmatrix}^{\mathrm{T}}$, which contains the state variables that describe the underlying dynamics of the signal, as well as the state variables of the various latent forcing terms $\mathbf{z}_k(t)$.

$$\dot{\mathbf{z}}^{\mathrm{a}}(t) = \mathbf{F}_{\mathrm{c,a}}\mathbf{z}^{\mathrm{a}}(t) + \mathbf{L}_{\mathrm{c,a}}\mathbf{w}^{\mathrm{a}}(t) \tag{12a}$$

$$y(t) = \mathbf{H}_{\mathrm{c,a}}\mathbf{z}^{\mathrm{a}}(t) + v(t) \tag{12b}$$

where $\mathbf{w}^{\mathrm{a}}(t)$ denotes the augmented process noise, consisting of the stacked noise terms for all latent terms. Each one of these noise variables is assumed to be Gaussian and white with spectral density $\sigma_{w,k}$, for $k = 1, 2, \ldots, K$. Lastly, $v(t)$ is the measurement noise, which is also a white process of variance σ_v. The state matrices of the augmented model, which consists of K latent variables are

$$\mathbf{F}_{\mathrm{c,a}} = \begin{bmatrix} \mathbf{A}_{\mathrm{c}} & \mathbf{H}^*_{\mathrm{c},1} & \mathbf{H}^*_{\mathrm{c},2} & \ldots & \mathbf{H}^*_{\mathrm{c},K} \\ \mathbf{0} & \mathbf{F}_{\mathrm{c},1} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{F}_{\mathrm{c},2} & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{F}_{\mathrm{c},K} \end{bmatrix}, \; \mathbf{L}_{\mathrm{c,a}} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{\mathrm{c},1} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{L}_{\mathrm{c},2} & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{L}_{\mathrm{c},K} \end{bmatrix} \tag{13}$$

where $\mathbf{H}^*_{\mathrm{c},k} = \begin{bmatrix} \mathbf{0} & \mathbf{H}^{\mathrm{T}}_{\mathrm{c},k} \end{bmatrix}^{\mathrm{T}}$, while the output matrix takes the following form

$$\mathbf{H}_{\mathrm{c,a}} = \begin{bmatrix} -\omega^2 & -2\xi\omega & \mathbf{H}_{\mathrm{c},1} & \ldots & \mathbf{H}_{\mathrm{c},K} \end{bmatrix} \tag{14}$$

which without loss of generality implies the modeling of acceleration signals due to the presence of the harmonic oscillator parameters in the first two entries. These can be adjusted accordingly for the modeling of displacement, velocity or strain signals. Lastly, the state-space representation of the augmented system can be transferred to the discrete-time domain by means of a zero-order hold scheme, similar to the one applied in Section .

Hyperparameter identification

The identification of the hyperprameters can be seen as an optimization problem, whose solution can be obtained by means of the Maximum Likelihood (ML) estimate. In this contribution, a least-squares estimate of the hyperparameters is adopted, which relies on the analytical expression of the spectral density as defined from the Gaussian Process Latent Force model. Similar to Eq. (11), the analytical form of the spectral density is derived using the augmented system matrices, which take into account all different latent forces

$$S_{\mathrm{h}}(\omega) = \mathbf{H}_{\mathrm{c,a}}(\mathbf{F}_{\mathrm{c,a}} + i\omega\mathbf{I})^{-1}\mathbf{L}_{\mathrm{c,a}}\,\mathbf{\Sigma}_w\,\mathbf{L}^{\mathrm{T}}_{\mathrm{c,a}}(\mathbf{F}_{\mathrm{c,a}} + i\omega\mathbf{I})^{-\mathrm{T}}\mathbf{H}^{\mathrm{T}}_{\mathrm{c,a}} \tag{15}$$

where $\mathbf{\Sigma}_w \in \mathbb{R}^{K\times K}$ is the diagonal covariance matrix that contains the variance terms σ_w^k of the noise components that act on the signal $h(t)$ as latent forces and as such, it is the quantity to be identified.

The expression of the spectral density described by Eq. (15) can be reformulated in a vector form with respect to the unknown variance parameters in order to obtain the values using a least squares solution of the form $\tilde{\mathbf{A}}\hat{\boldsymbol{\sigma}}_w = \tilde{\mathbf{b}}$. This is accomplished by evaluating Eq. (15) for a number of sampling frequencies ω_l, for $l = 1, 2, \ldots, L$, thus resulting to a set of L equations for the calculation of the unknown variance terms. The least squares solution is built according to the following equations

$$\tilde{\mathbf{A}}_l = \mathbf{G}_l \otimes \mathbf{G}_l \tag{16a}$$

$$\mathbf{G}_l = \mathbf{H}_{\mathrm{c,a}}(\mathbf{F}_{\mathrm{c,a}} + i\omega_l\mathbf{I})^{-1}\mathbf{L}_{\mathrm{c,a}} \tag{16b}$$

$$\tilde{\mathbf{b}}_l = S_{\mathrm{h,o}}(\omega_l) \tag{16c}$$

where $\tilde{\mathbf{A}}_l$ denotes the l-th row of matrix $\tilde{\mathbf{A}}$ and accordingly, $\tilde{\mathbf{b}}_l$ is the l-th entry of $\tilde{\mathbf{b}}$. Therefore, the vector $\boldsymbol{\sigma}_{\mathrm{w}}$ of unknown variance terms is computed using the generalized inverse of $\tilde{\mathbf{A}}$.

Application

The methodology presented in Section is herein applied for the modeling of vibration response signals obtained from wind turbine systems. To this end, the NREL 5MW onshore wind turbine [13] is simulated using FAST software for various operational and environmental conditions, which are defined by the mean wind speed and the turbulence intensity. For each sample of these conditional variables, a 10 minute period is simulated and the acceleration signals in both x and y directions are extracted from two positions of the tower. The signals are sampled at 50 Hz and are subsequently truncated to 1-minute intervals and transformed into the Fore-Aft (FA) and Side-to-Side (SS) directions. The corresponding SCADA variables are also extracted for the 1-minute intervals. Thereafter, a GPLF model is learned in order to be used as generative model that can translate SCADA variables into 1-min time series, whose spectral characteristics are in accordance with those obtained from simulations. The aim of the generated time series is to emulate the original signal, not only in terms of its spectral content and statistical properties, but also in terms of the loading cycles, which are responsible for the accumulation of fatigue damage. A schematic overview of the workflow followed in this case study is presented in Fig. 1

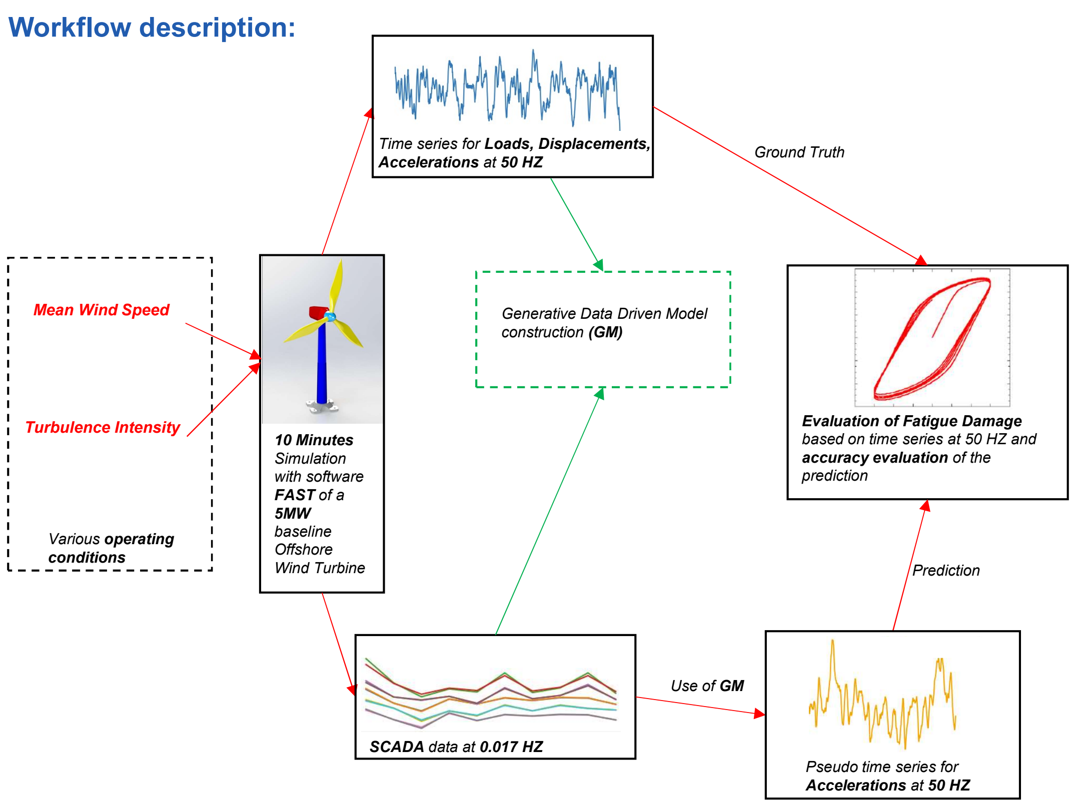

Fig. 1 General workflow description

As underlined in the section of the methodology, the optimization algorithm implies the knowledge of certain hyperparameters, which can be extracted from statistical metric of the signal. In this specific case study, it is assumed that certain characteristics of the response are available from the design, such as the eigenfrequencies and the damping of the system, while some further information can be extracted from the available SCADA data. As such, the dynamics of the acceleration signals of the two modes in each direction, namely FA and SS, are assumed to be governed by a harmonic oscillator kernel, with ω_m and ξ_m being the eigenfrequency and damping ratio of the m-th mode. The dynamics of the modelled signals are driven by a number of latent forces, each of which represents a specific feature of the response. Theses latent forces are mainly driven by the rotation of the wind turbine, and should therefore have frequency components as follows:

- Rotor frequency: $\omega_r = \frac{2\pi}{60} \cdot$ RPM rad/s
- Blade passage frequencies: $\omega_b^j = 3 \cdot j \cdot \omega_r,\ j \in \mathbb{N}$.

In order to account for the variability of the response, which is induced by the change in environmental and operational conditions, the identified hyperparameters are conditioned on the SCADA variables, which apart from the rotor speed (RPM), contain the following channels:

- **Wind velocity**: The absolute value of wind velocity is computed as the norm of the velocity vector.
- **Turbulence intensity**: The turbulence intensity used for the corresponding simulation.
- **Yaw misalignment**: The misalignment of the yaw position with respect to the wind direction, computed from the difference between the velocity vector and the yaw position angle.

Within this context, a mapping between the aforementioned SCADA variables and the hyperparameters of the GPLF model is identified, which is herein implemented using a Multi-Layer Perceptron (MLP) model, in order to extend the validity of the model to the entire operational spectrum.

Results

In this section the results from the simulated case study, which is described in Section , are presented. Upon simulation of the wind turbine system for different operational conditions, and decomposition of the acceleration response into modal signals in Fore-Aft and Side-to-Side directions, the GPLF model is fitted for each 1-min time series and the identified model is used as a generative surrogate, which aims at preserving the spectral and statistical information of the simulated response.

The performance of the model is assessed in both directions, namely Fore-Aft and Side-to-Side, of the wind turbine and for the first two modes of the structure, which typically contain most of the vibration energy. The comparison is presented in terms of the spectra between the generated and the original time series. In order to avoid having too many hyperparameters for the surrogate model, the damping ratios are considered equal for all cases, namely $\xi^k = 0.01$ for all latent forces. Therefore, the only hyperparameters identified by the algorithm are the measurement noise level σ_v and the spectral densities of the noise driving the latent forces σ_w^k, for $k = 1, 2, \ldots, K$.

The results of the time series generated by the identified models are shown in Figs. 2-3. Concretely, Fig. 2 depicts the spectra and the time series between the original signal and the one generated by the GPLF models for the first FA mode. The corresponding comparison of the signals in SS direction is presented in Fig. 3. For both directions, the generated pseudo time series seem to encapsulate the dynamics of the observations quite well, with a very limited number of hyperparameters. Indeed, as depicted by the power spectra, for both modes and both directions, most of the energy of the signal is concentrated in the frequency range between 0 and 5Hz. Moreover, the energy peaks are concentrated around the physically known

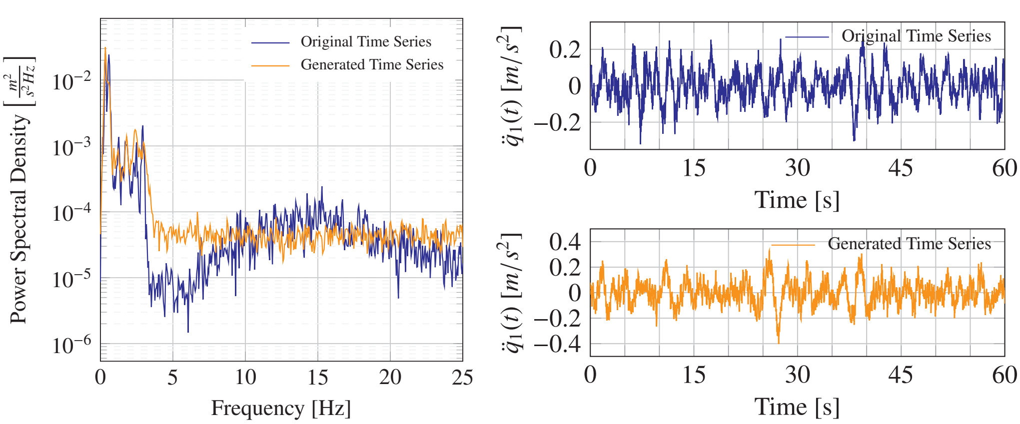

Fig. 2 Comparison of the spectra (left) between the original and the generated time series (right) of the first mode in FA direction

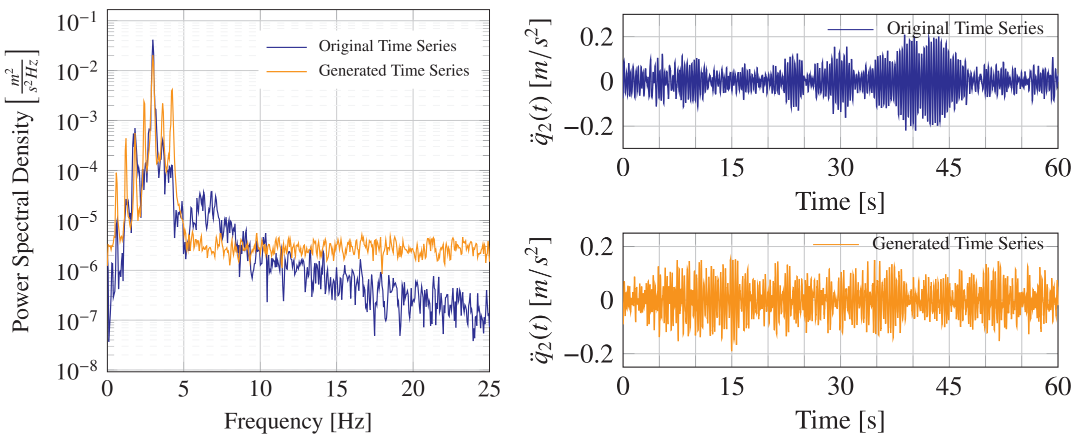

Fig. 3 Comparison of the spectra (left) between the original and the generate time series (right) of the second mode in FA direction

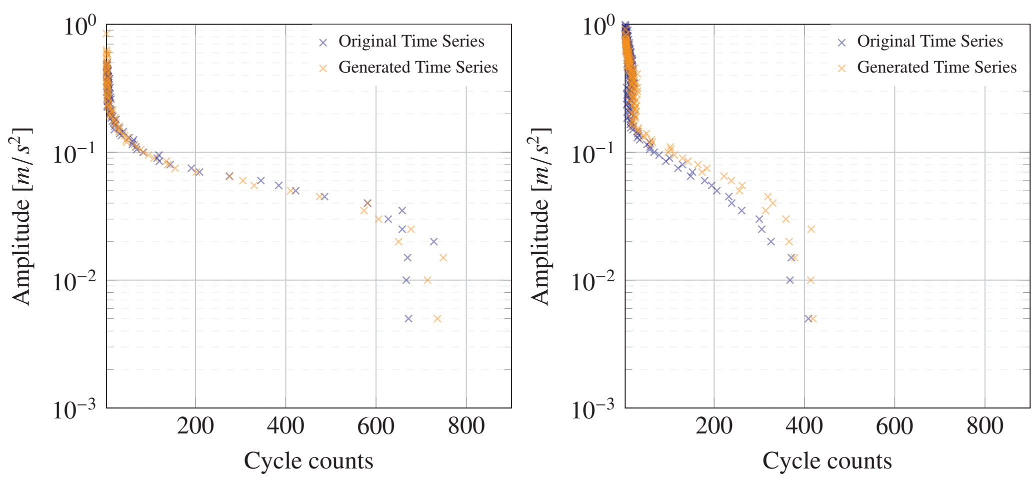

Fig. 4 Comparison of the modal strain cycle counts for the first (left) and second (right) mode in the FA direction

frequencies (eigenfrequencies of the system and multiples of rotational frequency). These two particular characteristics of the physical problem allow the surrogate model to replicate the real-world complex dynamics in a satisfactory manner with a very limited number of hyperparameters. One could argue that more complex models could replicate the dynamics in a better way, which is undeniably the case, however this would overlook the final goal of the surrogate.

The final target being to find a mapping between scarce SCADA data and the hyperparameters of the surrogate model, simpler models are preferred, as fewer parameters usually exhibit higher interpretability and predictabilty. Moreover, the presented approach allows to find analytically the hyperparameters of the model, avoiding the need for optimization algorithms, which, based on the implemented cases, seem to induce a lot more variability in the identified parameters.

Lastly, the quantity of interest is neither the spectra of the obtained signals nor the pseudo time series directly, but their predictive capabilities for structural analysis. A very critical example is the quantification of fatigue accumulation, and the applicability of the generated pseudo time series for this purpose is investigated in the remaining of this section. Fatigue damage is related to the strain cycles history in the tower basis. Following the methodology proposed in [14], the strain history is retrieved from the modal contributions $q(t)$, by initially integrating the modal accelerations in the Laplace domain [15] and subsequently transforming the model displacements to strains through the deformation matrix of the Finite Element model of the system. Once the strain time series are obtained, the number of cycles for each strain level is counted, which is herein performed using the rainflow counting algorithm. The obtained results are presented in Figs. 4, where a satisfactory agreement between the generated and the original time series is achieved.

Conclusions

In this contribution, a generative modeling approach is proposed for the translation of vibration signal statistics to short terms dynamics. The methodology is demonstrated on a simulated wind turbine system, where a Gaussian Process Latent Force (GPLF) model is used to map one-minute SCADA data to the GPLF parameters. With fatigue damage accumulation being a critical phenomenon in wind turbine systems, the generated pseudo time series have a satisfactory predictive capability for the quantification of loading cycles at different amplitudes. The relative simplicity of the surrogate model allows still to have a very decent accuracy compared to a high complexity dynamic model, requiring a lot of computational power and parameters. Moreover, another advantage of the proposed methodology is the connection between the scarce SCADA variables and the short-term dynamics. The major advantage of such a methodology is the generative approach that can be used to estimate the fatigue accumulation in critical parts of the structure based on a relatively general and simple implementation, avoiding the pitfall of data availability encountered by many purely data-driven models.

References

1. Agathos, K., Tatsis, K., Nicoli, S., Bordas, S.P., and Chatzi, E. "Crack detection in mindlin-reissner plates under dynamic loads based on fusion of data and models". *Computers & Structures*, 246:106475 (2021)
2. Gres, S., Döhler, M., Andersen, P., and Mevel, L. "Subspace-based mahalanobis damage detection robust to changes in excitation covariance". *Structural Control and Health Monitoring*, 28(8):e2760 (2021)
3. Tatsis, K.E., Ou, Y., Dertimanis, V.K., Spiridonakos, M.D., and Chatzi, E.N. "Vibration-based monitoring of a small-scale wind turbine blade under varying climate and operational conditions. part ii: A numerical benchmark". *Structural Control and Health Monitoring*, 28(10):e2813 (2021)
4. Tatsis, K., Lourens, E., and Chatzi, E. "A general substructure-based framework for input-state estimation using limited output measurements". *Mechanical Systems and Signal Processing*, 150:107223 (2021)
5. Tatsis, K.E., Agathos, K., and Chatzi, E.N. "A hierarchical output-only bayesian approach for online vibration-based crack detection using parametric reduced-order models". *Mechanical Systems and Signal Processing*, 167:108558 (2022)
6. Zhao, Y., Zhang, Y., Li, Z., Bu, L., and Han, S. "Ai-enabled and multimodal data driven smart health monitoring of wind power systems: A case study". *Advanced Engineering Informatics*, 56:102018 (2023)
7. Gopalakrishnan, S., Laflamme, S., and Hu, C. "Deep cnn-based learning for automated structural damage detection and localization". *Smart Materials and Structures*, 26(5):055027 (2017)
8. Malekjafarian, A. and O'Brien, E.J. "Structural damage detection using long short-term memory networks". *Journal of Civil Structural Health Monitoring*, 11:765–779 (2021)
9. Choe, D.E., Kim, H.C., and Kim, M.H. "Sequence-based modeling of deep learning with lstm and gru networks for structural damage detection of floating offshore wind turbine blades". *Renewable Energy*, 174:218–235 (2021)
10. Azimi, M., Eslamlou, A.D., and Pekcan, G. "Data-driven structural health monitoring and damage detection through deep learning: State-of-the-art review". *Sensors*, 20 (2020)
11. Sarkka, S., Solin, A., and Hartikainen, J. "Spatiotemporal learning via infinite-dimensional bayesian filtering and smoothing: A look at gaussian process regression through kalman filtering". *IEEE Signal Processing Magazine*, 30(4):51 – 61 (2013)
12. Álvarez, M., Luengo, D., and Lawrence, N.D. "Latent force models". *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, pages 9–16 (2009)
13. J.M. Jonkman, M.B.J. *FAST User's Guide*. National Renewable Energy Laboratory (2005)
14. Mozafari, V.P.R.J., S. and Dykes, K. "Sensitivity of fatigue reliability in wind turbines: effects of design turbulence and the wöhler exponent". *Wind Energ. Sci.*, pages 799–820 (2024)
15. Iliopoulos, A., Weijtjens, W., Van Hemelrijck, D., and Devriendt, C. "Fatigue assessment of offshore wind turbines on monopile foundations using multi-band modal expansion: Fatigue assessment of monopile owts using multi-band modal expansion". *Wind Energy*, 20 (2017)