Gaussian Process Time-Series Models for Structures under Operational Variability

Luis David Avendaño-Valencia[1]*, Eleni N. Chatzi[1], Ki Young Koo[2] and James M. W. Brownjohn[2]

[1]*Department of Civil, Environmental and Geomatic Engineering, Institute for Structural Mechanics, ETH Zürich, Zürich, Switzerland,* [2]*College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, United Kingdom*

A wide range of vibrating structures are characterized by variable structural dynamics resulting from changes in environmental and operational conditions, posing challenges in their identification and associated condition assessment. To tackle this issue, the present contribution introduces a stochastic modeling methodology *via* Gaussian Process (GP) time-series models. In the presently introduced approach, the vibration response is represented by means of a random coefficient time-series model, whose coefficients comply with a GP regression on the environmental and operational parameters. The approach may be implemented in conjunction to any type of linear-in-the-parameters time-series model, ranging from simple AR models to more complex non-linear or non-stationary time-series models. The obtained GP time-series modeling approach provides an effective and compact global representation of the vibrational response of a structure under a wide span of environmental and operational conditions. The effectiveness of the postulated GP time-series models is demonstrated through two case studies: the first involves the identification of the vertical vibration response of the Humber bridge, evaluated over a period of three years; the second considers the long-term simulated vibration response of a wind turbine featuring non-stationary dynamics stemming from the rotor speed. In both cases, the variation of the average wind speed is the main driver of uncertainty, while, through application of the proposed GP time-series models, it is possible to track the resulting variation in modal quantities.

Keywords: time-series models, uncertainty, metamodels, random coefficient, gaussian process

1. INTRODUCTION

Several types of vibrating structures by default operate in constantly varying environmental and operational conditions, which inevitably results in variability of the induced structural dynamics. This is the case for wind turbines, bridges, high rise buildings and more. This issue poses a practical challenge related to the identification and analysis of the vibrational response of these structures, as well as for the health monitoring, fatigue assessment and control of the induced vibrations.

In order to construct a robust model of the dynamic response of the structure, it is not only necessary to accurately model the *short-term* response of the structure, but it is further necessary to effectively capture the *long-term* trends underlying the induced dynamics. This issue, in the particular case of *data-based time-series models*, has been extensively researched in recent years, resulting in the formulation of different strategies, including *projection methods*, and

deterministic or *stochastic* functional dependence models. *Projection methods*, also referred to as *data normalization methods*, aim at projecting characteristic quantities associated with the time-series model representing the short-term response of the structure, into a subspace where the influence of *Environmental and Operational Parameters* (EOPs) may be easily removed (Yan et al., 2005; Sohn, 2007; Deraemaeker et al., 2008). On the other hand, *deterministic functional dependence models* aim at capturing the long-term variability in the dynamics by assuming a deterministic functional relationship from EOPs to the characteristic quantities of the time-series model describing the dynamics of the response. Typically, such a deterministic functional relationship is captured *via* a functional series expansion. Methods falling into this class include the regression/interpolation methods discussed in Worden et al. (2002) and Sohn (2007), as well as the *Functionally Pooled* (FP) time-series models explained in Kopsaftopoulos et al. (2018) and Sakellariou and Fassois (2016). These methods are particularly effective when a direct relationship exists between measurable input EOPs and the characteristic quantities of the time-series models. Nonetheless, uncertainty in the EOPs and in the dynamic response introduces variability in the characteristic quantities of the basic time-series model, which may not be effectively captured by means of a deterministic relationship. Instead, random or stochastic functions may be more effective in capturing the uncertainty on the characteristic quantities of the time-series model.

In this sense, a third class of methods, referred to as *stochastic functional dependence models*, aim at capturing the long-term variability by assuming that the characteristic quantities of the basic time-series model are stochastic variables depending on the EOPs. Toward this end, recent works have postulated either *Random Coefficient* (RC) (Avendaño-Valencia and Fassois, 2015, 2017a,b; Avendaño-Valencia et al., 2015a) or *Polynomial Chaos Expansions* (PCE) (Spiridonakos and Chatzi, 2014, 2015; Avendaño-Valencia et al., 2015b; Spiridonakos et al., 2016) to represent the variability of the characteristic quantities of a model. In particular, RC time-series models represent the variability in the dynamics as randomness in the parameters of the time-series model. Then, apart from the selection of the specific time-series model, a further user-defined choice pertains to the definition of an appropriate distribution model for its respective coefficients, which in several cases can become very complex. On the other hand the PCE approach exploits the probabilistic knowledge of the EOPs to build the most effective functional representation of the time-series parameters. However, it is also considered that the randomness in the model parameters originates solely on the randomness of the EOPs. Therefore, other sources of uncertainty may be misrepresented.

In this regard, this work provides a framework for the global (short and long term) identification of the dynamic response of a structure, of unknown properties or a given *a priori* numerical model, under variable operational and environmental conditions by representing the short-term dynamics *via* a linear-in-the-parameters regressive time-series model (which may assume the form of an AutoRegressive, AutoRegressive with eXogenous input or similar model), and a *Gaussian Process* (GP) regression to represent the stochastic dependence of the parameters of the basic time-series model on the EOPs, which in turn, describes the long-term variability on the dynamics of the structural response.

Contrary to deterministic functional dependence models and PCE-based methods, where the EOPs are considered as the sole source of variability on the time-series model parameters, the appraised GP approach is further capable of capturing and quantifying the additional uncertainty stemming from other unmeasurable sources. Likewise, the obtained GP time-series model is totally linear-in-the-parameters, which facilitates the identification, parameter estimation and posterior model-based analysis. The issue of model identification is addressed by the Maximum Likelihood principle, which is solved by means of an Expectation-Maximization method adapted to the particular structure of the Gaussian Process time-series model. In addition, *Gaussian Process Principal Component Regression* (GP-PCR) is introduced as an optional improvement to the basic GP time-series model aiming at reducing the number of variables in the representation and to improve the numerical stability of the parameter estimation and optimization.

The methods discussed here are demonstrated on two dedicated case-studies. The first one pertains to the identification of actual data corresponding to the vertical acceleration response in the Humber bridge measured over 21 non-consecutive days in the period from May 19, 2011, to March 24, 2013. The second one corresponds to the identification of the long-term vibration response of a wind turbine, employing simulations obtained *via* the FAST wind turbine aeroelastic simulation code, which are characterized by non-stationary dynamics and long-term variability induced by variable wind speed.

The remainder of this paper is organized as follows: Section 2 initially provides a summary of traditional linear-in-the-parameters time-series models, pointing out their limitations in long-term identification, and offering their natural extension *via* GP regression. In addition, principal component regression is introduced as an alternative to reduce the number of identified parameters. Subsequently, Section 3 is devoted to the identification of the GP time-series model based on a set of dynamic responses, including the estimation of the parameters of individual realizations, the estimation of the hyper-parameters of the representation and the assessment and validation of the obtained model. Finally, Section 4 provides the two aforementioned case studies, i.e., the Humber bridge and wind turbine simulated vibrational response, while Section 5 concludes the study.

2. MODELS OF THE DYNAMIC RESPONSE OF STRUCTURES

2.1. Traditional Linear-in-the-Parameters Regressive Time-Series Models

Consider the dynamic response of a structure $y[t] \in \mathbb{R}$ defined over the normalized discrete time $t = 1, 2, \ldots, N$ and sampled with a sampling rate f_s. The response is assumed to obey the linear-in-the-parameters regressive time-series model:

$$y[t] = \boldsymbol{\phi}^T(\boldsymbol{z}[t]) \cdot \boldsymbol{\theta} + w[t], \qquad w[t] \sim \text{NID}(0, \sigma_w^2) \quad (1)$$

where $\boldsymbol{\phi}(\boldsymbol{z}[t]) \in \mathbb{R}^n$ is the *regression vector*, $\boldsymbol{z}[t] \in \mathbb{R}^{n_z}$ is the vector of *regressed variables*, $\boldsymbol{\theta} \in \mathbb{R}^n$ is the *parameter vector* and

$w[t]$ is a *Normally and Identically Distributed* (NID) innovations of mean zero and variance σ_w^2. The vector of regressed variables $z[t]$ may contain previous values of the dynamic response $y[t]$, excitation inputs $x[t]$ and innovations $w[t]$, according to the model type. Accordingly, equation (1) corresponds to the case of either an *output-only* or a *Multiple Input Single Output* (MISO) model. However, the *Multiple Input Multiple Output* (MIMO) case can be cast into the presently discussed framework through the proper rearrangement of the regression and parameter vectors, as well as the redefinition of the dynamic response and innovations as column vectors. In addition, linear state space representations may be considered when observation errors are an important component of the dynamic response measurements.

The linear-in-the-parameters regressive time-series model of equation (1) encompasses a large group of time-series models, which differ in the specific form of the regression vector and the vector of regressed variables. A few important examples are summarized next:

- AutoRegressive (AR) models: the simplest case corresponds to the AR model, for which the regression vector is of the form ((Ljung, 1999) Ch. 4):

$$\boldsymbol{\phi}(z[t]) = z[t] = \begin{bmatrix} y[t-1] & y[t-2] & \cdots & y[t-n_a] \end{bmatrix}^T \quad (2)$$

where n_a is the order of the AR model.

- AutoRegressive Moving Average (ARMA) models: ARMA models further include previous values of the innovations in the vector of regressed variables, and thus (Ljung, 1999):

$$\begin{aligned} \boldsymbol{\phi}(z[t]) &= z[t] \\ &= \begin{bmatrix} y[t-1] & \cdots & y[t-n_a] & w[t-1] & \cdots & w[t-n_c] \end{bmatrix}^T \end{aligned}$$
(3)

where n_a and n_c are the orders of the AR and MA parts of the model.

- AutoRegressive with eXogenous variable (ARX) models: ARX models combine previous values of the dynamic response and the excitation vector $x[t] \in \mathbb{R}^{n_x}$, and thus ((Ljung, 1999) Ch. 4):

$$\begin{aligned} \boldsymbol{\phi}(z[t]) &= z[t] \\ &= \begin{bmatrix} y[t-1] & \cdots & y[t-n_a] & x^T[t-1] & \cdots & x^T[t-n_b] \end{bmatrix}^T \end{aligned}$$
(4)

where n_a and n_b are the orders of the AR and exogenous parts of the model.

- Linear Parameter Varying AR (LPV-AR) models: LPV-AR models are a class of time-dependent AR models, which correspond to a generalization of the simple AR model, where the parameters of the AR model are dependent on an external *scheduling variable* $\beta[t]$ determining the values of these parameters at time t. The regression vector in the case of LPV-AR models is of the form (Avendaño-Valencia and Fassois, 2017b):

$$z[t] = \begin{bmatrix} y[t-1] & y[t-2] & \cdots & y[t-n_a] \end{bmatrix}^T \quad (5a)$$

$$\boldsymbol{\phi}(z[t]) = z[t] \otimes \begin{bmatrix} g_{b_1}(\beta[t]) & g_{b_2}(\beta[t]) & \cdots & g_{b_p}(\beta[t]) \end{bmatrix}^T \quad (5b)$$

where n_a is the AR order, $\otimes$ denotes the Kronecker product, $g_{b_j}(\beta[t])$ is the *j*-th functional expansion basis, and p is the order of the functional expansion basis. The closely related *Functional Series TAR* (FS-TAR) models form a special case of the LPV-AR model where the scheduling variable is time, i.e., $\beta[t] \equiv t$ (Poulimenos and Fassois, 2006).

- Non-linear AR (NAR) models: NAR models correspond to the non-linear counterpart of AR models, where the dynamic response is regressed on non-linear functions of its previous values, so that the regression vector assumes the form (Spiridonakos and Chatzi, 2015):

$$z[t] = \begin{bmatrix} y[t-1] & y[t-2] & \cdots & y[t-n_a] \end{bmatrix}^T \quad (6a)$$

$$\boldsymbol{\phi}(z[t]) = \begin{bmatrix} g_1(z[t]) & g_2(z[t]) & \cdots & g_n(z[t]) \end{bmatrix}^T \quad (6b)$$

where n_a is the AR order, and $g_j(z[t])$ is the *j*–th non-linear term of the vector of regressed variables.

Equation (1) may be expressed alternatively as follows:

$$w[t] = y[t] - \boldsymbol{\phi}^T(z[t]) \cdot \boldsymbol{\theta} = y[t] - y[t|t-1] \quad (7)$$

where $y[t|t-1] := \mathrm{E}\{y[t]|\boldsymbol{\theta}, \boldsymbol{\phi}(z[t])\}$ is the *one-step-ahead prediction* of the dynamic response, with associated variance $\mathrm{E}\{(y[t_1] - y[t_1|t_1-1]) \cdot (y[t_2] - y[t_2|t_2-1])\} = \sigma_w^2 \cdot \delta[t_1 - t_2]$, where t_1, t_2 are two analysis instants, and $\delta[t]$ denotes the Kronecker delta. Hence, under the NID assumption of the innovations $w[t]$, the conditional probability of the dynamic response $y[t]$ given the parameter vector $\boldsymbol{\theta}$ and the regression vector $\boldsymbol{\phi}(z[t])$ is Gaussian with mean $y[t|t-1]$ and variance σ_w^2, or more specifically:

$$\begin{aligned} p(y[t] \mid \boldsymbol{\theta}, \boldsymbol{\phi}(z[t])) &= \mathcal{N}_{y[t]}(y[t|t-1], \sigma_w^2), \\ y[t|t-1] &= \boldsymbol{\phi}^T(z[t]) \cdot \boldsymbol{\theta} \end{aligned}$$
(8)

where $\mathcal{N}_x(x_o, \sigma_x^2)$ denotes a Gaussian distribution for the random variable x with mean x_o and variance σ_x^2. Moreover, by virtue of the NID nature of the innovations, it follows that the probability for an entire vibration response realization of length N aggregated in the vector $\boldsymbol{y} = [y[1] \ y[2] \cdots y[N]]^T$, is:

$$p(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{\Phi}) = \prod_{t=1}^{N} p(y[t] \mid \boldsymbol{\theta}, \boldsymbol{\phi}(z[t])) = \prod_{t=1}^{N} \mathcal{N}_{y[t]}(y[t|t-1], \sigma_w^2) \quad (9)$$

where $\boldsymbol{\Phi} = [\boldsymbol{\phi}(z[1]) \ \boldsymbol{\phi}(z[2]) \ \cdots \ \boldsymbol{\phi}(z[N])] \in \mathbb{R}^{n \times N}$ is the *N*-sample long *regression matrix*. Then, by introduction of equation 8, and by virtue of the properties of exponential functions, it follows that (see Appendix A.1):

$$p(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{\Phi}) = \mathcal{N}_y(\boldsymbol{\Phi}^T \cdot \boldsymbol{\theta}, \sigma_w^2 \cdot \boldsymbol{I}_N) \quad (10)$$

where $\boldsymbol{I}_N$ indicates the *N*-size identity matrix. The conditional PDF $p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\Phi})$, seen as a function of the parameter vector $\boldsymbol{\theta}$, determines the *likelihood of the parameter vector* $\mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{\Phi})$. Furthermore, after assuming that the coefficient vector $\boldsymbol{\theta}$ is a deterministic variable, *Maximum Likelihood* (ML) estimates may be obtained by determining the values that maximize the likelihood $\mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{\Phi})$ ((Ljung, 1999), Sec. 7.4).

2.2. Limitations of the Traditional Linear Regressive Models

Although the linear-in-the-parameters regressive time-series model shown in equation (1) is useful for representing diverse classes of time-series, including stationary, non-stationary and non-linear, it lacks the flexibility to effectively represent variable dynamics stemming from variable *Environmental and Operational Conditions* (EOCs). For instance, it is known that the elasticity modulus of a material may change with temperature, and in turn, a change in this variable would modify the natural frequencies and damping ratios associated with the dynamic response of the structure. Hence, the model in equation (1) with a fixed parameter vector $\boldsymbol{\theta}$, would fail to effectively represent the dynamic response of the structure over a long period of analysis.

Instead, it may be considered that during a given period of time, say $T = N \cdot f_s$, the EOCs remain more or less constant, and consequently, the physical parameters of the structure would also remain constant. Under these conditions, the linear regression model of equation (1) is a valid representation of the dynamic response of the structure for such analysis period, while the parameter vector $\boldsymbol{\theta}$ would change according to the EOCs. Two main questions may be identified in this regard, the first is on how to select the length of the period where it is considered that the structural parameters remain more or less constant; the second is on how to represent the variability in the parameters as a function of the EOCs. The selection of the period of pseudoconstant dynamics may be obtained empirically by means of stationarity tests, as described for example in Kay (2008), Basu et al. (2009), and Borgnat et al. (2010). On the other hand, the representation of the variability in the dynamics of the structure as an effect of the EOCs is the main problem addressed in this work, for which a Gaussian Process Regression approach is postulated, as shown in the remainder of this work.

2.3. Gaussian Process Time-Series Model

The key assumption in the *Gaussian Process* (GP) regression approach is that the parameter vector of the time-series model follows a Gaussian distribution. Therefore, the linear-in-the-parameters regressive time-series model of equation (1), is complemented as follows:

$$y[t] = \boldsymbol{\phi}^T(z[t]) \cdot \boldsymbol{\theta} + w[t], \quad w[t] \sim \texttt{NID}(0, \sigma_w^2) \quad (11a)$$

$$\boldsymbol{\theta} = \boldsymbol{\theta}_o(\boldsymbol{\xi}, \boldsymbol{M}) + \boldsymbol{u}, \quad \boldsymbol{u} \sim \texttt{NID}(\boldsymbol{0}_{n \times 1}, \boldsymbol{\Sigma}_\theta) \quad (11b)$$

where $\boldsymbol{\xi} \in \mathbb{R}^m$ is the *Environmental and Operational Parameter* (EOP) vector determining the EOCs in the analysis period, $\boldsymbol{\theta}_0(\boldsymbol{\xi}, \boldsymbol{M}) = \text{E}\{\boldsymbol{\theta}|\boldsymbol{\xi}, \boldsymbol{M}\}$ is the mean parameter vector indicating the expected value of the parameter vector given the EOP $\boldsymbol{\xi}$ and the matrix of projection coefficients $\boldsymbol{M} \in \mathbb{R}^{n \times p}$, and $\boldsymbol{u}$ is an NID random vector with mean zero and covariance $\boldsymbol{\Sigma}_\theta$. The model is completed by the following functional series expansion of the mean parameter vector:

$$\boldsymbol{\theta}_o(\boldsymbol{\xi}, \boldsymbol{M}) = \sum_{j=1}^{p} \boldsymbol{\mu}_j \cdot g_{b_j}(\boldsymbol{\xi}) = \boldsymbol{M} \cdot \boldsymbol{g}(\boldsymbol{\xi}) \quad (12)$$

where p is the order of the functional series expansion and:

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{\mu}_1 & \cdots & \boldsymbol{\mu}_p \end{bmatrix} \in \mathbb{R}^{n \times p} \quad \boldsymbol{g}(\boldsymbol{\xi}) = \begin{bmatrix} g_{b_1}(\boldsymbol{\xi}) \\ \vdots \\ g_{b_p}(\boldsymbol{\xi}) \end{bmatrix} \in \mathbb{R}^{p \times 1} \quad (13)$$

are the matrix of projection coefficients and the functional basis vector containing the basis with indices $\boldsymbol{b} = \begin{bmatrix} b_1 & \cdots & b_p \end{bmatrix}^T_{p \times 1}$, $b_j \in \mathbb{N}$. Note that any type of non-linear function may have been used to describe the parameter variation, however, a linear-in-the-parameter structure has been selected—again—in equation (12) to facilitate the estimation and analysis of the model. A time-series model obeying equation (11) shall be referred to as a *Gaussian Process* (GP) time-series model and is characterized by a set of deterministic parameters, referred to as *hyper-parameters* $\mathcal{P} = \{\boldsymbol{M}, \boldsymbol{\Sigma}_\theta, \sigma_w^2\}$, consisting of the matrix of projection coefficients, the parameter covariance matrix and the innovations variance.

For a GP time-series model, the instantaneous value of the dynamic response $y[t]$ and the parameter vector $\boldsymbol{\theta}$ are jointly distributed Gaussian variables, with the joint PDF conditioned on the regression vector, the EOP vector and the hyper-parameters shown next:

$$p(y[t], \boldsymbol{\theta} \mid \boldsymbol{\phi}[t], \boldsymbol{\xi}, \mathcal{P}) = p(y[t] \mid \boldsymbol{\phi}[t], \boldsymbol{\theta}, \boldsymbol{\xi}, \mathcal{P}) \cdot p(\boldsymbol{\theta} \mid \boldsymbol{\phi}[t], \boldsymbol{\xi}, \mathcal{P}) \quad (14)$$

where, from equation (11), it follows that:

$$p(y[t] \mid \boldsymbol{\theta}, \boldsymbol{\phi}[t], \boldsymbol{\xi}, \mathcal{P}) = \mathcal{N}_{y[t]}(y[t|t-1], \sigma_w^2) \quad (15a)$$

$$p(\boldsymbol{\theta} \mid \boldsymbol{\phi}[t], \boldsymbol{\xi}, \mathcal{P}) = \mathcal{N}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_o(\boldsymbol{\xi}, \mathcal{M}), \boldsymbol{\Sigma}_{\boldsymbol{\theta}}) \quad (15b)$$

Similarly, when the dynamic response over the complete period of analysis of length N, namely $\boldsymbol{y}$, is considered, the respective joint conditional PDF takes the form:

$$p(\boldsymbol{y}, \boldsymbol{\theta} \mid \boldsymbol{\Phi}, \boldsymbol{\xi}, \mathcal{P}) = \prod_{t=1}^{N} p(y[t], \boldsymbol{\theta} \mid \boldsymbol{\phi}[t], \boldsymbol{\xi}, \mathcal{P})$$

$$= \prod_{t=1}^{N} p(y[t] \mid \boldsymbol{\theta}, \boldsymbol{\phi}[t], \boldsymbol{\xi}, \mathcal{P}) \cdot p(\boldsymbol{\theta} \mid \boldsymbol{\phi}[t], \boldsymbol{\xi}, \mathcal{P}) \quad (16)$$

which, under the Gaussianity of both $p(y[t]|\boldsymbol{\theta}, \boldsymbol{\phi}[t], \boldsymbol{\xi}, \mathcal{P})$ and $p(\boldsymbol{\theta} \mid \boldsymbol{\phi}[t], \boldsymbol{\xi}, \mathcal{P})$, becomes:

$$p(\boldsymbol{y}, \boldsymbol{\theta} \mid \boldsymbol{\Phi}, \boldsymbol{\xi}, \mathcal{P}) =$$

$$\underbrace{\mathcal{N}_y(\boldsymbol{\Phi}^T \cdot \boldsymbol{\theta}, \sigma_w^2 \cdot \boldsymbol{I}_N)}_{p(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{\Phi}, \boldsymbol{\xi}, \mathcal{P})} \cdot \underbrace{\mathcal{N}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_o(\boldsymbol{\xi}, \boldsymbol{M}), N^{-1} \cdot \boldsymbol{\Sigma}_{\boldsymbol{\theta}})}_{p(\boldsymbol{\theta} \mid \boldsymbol{\Phi}, \boldsymbol{\xi}, \mathcal{P})} \quad (17)$$

In addition, according to the conditional density axiom, the joint PDF may be decomposed as follows ((Rasmussen and Williams, 2006), p. 9):

$$p(\boldsymbol{y}, \boldsymbol{\theta} \mid \boldsymbol{\Phi}, \boldsymbol{\xi}, \mathcal{P}) = p(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{\Phi}, \boldsymbol{\xi}, \mathcal{P}) \cdot p(\boldsymbol{\theta} \mid \boldsymbol{\Phi}, \boldsymbol{\xi}, \mathcal{P})$$

$$= p(\boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{\Phi}, \boldsymbol{\xi}, \mathcal{P}) \cdot p(\boldsymbol{y} \mid \boldsymbol{\Phi}, \boldsymbol{\xi}, \mathcal{P}) \quad (18)$$

where $p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{\Phi}, \boldsymbol{\xi}, \mathcal{P})$ is the *posterior* PDF of the parameter vector after observing the dynamic response, and $p(\boldsymbol{y} \mid \boldsymbol{\Phi}, \boldsymbol{\xi}, \mathcal{P})$ is the *marginal probability of the dynamic response*, comprising a function of the hyper–parameters and is referred to as the *marginal likelihood* of the hyper-parameters $\mathcal{L}(\mathcal{P} \mid \boldsymbol{y}, \boldsymbol{\Phi}, \boldsymbol{\xi})$. The latter is obtained by marginalizing the joint conditional PDF $p(\boldsymbol{y}, \boldsymbol{\theta} \mid \boldsymbol{\Phi}, \boldsymbol{\xi}, \mathcal{P})$ with respect to all the possible values of $\boldsymbol{\theta}$ (Rasmussen and Williams, 2006), p. 9). Given that both distributions $p(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{\Phi}, \boldsymbol{\xi}, \mathcal{P})$ and $p(\boldsymbol{\theta} \mid \boldsymbol{\Phi}, \boldsymbol{\xi}, \mathcal{P})$ are Gaussian, the posterior parameter PDF and marginal likelihood are Gaussian as well, of the form (Rasmussen and Williams, 2006), p. 9):

$$p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{\Phi}, \boldsymbol{\xi}, \mathcal{P}) = \mathcal{N}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}, \boldsymbol{P}_{\boldsymbol{\theta}}) \tag{19a}$$

$$\mathcal{L}(\mathcal{P} \mid \boldsymbol{y}, \boldsymbol{\Phi}, \boldsymbol{\xi}) = p(\boldsymbol{y}|\boldsymbol{\Phi}, \boldsymbol{\xi}, \mathcal{P}) = \mathcal{N}_{\boldsymbol{y}}(\boldsymbol{\Phi}^T \cdot \boldsymbol{\theta}_o(\boldsymbol{\xi}, M), \boldsymbol{\Sigma}_{\varepsilon_k}) \tag{19b}$$

where

$$\hat{\boldsymbol{\theta}} = \mathrm{E}\{\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{\Phi}, \boldsymbol{\xi}, \mathcal{P}\} = \boldsymbol{\theta}_o(\boldsymbol{\xi}, M) + \boldsymbol{K} \cdot (\boldsymbol{y} - \boldsymbol{\Phi}^T \cdot \boldsymbol{\theta}_o(\boldsymbol{\xi}, M)) \tag{20a}$$

$$\boldsymbol{P}_{\boldsymbol{\theta}} = \mathrm{E}\{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \cdot (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T|\boldsymbol{y}, \boldsymbol{\Phi}, \boldsymbol{\xi}, \mathcal{P}\} = (\boldsymbol{I}_n - \boldsymbol{K} \cdot \boldsymbol{\Phi}^T) \cdot \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \tag{20b}$$

$$\boldsymbol{K} = \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \cdot \boldsymbol{\Phi} \cdot \boldsymbol{\Sigma}_{\varepsilon}^{-1} \tag{20c}$$

$$\boldsymbol{\Sigma}_{\varepsilon} = \sigma_w^2 \cdot \boldsymbol{I}_N + \boldsymbol{\Phi}^T \cdot \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \cdot \boldsymbol{\Phi} \tag{20d}$$

In the previous equations $\hat{\boldsymbol{\theta}} \in \mathbb{R}^n$ corresponds to the *posterior parameter mean* with the associated *posterior parameter covariance matrix* $\boldsymbol{P}_{\boldsymbol{\theta}} \in \mathbb{R}^{n \times n}$. Moreover, the quantities:

$$\tilde{y}[t|t-1] := \boldsymbol{\phi}^T(\boldsymbol{z}[t]) \cdot \boldsymbol{\theta}_o(\boldsymbol{\xi}, M), \quad \varepsilon[t] := y[t] - \tilde{y}[t|t-1] \tag{21a}$$

$$y[t|t-1] := \boldsymbol{\phi}^T(\boldsymbol{z}[t]) \cdot \hat{\boldsymbol{\theta}}, \quad e[t] := y[t] - y[t|t-1] \tag{21b}$$

are referred to as the *prior* and *posterior* one-step-ahead predictions, respectively, with associated prior and posterior prediction errors $\varepsilon[t]$ and $e[t]$. In addition, $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} \in \mathbb{R}^{N \times N}$ is the covariance matrix of the prior prediction error. The main difference between the prior and posterior predictions, is that the prior predictions correspond to the best guess of the dynamic response in the absence of knowledge of the actual parameter vector at the period of analysis, while the posterior predictions are the best guess of the dynamic response based on (an estimate of) the actual parameter vector.

2.3.1. Remark – Unknown or Non-Measurable Sources of Uncertainty

The GP time-series model may also be used in the context where the EOP vector is either unknown or unmeasurable. If that is the case, the functional series expansion of the mean parameter vector in equation (12) is limited to a single constant term, so that $\boldsymbol{\theta}_0(\boldsymbol{\xi}, M) := \boldsymbol{\theta}_0 = \boldsymbol{\mu}_1 \cdot 1$, while the parameter PDF reduces to the conventional multivariate Gaussian model, so that $\boldsymbol{\theta} \sim \mathcal{N}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$. Such a case has been explored in Avendaño-Valencia et al. (2015a).

2.4. Regression on a Reduced Parameter Set *via* Principal Component Regression

A potential difficulty in the adoption of the GP time-series models lies in the computational cost due to the large number of parameters that need to be estimated. Moreover, a coefficient vector covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ with full structure implies that several of the estimated parameters are redundant and/or unnecessary. A potential solution to this problem is to use a dimensionality reduction scheme for the regression matrix, such as in *Principal Component Regression* (PCR), which is explained next.

To start with, consider the matrix $\bar{\bar{\boldsymbol{\Phi}}}_K = [\boldsymbol{\Phi}_1 \, \boldsymbol{\Phi}_2 \, \cdots \, \boldsymbol{\Phi}_K]^T \in \mathbb{R}^{n \times (N \cdot K)}$ containing the regression matrices associated with the set of dynamic responses $\mathcal{Y}_K$, which possesses the *Singular Value Decomposition* (SVD) $\bar{\boldsymbol{\Phi}} = \boldsymbol{U} \cdot \boldsymbol{\Lambda} \cdot \boldsymbol{V}^T$, where $\boldsymbol{U} \in \mathbb{R}^{n \times m}$ and $\boldsymbol{V} \in \mathbb{R}^{(N \cdot K) \times m}$ are orthogonal matrices, and $\boldsymbol{\Lambda} \in \mathbb{R}^{m \times m}$, with $m = \min\{(N \cdot K - 1), n\}$ designating the rank of $\bar{\boldsymbol{\Phi}}_K$, is a diagonal singular value matrix with entries $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$. Moreover, consider the vector $\boldsymbol{d} = \begin{bmatrix} d_1 & d_2 & \cdots & d_{\tilde{m}} \end{bmatrix}^T$ of dimension $\tilde{m} \leq m$ containing the indices of selected singular values, and the truncated eigenvector matrices $\tilde{\boldsymbol{U}} \in \mathbb{R}^{n \times \tilde{m}}$ and $\tilde{\boldsymbol{V}} \in \mathbb{R}^{(N \cdot K) \times \tilde{m}}$ built from the columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ corresponding to the indices in $\boldsymbol{d}$.

Hence, if the regression vector $\boldsymbol{\phi}(\boldsymbol{z}[t])$ is projected into the column Eigen-space, so that:

$$\tilde{\boldsymbol{U}} \cdot \widetilde{\boldsymbol{\phi}}(\boldsymbol{z}[t]) = \boldsymbol{\phi}(\boldsymbol{z}[t]) \quad \leftrightarrow \quad \widetilde{\boldsymbol{\phi}}(\boldsymbol{z}[t]) = \tilde{\boldsymbol{U}}^T \cdot \boldsymbol{\phi}(\boldsymbol{z}[t]) \tag{22}$$

Then, upon replacement of the previous result into the original regression model, the alternative *Principal Component Gaussian Process Regression* (PC-GP) time-series model is obtained:

$$y[t] = \widetilde{\boldsymbol{\phi}}^T(\boldsymbol{z}[t]) \cdot \boldsymbol{\vartheta} + w[t], \qquad w[t] \sim \mathtt{NID}(0, \sigma_w^2) \tag{23a}$$

$$\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_o(\boldsymbol{\xi}, \tilde{M}) + \tilde{\boldsymbol{u}}, \qquad \tilde{\boldsymbol{u}} \sim \mathtt{NID}(\mathbf{0}_{\tilde{m} \times 1}, \boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}) \tag{23b}$$

where $\boldsymbol{\vartheta} = \tilde{\boldsymbol{U}}^T \cdot \boldsymbol{\theta} \in \mathbb{R}^{\tilde{m}}$ is a reduced dimensionality parameter vector. Note that the matrix $\boldsymbol{U}$ and the scaled singular values $\boldsymbol{\Lambda}^2/N$ correspond to the matrix of principal vectors and the matrix of principal values of the Principal Component Analysis (PCA) of the covariance matrix estimate $\bar{\boldsymbol{\Phi}} \cdot \bar{\boldsymbol{\Phi}}^T/N$, thus the name *Principal Component Regression* (PCR) ((Bair et al., 2006; Hastie et al., 2009), Section 3.5).

Two main advantages are obtained through the use of the PC-GP method (Hastie et al., 2009): (i) the dimension of the parameter vector is reduced from n to $\tilde{m}$; (ii) since the regression is built on orthogonal regressors (contained in the matrix $\tilde{\boldsymbol{U}}$), the reduced parameter vector $\boldsymbol{\vartheta}$ is also uncorrelated, and thus the corresponding covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}$ is diagonal. Consequently, only the diagonal elements of the matrix need to be estimated. Additionally, the original parameter vector may be retrieved *via* the operation:

$$\boldsymbol{\theta} = \tilde{\boldsymbol{U}} \cdot \boldsymbol{\vartheta} \tag{24}$$

3. IDENTIFICATION OF THE GP TIME-SERIES MODEL

The identification of a GP time-series model may be stated as the problem of determining the hyper-parameters $\mathcal{P}$ and the

structural parameters (consisting of the model and basis orders) that best fit a given a set of K dynamic responses, say $\mathcal{Y}_K = \{y_1, y_2, \ldots, y_K\}$. Additionally, according to the model type, a corresponding set of excitation inputs $\mathcal{X}_K = \{x_1, x_2, \ldots, x_K\}$ and EOP vectors $\Xi_K = \{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_K\}$ are provided. Additionally, it may be of interest to determine the parameter vectors associated with each one of the realizations, i.e., each one of the parameter vectors $\boldsymbol{\theta}_k$ for all $k = 1, \ldots, K$. These three topics are analyzed next.

3.1. Estimation of the Parameter Vectors of Individual Realizations

Maximum A Posteriori (MAP) estimates of the parameter vector of the GP time-series model for a single realization y_k are obtained by evaluating the values that maximize the posterior PDF $p(\boldsymbol{\theta}_k \mid y_k, \Phi_k, \boldsymbol{\xi}_k, \mathcal{P})$ in equation (19a). Given that the posterior distribution is Gaussian, the MAP estimates correspond to the posterior mean (equal to the mode of the Gaussian distribution), which may be computed *via* equation (20) for given values of $\mathcal{P}$.

3.2. Estimation of the Hyperparameters
3.2.1. Maximum Likelihood Estimation of the Hyperparameters

Maximum likelihood estimates of the hyperparameters are obtained *via* optimization of the marginalized hyper-parameter likelihood for the complete set of data. Accordingly, ML estimates are obtained from the optimization problem ((Rasmussen and Williams, 2006), ch. 5; (Shumway and Stoffer, 2011), ch. 6):

$$\hat{\mathcal{P}} = \arg\max_{\mathcal{P}} \sum_{k=1}^{K} \ln \mathcal{L}(\mathcal{P}|y_k, \Phi_k, \boldsymbol{\xi}_k) \tag{25a}$$

$$\sum_{k=1}^{K} \ln \mathcal{L}(\mathcal{P}|y_k, \Phi_k, \boldsymbol{\xi}_k) = -\frac{N \cdot K}{2} \ln 2\pi$$
$$- \frac{1}{2} \sum_{k=1}^{K} \left(\ln |\Sigma_{\varepsilon_k}| + \boldsymbol{\varepsilon}_k^T \cdot \Sigma_{\varepsilon_k}^{-1} \cdot \boldsymbol{\varepsilon}_k \right) \tag{25b}$$

where $\boldsymbol{\varepsilon}_k = \begin{bmatrix} \varepsilon_k[1] & \varepsilon_k[2] & \cdots & \varepsilon_k[N] \end{bmatrix}^T_{N \times 1}$ is the vector of prior prediction errors, with $\varepsilon_k[t]$ defined in equation (21a). Although it is possible to analytically solve the ML optimization problem in equation (25a) for some of the hyper-parameters (in particular for a constant parameter mean), the problem becomes intractable for other quantities. Alternatively, the Expectation-Maximization (EM) algorithm constitutes a powerful tool to solve this optimization problem.

3.2.2. Expectation-Maximization Algorithm for Efficient Computation of the ML Estimates

The *Expectation-Maximization* (EM) algorithm attempts to maximize the conditional expectation of the logarithm of the joint conditional PDF $p(y_k, \boldsymbol{\theta}_k \mid \Phi_k, \boldsymbol{\xi}_k, \mathcal{P})$, with respect to available data ((Shumway and Stoffer, 2011), ch. 6). Formally expressed, the EM algorithm aims at maximizing the expected log-likelihood:

$$Q(\mathcal{P}|\mathcal{P}^{(-)}) = \mathrm{E}_{\boldsymbol{\theta}|\mathcal{D}, \mathcal{P}^{(-)}} \left\{ -\sum_{k=1}^{K} \ln p(y_k, \boldsymbol{\theta}_k \mid \Phi_k, \boldsymbol{\xi}_k, \mathcal{P}) \right\} \tag{26}$$

where $\mathrm{E}_{\boldsymbol{\theta}|\mathcal{D}, \mathcal{P}^{(-)}} \{\cdot\}$ denotes the conditional expectation of the argument with respect to the space of $\boldsymbol{\theta}$ given data $\mathcal{D} = \{y_k, \Phi_k, \boldsymbol{\xi}_k\}, \forall\, k = 1, \ldots, K$ and hyper-parameters $\mathcal{P}^{(-)}$, and where:

$$\sum_{k=1}^{K} \ln p(y_k, \boldsymbol{\theta}_k \mid \Phi_k, \boldsymbol{\xi}_k, \mathcal{P})$$
$$= -\frac{1}{2} \sum_{k=1}^{K} \left(N \cdot \ln \sigma_w^2 + \sigma_w^{-2}(y_k - \Phi_k^T \cdot \boldsymbol{\theta}_k)^T \cdot (y_k - \Phi_k^T \cdot \boldsymbol{\theta}_k) \right.$$
$$\left. + \ln |\Sigma_{\boldsymbol{\theta}}| + (\boldsymbol{\theta}_k - \boldsymbol{\theta}_o(\boldsymbol{\xi}_k, M))^T \cdot \Sigma_{\boldsymbol{\theta}}^{-1} \cdot (\boldsymbol{\theta}_k - \boldsymbol{\theta}_o(\boldsymbol{\xi}_k, M))) \right) \tag{27}$$

Thus, after evaluating the expectation, the expected log-likelihood of the GP time-series model becomes:

$$Q(\mathcal{P}|\mathcal{P}^{(-)}) = Q_1(\mathcal{P}|\mathcal{P}^{(-)}) + Q_2(\mathcal{P}|\mathcal{P}^{(-)}) \tag{28}$$

$$Q_1(\mathcal{P}|\mathcal{P}^{(-)}) = -\frac{K \cdot N}{2} \ln \sigma_w^2$$
$$- \frac{1}{2\sigma_w^2} \sum_{k=1}^{K} \left(\left(e_k^{(-)} \right)^T \cdot e_k^{(-)} + \mathrm{tr}\left(\Phi_k^T \cdot P_k^{(-)} \cdot \Phi_k \right) \right)$$

$$Q_2(\mathcal{P}|\mathcal{P}^{(-)}) = -\frac{K}{2} \ln |\Sigma_{\boldsymbol{\theta}}|$$
$$- \frac{1}{2} \sum_{k=1}^{K} \left(\left(\boldsymbol{\delta}_k^{(-)} \right)^T \cdot \Sigma_{\boldsymbol{\theta}}^{-1} \cdot \boldsymbol{\delta}_k^{(-)} + \mathrm{tr}\left(\Sigma_{\boldsymbol{\theta}}^{-1} \cdot P_k^{(-)} \right) \right)$$

where $\mathrm{tr}(\cdot)$ denotes the trace operation, and:

$$e_k^{(-)} = y_k - \Phi_k^T \cdot \hat{\boldsymbol{\theta}}_k^{(-)}, \quad e_k^{(-)} \in \mathbb{R}^N \tag{29a}$$

$$\boldsymbol{\delta}_k^{(-)} = \hat{\boldsymbol{\theta}}_k^{(-)} - \boldsymbol{\theta}_o\left(\boldsymbol{\xi}_k, M^{(-)} \right), \quad \boldsymbol{\delta}_k^{(-)} \in \mathbb{R}^n \tag{29b}$$

and $\hat{\boldsymbol{\theta}}_k^{(-)}$ and $P_k^{(-)}$ are the MAP estimates of the mean and covariance of the coefficient vector given the hyperparameter values $\mathcal{P}^{(-)} = \{M^{(-)}, \Sigma_{\boldsymbol{\theta}}^{(-)}, \sigma_w^{2(-)}\}$ obtained with equation (20).

The Expectation-Maximization algorithm operates by selecting some initial hyperparameter values $\mathcal{P}^{(0)}$, and then, at each iteration $i = 1, 2, \ldots$, the following two steps are performed:

3.2.3. Expectation Step (E-Step)

The expected log-likelihood is evaluated based on the previous hyper-parameter values, $\mathcal{P}^{(i-1)}$. This translates into the evaluation of the mean and covariance matrix of the posterior coefficient PDF, by applying equation (20) on all the available dynamic response realizations $k = 1, \ldots, K$.

3.2.4. Maximization Step (M-Step)

Updated hyper-parameter values are obtained by computing the values that maximize the expected log-likelihood function, this is to say:

$$\mathcal{P}^{(i)} = \arg\max_{\mathcal{P}} Q\left(\mathcal{P}|\mathcal{P}^{(i-1)} \right) \tag{30}$$

which leads to the update equations:

$$M^{(i)} = \left(\sum_{k=1}^{K} \hat{\boldsymbol{\theta}}_k^{(i-1)} \cdot g^T(\boldsymbol{\xi}_k) \right) \cdot \left(\sum_{k=1}^{K} g(\boldsymbol{\xi}_k) \cdot g^T(\boldsymbol{\xi}_k) \right)^{-1} \quad (31a)$$

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}^{(i)} = \frac{1}{K} \sum_{k=1}^{K} \left(\boldsymbol{\delta}_k^{(i-1)} \cdot (\boldsymbol{\delta}_k^{(i-1)})^T + \boldsymbol{P}_k^{(i-1)} \right) \quad (31b)$$

$$\hat{\sigma}_w^{2(i)} = \frac{1}{K \cdot N} \sum_{k=1}^{K} \left((e_k^{(i-1)})^T \cdot e_k^{(i-1)} + \mathrm{tr}\left(\boldsymbol{\Phi}_k^T \cdot P_k^{(i-1)} \cdot \boldsymbol{\Phi}_k \right) \right) \quad (31c)$$

Moreover, if the parameter vector is constant, then the update equation for the mean parameter vector reduces to:

$$\boldsymbol{\theta}_o^{(i)} = \frac{1}{K} \sum_{k=1}^{K} \hat{\boldsymbol{\theta}}_k^{(i-1)} \quad (32)$$

In addition, in the case of the PC-GP time-series model, the diagonal structure of the coefficient covariance matrix leads to the simplified update equation:

$$\sigma_{\vartheta_j}^{2(i)} = \frac{1}{K} \sum_{k=1}^{K} \left((\vartheta_j^{(i-1)} - \vartheta_{j,o}^{(i-1)})^2 + \left[\boldsymbol{P}_k^{(i-1)} \right]_{j,j} \right),$$
$$\forall j = 1, 2, \ldots, \tilde{m} \quad (33)$$

where $[\boldsymbol{M}]_{a,b}$ is the entry of matrix $\boldsymbol{M}$ on row a and column b.

The E and M steps are iterated until a specific number of iterations, say N_{iter}, is reached, or until convergence, which may be assessed by evaluating if the norm of the difference between the current and previous values of the marginal likelihood and hyperparameter estimates is lower than a pre-specified threshold. The later translates into monitoring if any of the following conditions is true:

$$\Delta_P \geq |\mathcal{P}^{(i)} - \mathcal{P}^{(i-1)}| \quad (34a)$$

$$\Delta_{\mathcal{L}} \geq |\ln \mathcal{L}(P^{(i)}|\mathcal{Y}_K) - \ln \mathcal{L}(\mathcal{P}^{(i-1)}|\mathcal{Y}_K)| \quad (34b)$$

$$N_{\text{iter}} \geq i \quad (34c)$$

where Δ_p and $\Delta_{\mathcal{L}}$ are thresholds on the absolute difference of hyperparameters and the marginal likelihood updates.

The EM algorithm has been demonstrated to maximize the marginal likelihood (equation (25b)) at every step and to converge to a local maximum of the marginal likelihood located in the neighborhood of the given initial values (Shumway and Stoffer, 2011). In order to facilitate the convergence toward the global maximum, it is essential to provide a suitable set of initial hyperparameter values, which may be derived from an initial set of estimates of the coefficient vectors $\boldsymbol{\theta}_k$ obtained with traditional least squares or maximum likelihood methods as explained for example in Ljung (1999).

3.3. Model Assessment and Validation

Once the GP time-series model has been estimated, it is important to determine the performance of the model. Likewise, it may be of interest to compare with other model structures and determine which one is best for the data. For that purpose, the main tool for evaluating the performance is the marginal likelihood shown in equation (25b). However, precise evaluation of the marginal likelihood may be non-trivial, in particular because the prior prediction error covariance matrix is unknown. Instead, it may be preferable to evaluate the *Residual Sum of Squares* normalized by the *Series Sum of Squares* (RSS/SSS) based on the prior estimation residuals $\hat{\varepsilon}[t]$, as follows:

$$RSS/SSS_{(prior)} = \left(\sum_{k=1}^{K} \sum_{t=1}^{N} \hat{\varepsilon}_k^2[t] \right) \Big/ \left(\sum_{k=1}^{K} \sum_{t=1}^{N} y_k^2[t] \right),$$
$$\hat{\varepsilon}_k[t] = y_k[t] - \boldsymbol{\phi}_k^T(z[t]) \cdot \boldsymbol{\theta}_o(\boldsymbol{\xi}_k, \hat{\boldsymbol{M}}) \quad (35)$$

where $\hat{\boldsymbol{M}}$ corresponds to the estimates of the matrix of coefficients of projection.

Alternatively, a validation error can be evaluated in order to assess the representation effectiveness and generalization ability of the GP time-series model. In this sense, consider the *validation set* $\mathcal{Y}_L^{(v)} = \{ \boldsymbol{y}_1^{(v)}, \cdots, \boldsymbol{y}_L^{(v)} \}$ consisting of L dynamic responses, whose elements are independent from the set $\mathcal{Y}_K$ used for estimation (training) of the model. Then, the prior RSS/SSS in equation (35) may be evaluated based on the validation set, where the prior estimation residuals are replaced by the validation error $\hat{\varepsilon}_l^{(v)}[t]$, so that:

$$RSS/SSS_{(prior)}^{(v)} = \left(\sum_{l=1}^{L} \sum_{t=1}^{N} (\hat{\varepsilon}_l^{(v)}[t])^2 \right) \Big/ \left(\sum_{l=1}^{L} \sum_{t=1}^{N} (y_l^{(v)}[t])^2 \right),$$
$$\hat{\varepsilon}_l^{(v)}[t] = y_l^{(v)}[t] - \boldsymbol{\phi}_k^T(z_l^{(v)}[t]) \cdot \boldsymbol{\theta}_o(\boldsymbol{\xi}_l^{(v)}, \hat{\boldsymbol{M}}) \quad (36)$$

The validation RSS/SSS may be associated with the empirical risk ((Vapnik, 2000), ch. 1) for the loss function $L(y_k[t], y_k[t|t-1]) = \hat{\varepsilon}_k^2[t] \cdot \left(\sum_{k=1}^{K} \sum_{t=1}^{N} y_k^2[t] \right)^{-1}$.

4. CASE STUDIES

4.1. Long-term Identification of the Acceleration Response in the Humber Bridge

4.1.1. Data Description and Preprocessing

The Humber Bridge is a long span suspension bridge joining the small towns of Hessle (north) and Barton (south) in the UK. The main span of the bridge comprises 1,410 m and is built on aerodynamic steel box girders and inclined hangers, and supported by two reinforced concrete towers rising 155.5 m above the caisson foundations. The bridge is exposed to prevailing southwesterly cyclonic winds that can reach hurricane force (exceeding 32.7 m/s), with atmospheric temperatures ranging from −10 to 30°C. Further details of the structure of the Humber bridge may be found in Rahbari et al. (2015). The bridge has been instrumented with various sensors, including GPS antennas, accelerometers, inclinometers and extensometers. In addition, various environmental variables including wind speed and temperature at different locations of the bridge are also measured. The monitoring campaign comprised a three year period starting from January 11, 2011, to December 2, 2013. In the present study, the vertical

acceleration response signal measured at the midspan on the east side of the deck is selected for analysis, while the wind speed is used as EOP.

The vertical acceleration response signal is originally sampled at 20 Hz. For the present study however, it is down-sampled to 2 Hz in order to focus the analysis on the main structural frequencies, located under 1 Hz, while at the same time reducing the model complexity. Acceleration and wind speed signal segments of 250 s ($N = 500$ samples) are extracted every 30 min, thus resulting in a maximum of 48 segments per day. In the analysis presented here, the average wind speed on each analysis period is considered as the unique EOP for the construction of a GP-AR model of the vertical vibration of the bridge. Hence, in order to reduce the parameter uncertainty and the computational cost in the construction of the model, only the vibration records corresponding to the main wind direction (about $90 \pm 20°$) are utilized in the construction of the model. Note, however, that a more comprehensive representation of the vibration response of the bridge may be appraised by considering also the wind direction as an EOP in the GP-AR model. This issue shall be appraised in a future work. Thus, after the selection of the vibration records corresponding to the main wind direction and the removal of artifacts due to problems of the measuring system, a total of 7,000 signal segments are finally obtained for the construction of the model.

In **Figure 1** is provided a histogram of the average wind speed corresponding to the selected vibration signals. In addition, **Figure 2** displays a typical daily variation of the power spectral density (PSD) of the vertical acceleration response and the average wind speed (irrespective of the wind direction). The obtained PSDs demonstrate that the main natural frequencies remain relatively stable, although the amplitude of the vibration evidences large variations even during a single day. In particular, it is evident that during the low wind speed period observed on the first 4 h of the analyzed period, the power of the vibration is much lower in comparison with later hours where the wind speed increases.

4.1.2. Model Identification
4.1.2.1. Modeling of the Short-term Response
An AutoRegressive (AR) model structure is selected to represent the acceleration response on short, 250 s, intervals. The order of the AR model is selected by evaluating different model structures with orders in the range $n_a = [1, \ldots, 100]$. A subset of 7 days of data is selected to determine the model order. The prior RSS/SSS and Bayesian Information Criterion (BIC) curves, as well as the frequency stabilization plot, displaying the natural frequencies and damping ratios associated with the estimated AR models for increasing orders, are shown in **Figure 3**. It may be observed that the RSS/SSS tends to favor large AR orders, while the BIC clearly demonstrates several minima, and a global minimum found for the value $n_a = 72$. The frequency stabilization plot seems to confirm these findings, by demonstrating that the main frequency peaks found in the PSD are accommodated by AR models with order around $n_a = 72$. Thus, according to the frequency stabilization plot and the BIC curve, the selected model order for subsequent analysis is $n_a = 72$.
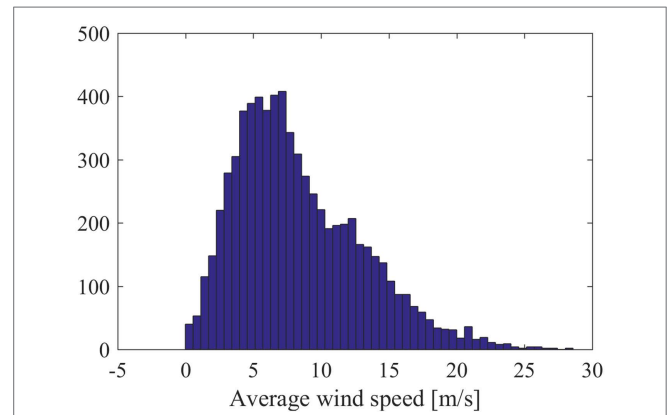

FIGURE 1 | Histogram displaying the distribution of the average wind speeds available for the identification of the model.

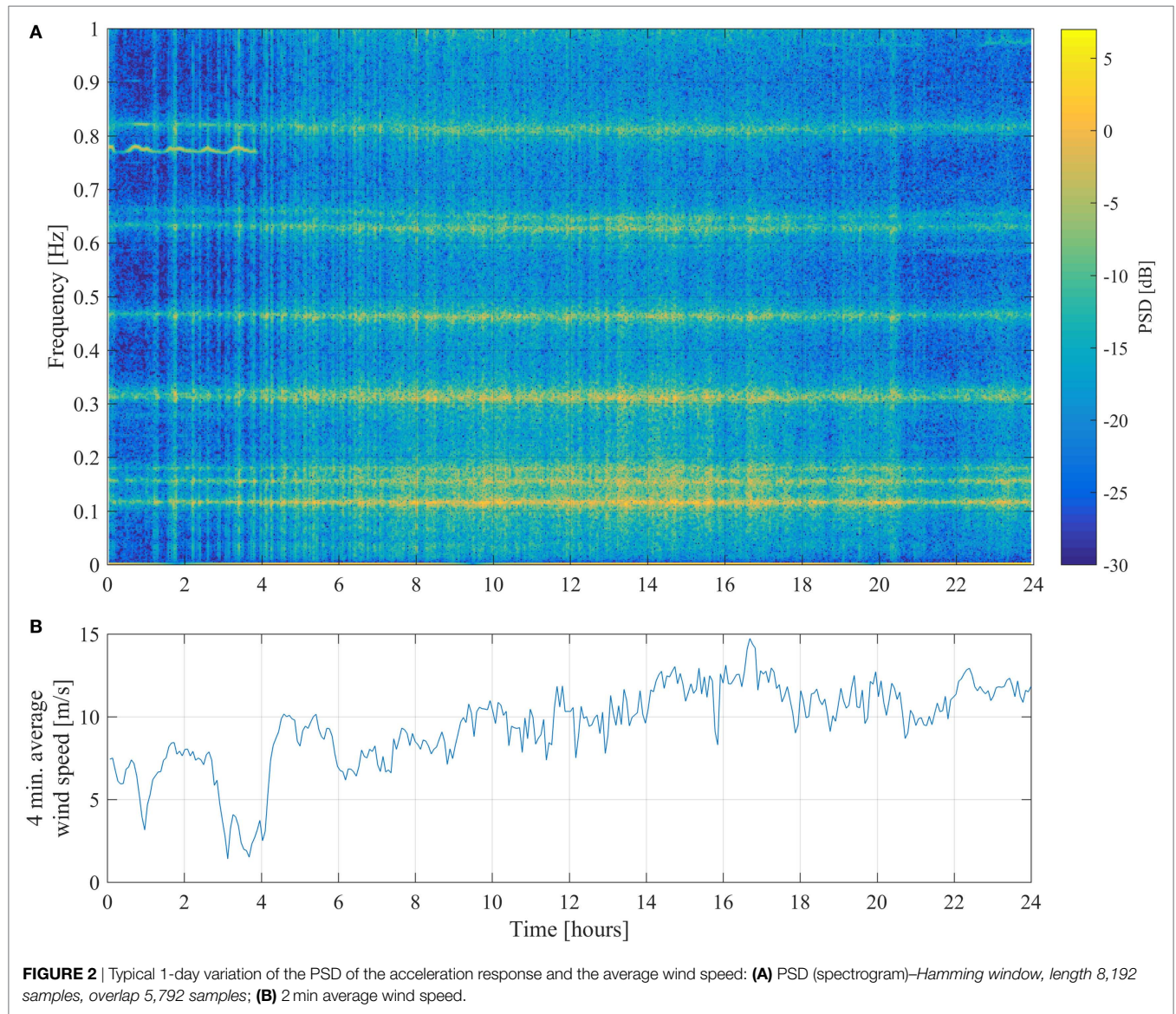#### 4.1.2.2. Modeling of the Long-term Response
The long-term acceleration response of the bridge is represented by means of a GP-AR model where the 250 s average wind speed is used as EOP, this is to say $\boldsymbol{\xi}$. For this purpose, the average wind speed is normalized from the range [0, 30] m/s to the range [0,1] by making $\boldsymbol{\xi} = AWS/30$, where AWS stands for the 250 s Average Wind Speed. The functional basis used for the expansion of the parameter vector of the model corresponds to the class of Hermite orthogonal polynomials, which satisfy the recurrence relation:

$$g_{j+1}(\boldsymbol{\xi}) = H_j(\boldsymbol{\xi}) = \boldsymbol{\xi} \cdot H_{j-1}(\boldsymbol{\xi}) - (j-1) \cdot H_{j-2}(\boldsymbol{\xi}), \quad \forall j \geq 2 \tag{37}$$

$$g_1(\boldsymbol{\xi}) = H_o(\boldsymbol{\xi}) = 1, \qquad g_2(\boldsymbol{\xi}) = H_1(\boldsymbol{\xi}) = \boldsymbol{\xi}$$

Thus, GP-AR(72) models are estimated using the EM algorithm for increasing basis orders in the range of $p = 1, \ldots, 6$. For the application of the EM algorithm and validation of the performance of the obtained models, the whole dataset consisting of 7,000 segments of 250 s is separated into training and validation subsets. The training subset is composed by the initial 3,000 segments, while the validation subset is composed by the remaining 4,000 segments. It should be noted that the GP-AR(72) model with $p = 1$ would correspond to the case when the EOP variable is ignored and the parameter vector is assumed to be Gaussian with constant mean and covariance matrix. The settings used for the application of the EM algorithm are shown in **Table 1**. Note that the threshold $\Delta_{RSS/SSS}$ is presently utilized instead of $\Delta_{\mathcal{L}}$, as suggested in equation (34), since the RSS/SSS is used to evaluate the convergence of the optimization procedure. **Figure 4** shows the RSS/SSS based on prior and posterior residuals, as well as for the training and validation subsets obtained with the GP-AR(72) model with basis orders $p = 1, \ldots, 6$. The obtained results demonstrate that a GP-AR(72) model with basis order $p = 4$ provides the best fit in all cases. Furthermore, the validation error results slightly elevated when compared against the training error, thus demonstrating the generalization capability of the model.

The prior and posterior estimates of the parameter vector as a function of the AWS are shown in **Figure 5**. The dependency of the parameter estimates on the AWS is evident and is consistent on both the prior and posterior parameter estimates. Nonetheless, in

FIGURE 2 | Typical 1-day variation of the PSD of the acceleration response and the average wind speed: **(A)** PSD (spectrogram)–*Hamming window, length 8,192 samples, overlap 5,792 samples*; **(B)** 2 min average wind speed.

some cases the posterior estimates tend to deviate from the prior at higher AWS values, especially for higher wind speeds. This effect may be due to the reduced amount of signal segments available for higher wind speeds in the construction of the model, as evident in the histogram of wind speeds shown in **Figure 1**.

4.1.3. Model-Based Analysis

Once the GP-AR$(72)_{p=4}$ model has been identified, it is possible to analyze the dynamic characteristics of the acceleration response of the bridge as a function of the average wind speed. In particular, the Power Spectral Density, the natural frequencies and damping ratios are extracted from the identified GP-AR$(72)_{p=4}$ model. Each one of these quantities are evaluated as follows:

$$\text{Characteristic polynomial}: A(z, \boldsymbol{\xi}) = 1 + \sum_{i=1}^{n_a} \hat{a}_i(\boldsymbol{\xi}) \cdot z^{-i} \tag{38a}$$

$$\text{PSD}: P_{yy}(f, \boldsymbol{\xi}) = \frac{\sigma_w^2}{|A(e^{j2\pi f}, \boldsymbol{\xi})|^2} \tag{38b}$$

$$\text{Poles}: \{\lambda_i(\boldsymbol{\xi}) \in \mathbb{C}, i = 1, \ldots, n_a : A(\lambda, \boldsymbol{\xi}) = 0\} \tag{38c}$$

$$\text{Natural frequencies}: f_{n,i}(\boldsymbol{\xi}) = \frac{f_s}{2\pi} \cdot |\ln \lambda_i(\boldsymbol{\xi})| \tag{38d}$$

$$\text{Damping ratios}: \zeta_i(\boldsymbol{\xi}) = -\cos(\arg(\ln \lambda_i(\boldsymbol{\xi}))) \tag{38e}$$

Figure 6 shows the GP-AR$(72)_{p=4}$ model-based PSD, natural frequencies and damping ratios obtained for the range of average wind speeds from 0 to 25 m/s, while frequencies and damping ratios are limited to the ranges [0,500] mHz and [0,10]%, respectively. The model-based PSD and modal quantities demonstrate that both the amplitude and frequency of the vibration are directly influenced by the average wind speed.

A more detailed analysis of the first six natural frequencies and damping ratios obtained from prior and posterior parameter estimates is presented in **Figure 7**. The posterior estimates are
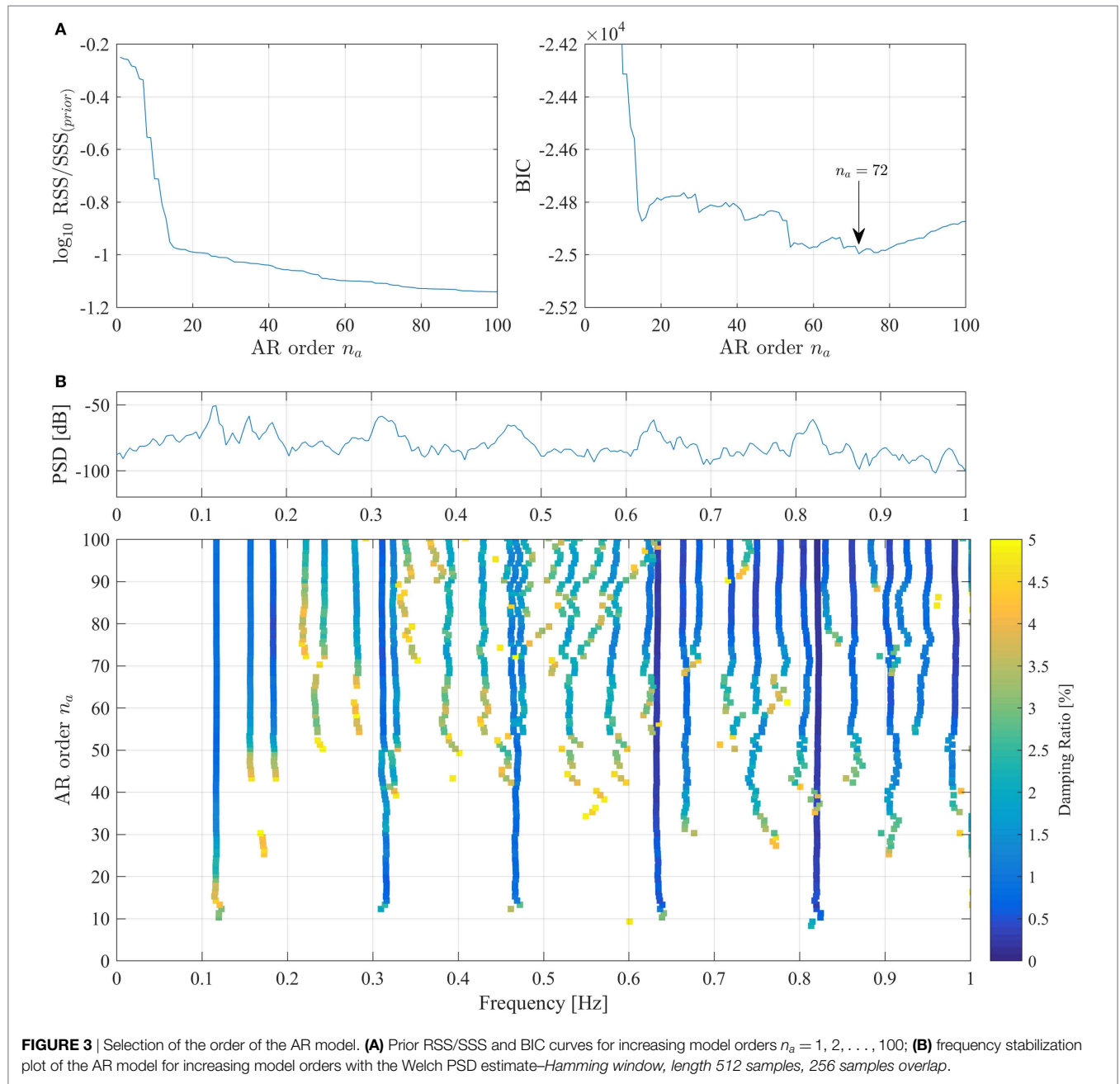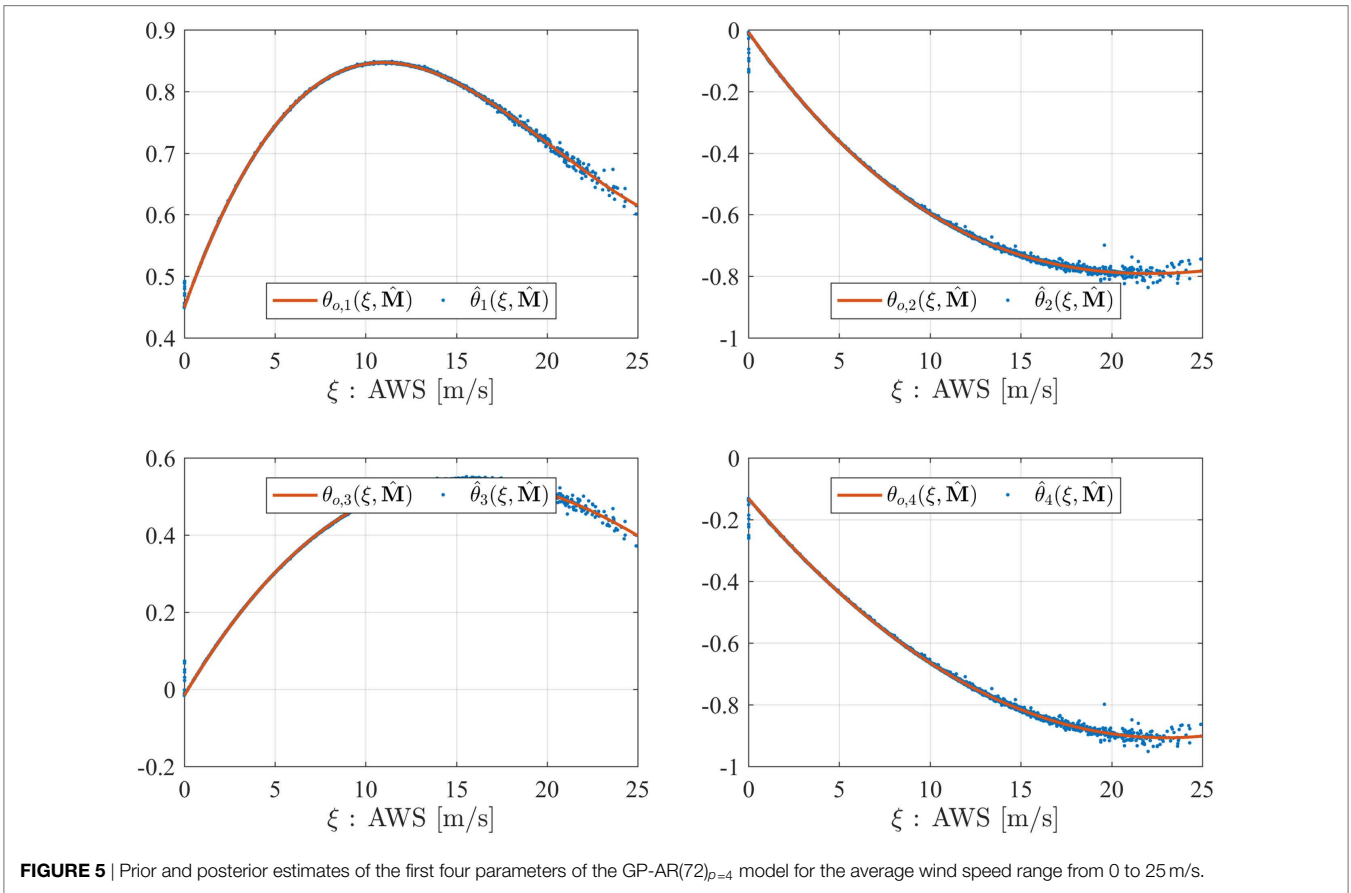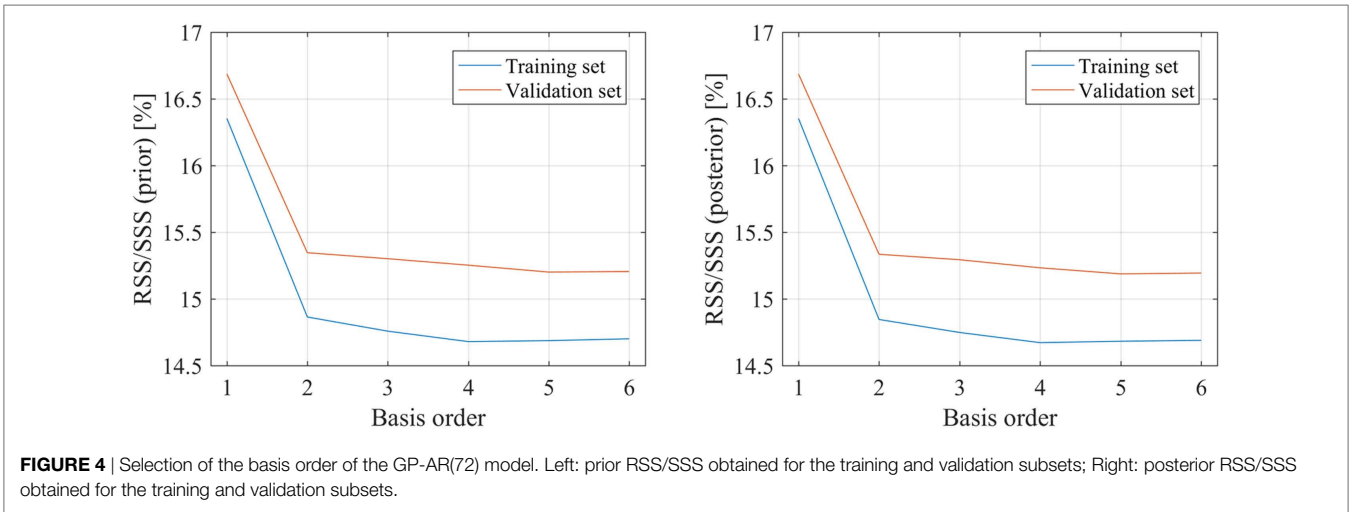
FIGURE 3 | Selection of the order of the AR model. **(A)** Prior RSS/SSS and BIC curves for increasing model orders $n_a = 1, 2, \ldots, 100$; **(B)** frequency stabilization plot of the AR model for increasing model orders with the Welch PSD estimate–*Hamming window, length 512 samples, 256 samples overlap*.

TABLE 1 | Settings of the EM algorithm.

Parameter	Value
Maximum number of iterations	$N_{iter} = 10^4$
Hyper-parameter difference threshold	$\Delta_p = 10^{-5}$
RSS/SSS difference threshold	$\Delta_{RSS/SSS} = 10^{-5}$
Initialization	Ordinary least squares estimates of the parameters of individual realizations

calculated for the complete set of vibration segments based on the estimated GP-AR model. The dispersion of the modal quantities tends to blow up for increasing wind speeds, which in part may be due to the effect of wind and turbulence on the structure. In

addition, the difference between prior and posterior estimates of the modal quantities tends to increase when the wind speed is larger than 20 m/s. This effect may be due to the lesser amount of samples acquired from higher wind speeds, which may be leading to increased uncertainty in the parameter estimates. The modal analysis results obtained with the GP-AR model may be contrasted with those previously reported on (Diana et al., 1992; Brownjohn et al., 1994, 2010). In particular, it appears that the modes displayed in **Figure 7** correspond to the first, second, third, and eight vertical modes ($f_{n,1}$, $f_{n,2}$, $f_{n,3}$, and $f_{n,5}$ respectively), and the first torsional mode of the bridge ($f_{n,4}$). Nonetheless, the predicted variation in the first torsional mode appears to be different to that one found with the GP-AR model.
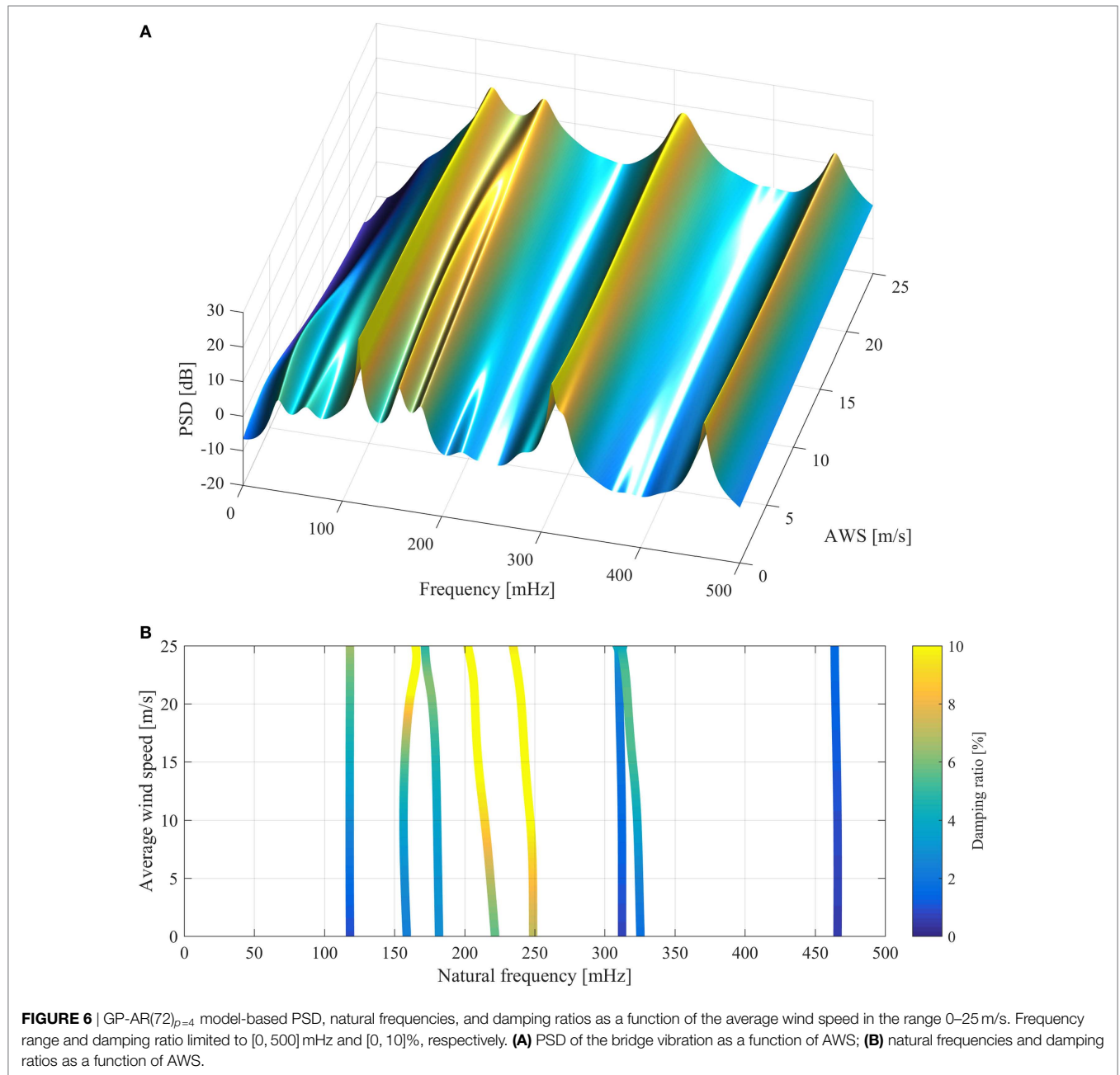
FIGURE 4 | Selection of the basis order of the GP-AR(72) model. Left: prior RSS/SSS obtained for the training and validation subsets; Right: posterior RSS/SSS obtained for the training and validation subsets.

FIGURE 5 | Prior and posterior estimates of the first four parameters of the GP-AR(72)$_{p=4}$ model for the average wind speed range from 0 to 25 m/s.

Confirmation of the modal analysis results shall be sought in a future work, where a vector AR model would be used to represent the two vertical and the horizontal vibration response of the bridge. However, the relatively simpler model utilized in this analysis can be used to track and assess the variability in the modes of the bridge when the wind is blowing from the main direction.

4.2. Simulated Vibration of Operating Wind Turbine Blades
4.2.1. Data Description and Preprocessing
This application focuses on the identification and analysis of the vibration acceleration signals obtained *via* simulations of a fully operational wind turbine. For a PCE-based treatment on actual tower measurements obtained from an operated wind turbine,

FIGURE 6 | GP-AR(72)$_{p=4}$ model-based PSD, natural frequencies, and damping ratios as a function of the average wind speed in the range 0–25 m/s. Frequency range and damping ratio limited to [0, 500] mHz and [0, 10]%, respectively. **(A)** PSD of the bridge vibration as a function of AWS; **(B)** natural frequencies and damping ratios as a function of AWS.

the interested reader is referred to Bogoevska et al. (2017). The analyzed wind turbine is the NREL 5 MW reference offshore wind turbine, fully described in Jonkman et al. (2009). The simulation is performed by means of the FAST wind turbine aeroelastic code, which uses a turbulent wind excitation simulated with TurbSim (Jonkman and Buhl, 2005). Acceleration signals are measured at different locations along the span of on one of the blades of the wind turbine on the flap wise direction, as depicted in **Figure 8**. From these, the acceleration signals measured on the tip of the blade (node 6) are used for further analysis.

Simulations of both turbulent wind and vibration response are computed for a period of 10 min (600 s) with a sampling rate of 200 Hz. Moreover, the instantaneous rotor azimuth is also extracted, which shall be used as the scheduling variable in the model for representation of the short-term response. The obtained acceleration signals are subsequently downsampled at 12.5 Hz for further analysis and processing. An antialiasing low-pass filter is applied before down-sampling. The filter consists of a 100 order FIR filter with cutoff frequency of 5 Hz. The cutoff frequency has been selected in order to preserve the structural modes which are under 4 Hz, and is applied in a forward–backward fashion to compensate the phase delay by using the MATLAB command `filtfilt`.

A set of 100 simulations is obtained under different 10 min average wind speeds in the range from 5 to 26 m/s. For that

FIGURE 7 | Distribution of selected natural frequencies and damping ratios obtained from the GP-AR(72)$_{p=4}$ model. Solid line, modal quantities from prior parameter estimates; dots, modal quantities from posterior parameter estimates; shaded areas, 50 and 90% confidence intervals derived from the parameter mean and covariance matrix. Left column: natural frequencies in mHz; right column: damping ratios.

purpose, the Latin hypercube sampling algorithm is used to create a set of 100 random average wind speeds within the specified range. Then, a turbulent wind speed time-series is simulated for each 10 min average wind speed, which is subsequently used as input to the FAST simulation software.

It is noted that the present simulated data bears important differences with those published in the previous work (Avendaño-Valencia and Fassois, 2017a), which are summarized as follows: (i) The turbulent wind excitation is simulated over an 8×8 grid, in comparison with the 2×2 grid used in the previous work. Therefore, the excitation is richer while the vibration response may be more complex. (ii) The sampling rate used for simulation is extended from 25 to 200 Hz so as to improve the convergence of the numerical integration algorithm. (iii) The analysis is performed in a sensor in the blade tip instead of the blade root. A consequence of the previous selections is that the structure of the obtained models may differ significantly with that reported (Avendaño-Valencia and Fassois, 2017a).

4.2.2. Model Identification
4.2.2.1. Modeling of the Short-term Response
The blade vibration signal is represented *via* a *Linear Parameter Varying AR* (LPV-AR) model, which uses the instantaneous rotor azimuth as scheduling variable (Avendaño-Valencia and Fassois, 2017a). The selection structure of the LPV-AR model follows the procedure described in Avendaño-Valencia and Fassois (2017a), while the selection of the LPV-AR model structure is guided by the BIC. The obtained BIC curves are shown in **Figure 9**, from which the model order $n_a = 17$ and basis order $p_a = 5$ are selected.

4.2.2.2. Modeling of the Long-term Response
A GP-LPV-AR model is constructed for the representation of the long term response of the blade. For this purpose, the EOP variable $\boldsymbol{\xi}$ is defined as the 10 minute average wind speed (lying in the range [0,30] m/s) normalized within the range [0,1]. The parameter vector of the model is expanded on the Hermite polynomials defined in equation (37).

GP-LPV-AR$(17)_5$ models with GP basis orders in the range $p = 0, 1, \ldots, 6$ are estimated using the EM algorithm. The settings of the EM algorithm are similar to those used in the previous example and summarized in **Table 1**; however, in the present case the thresholds for finalization of the optimization are defined as $\Delta_p = \Delta_{\mathrm{RSS/SSS}} = 10^{-6}$. **Figure 10A** shows the RSS/SSS based on the prior and posterior prediction errors for increasing orders of the functional basis expansion of the GP. The plot shows
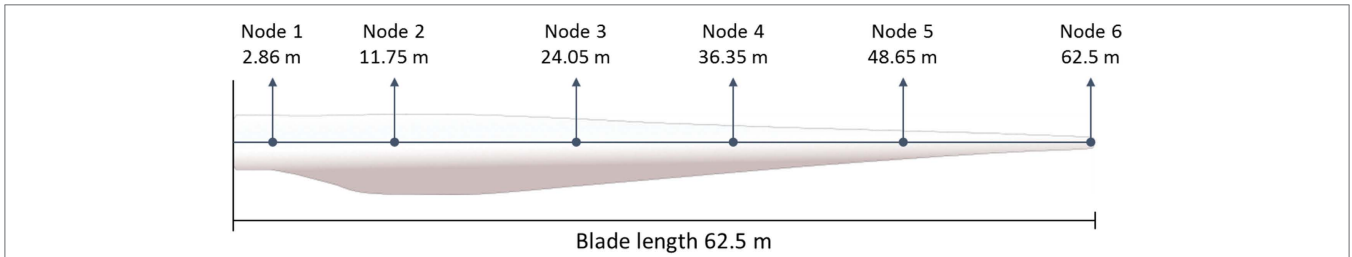
FIGURE 8 | Location of the sensors in the blade of the wind turbine. Acceleration signals are measured on the flapwise direction of the blade (normal to the surface of the page).
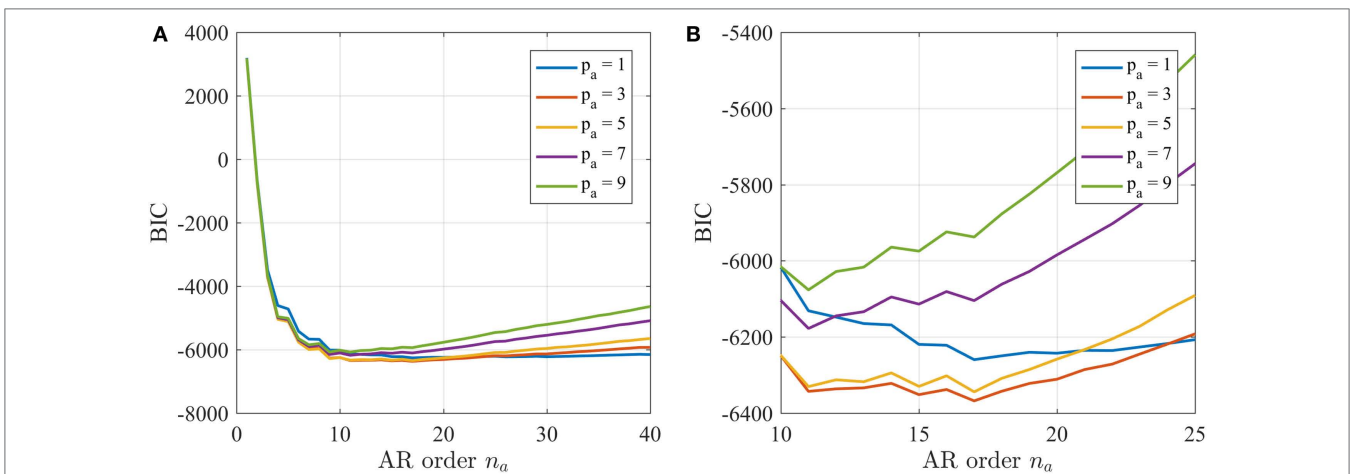
FIGURE 9 | Selection of the LPV-AR model structure: BIC curves for increasing model orders $n_a = 1, 2, \ldots, 40$ and for different basis orders $p_a = 1, 3, \ldots, 9$: **(A)** whole range; **(B)** detail in the range $n_a = [10, 25]$.
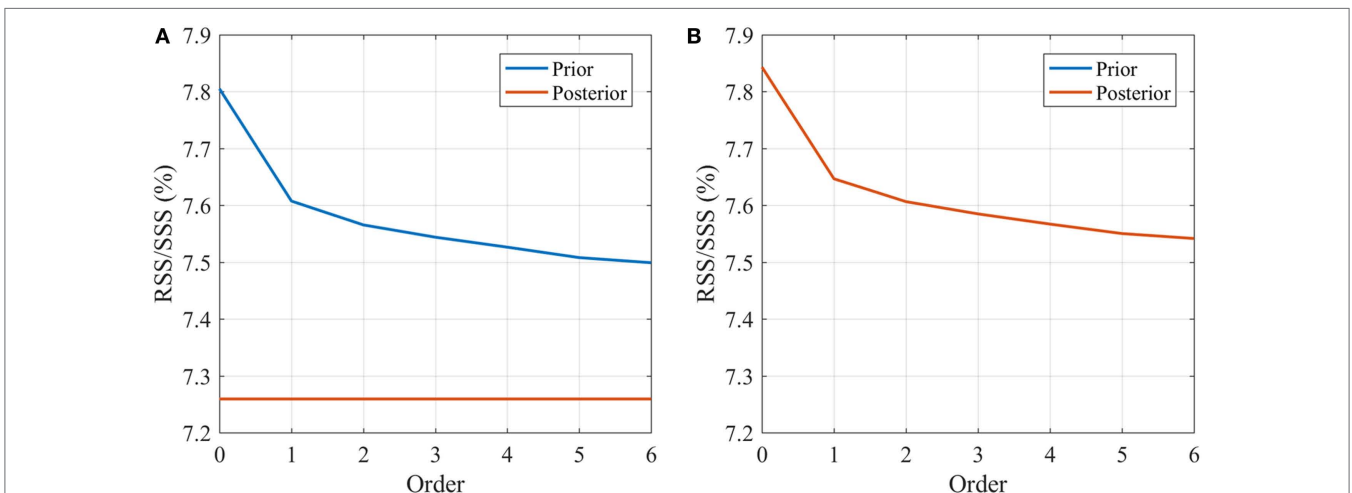
FIGURE 10 | Selection of the order of the GP-LPV-AR models: prior and posterior RSS/SSS curves for increasing GP basis orders $p = 0, 1, \ldots, 6$, for **(A)** GP-LPV-AR model; **(B)** GP-LPV-AR model with principal component regression.
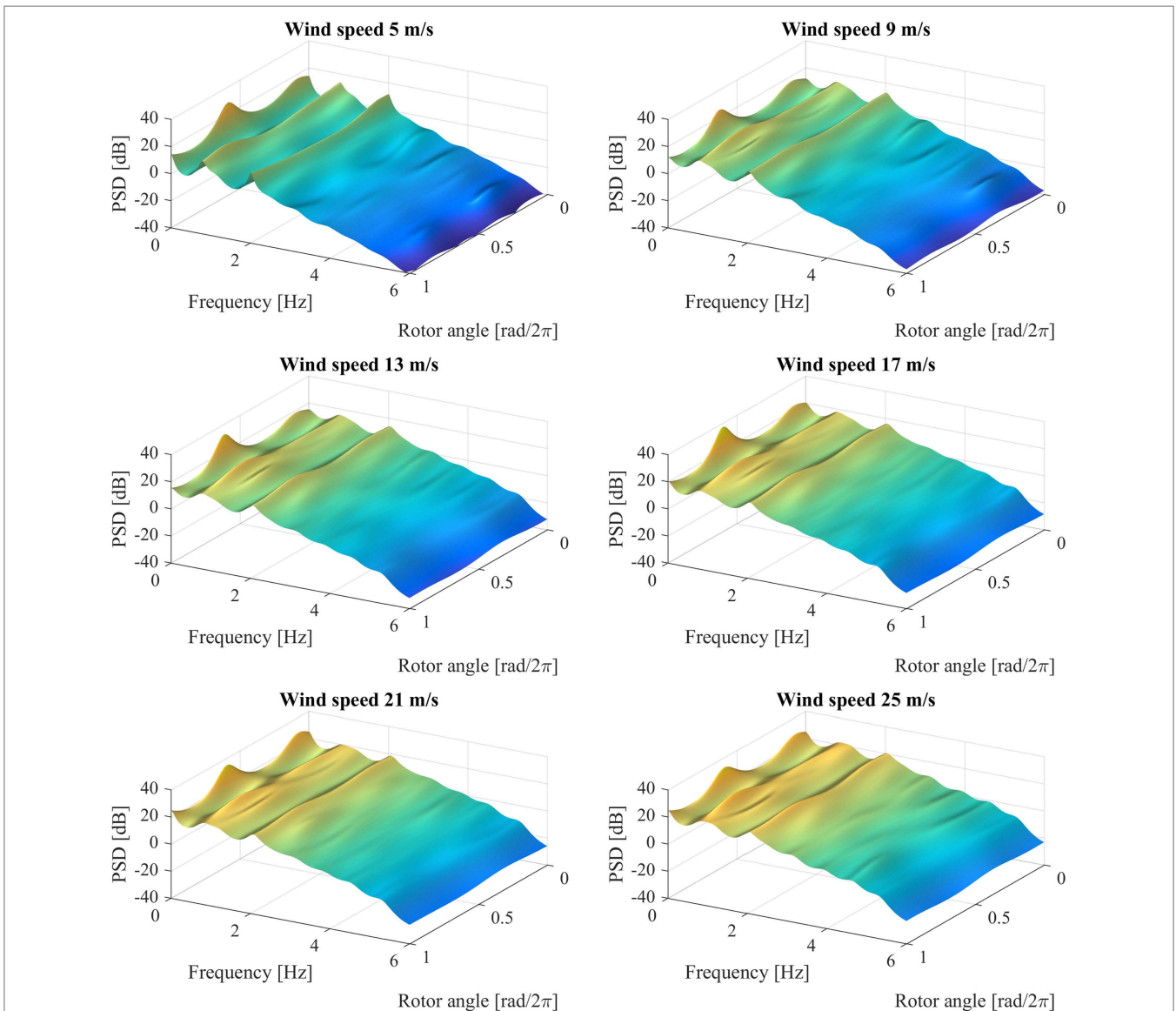
FIGURE 11 | GP-LPV-AR model-based "frozen" Power Spectral Density evaluated for a full rotation period of the wind turbine blade, and for increasing wind speeds.

totally different behaviors of the prior and posterior errors. In particular, it is evident that the prior prediction error depends on the order of the functional expansion, while the posterior error does not show a significant dependence. This behavior may be expected, since the prior estimates are based solely on the model predictions, while the posterior estimates are adjusted to the observed vibration response. For that same reason, the prior error is a better tool to evaluate the model capabilities. In the present case, according to the prior RSS/SSS curve, a basis order $p = 5$ is selected.

Similarly, a PC-GP-LPV-AR model (based on a Principal Component representation of the regression matrix of the LPV-AR model, as explained in Section 2.4) is used for the long-term identification of the response of the wind turbine. The PC-GP-LPV-AR model is further estimated by means of the EM

algorithm, using the same settings applied in the previous GP-LPV-AR model. The Principal Component representation of the regression matrix is carried out using all the components (without dimensionality reduction), however, the covariance matrix of the model parameters is in the present case, diagonal, thus a lower number of hyper-parameters have to be estimated. The obtained prior and posterior RSS/SSS curves obtained with the PC-GP-LPV-AR model are shown in **Figure 10B**. In this case, the prior and posterior error curves are almost the same, while the overall error is slightly higher than that obtained with the GP-LPV-AR model.

4.2.3. Model-Based Analysis
The dynamics of the wind turbine are analyzed based on the identified GP-LPV-AR(17)$_{5,5}$ model. For that purpose, the analysis of
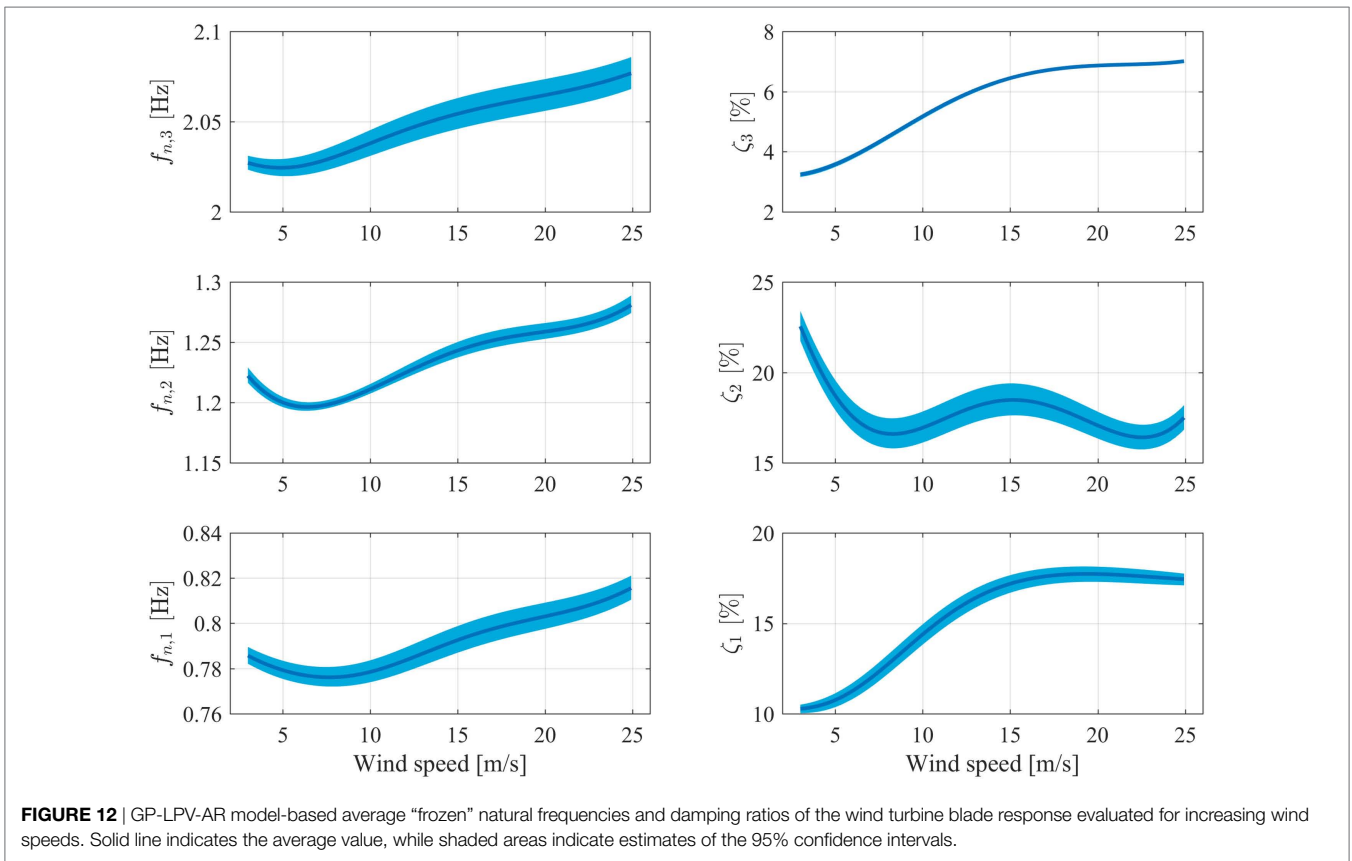
FIGURE 12 | GP-LPV-AR model-based average "frozen" natural frequencies and damping ratios of the wind turbine blade response evaluated for increasing wind speeds. Solid line indicates the average value, while shaded areas indicate estimates of the 95% confidence intervals.

the dynamics is based on the "frozen" Power Spectral Density, natural frequencies and damping ratios, which are calculated by means of the following equations:

"Frozen" characteristic polynomial :

$$A(z, \beta, \boldsymbol{\xi}) = 1 + \sum_{i=1}^{n_a} \hat{a}_i(\beta, \boldsymbol{\xi}) \cdot z^{-i} \qquad (39a)$$

$$\text{PSD} : P_{yy}(f, \beta, \boldsymbol{\xi}) = \frac{\sigma_w^2}{|A(e^{j2\pi f}, \beta, \boldsymbol{\xi})|^2} \qquad (39b)$$

$$\text{Poles} : \{\lambda_i(\beta, \boldsymbol{\xi}) \in \mathbb{C}, i = 1, \dots, n_a : A(\lambda, \beta, \boldsymbol{\xi}) = 0\} \quad (39c)$$

$$\text{Natural frequencies} : f_{n,i}(\beta, \boldsymbol{\xi}) = \frac{f_s}{2\pi} \cdot |\ln \lambda_i(\beta, \boldsymbol{\xi})| \qquad (39d)$$

$$\text{Damping ratios} : \zeta_i(\beta, \boldsymbol{\xi}) = -\cos(\arg(\ln \lambda_i(\beta, \boldsymbol{\xi}))) \qquad (39e)$$

For the present application, these quantities are functions of two variables, namely the instantaneous rotor angle β and the 10 minute average wind speed $\boldsymbol{\xi}$. In that sense, the "frozen" PSD, natural frequencies and damping ratios are calculated for a single period of rotation of the blades and for the whole range of wind speeds (3–25 m/s). **Figure 11** shows the obtained "frozen" PSDs for different wind speeds in the range. The figure indicates that the variability of the characteristics of the dynamic response of the wind turbine both as the rotor azimuth changes, and as the wind speed increases. In particular, it can be observed that for all the wind speeds, an important increment in the power of the vibration response is evident at about two-thirds of a full

rotation. This event may be associated with the blade passing in front of the tower. Moreover, it is also evident that as the wind speed increases, the overall power of the vibration response increases as well. The total power difference when the wind speed changes from 5 to 25 m/s is about 20 dB or a whole magnitude level.

Average values of the "frozen" natural frequencies and damping ratios and their respective confidence intervals may be drawn from the obtained GP-LPV-AR model. The procedure is similar to that performed in the previous application example. Three modes are selected for this analysis, namely the pair of modes located around 1 Hz, and the mode located around 2 Hz. The selected average "frozen" natural frequencies and damping ratios are shown in **Figure 12**. In the obtained curves is clear the dependency of both natural frequency and damping ratio on the wind speed. Particularly, for modes 1 and 3 there is an increase on the damping ratio, while for mode 2 the damping ratio tends to decrease as the wind speed does. In addition, the natural frequencies tend to increase with the wind speed as well. The confidence intervals are in general well bounded, and in particular it is clear that the estimates of the damping ratios are reliable.

5. CONCLUDING REMARKS

This work has been devoted to a Gaussian Process time-series modeling framework for the representation of the long-term dynamics of structures operating under variable environmental

and operational conditions. The model definitions plus an identification method based on the Expectation-Maximization algorithm have been presented. In addition, an optional parameter reduction technique based on Principal Component Regression has been introduced as a method to reduce the number of parameters to be estimated in the representation. The resulting GP time-series methods provide an appealing alternative for the representation of the complex dynamics of vibrating structures operating under variable environmental and operational conditions.

A potential limitation of the GP time-series modeling methodology, as presented in this work, is that the innovations variance of the time-series model is assumed to be constant. The adoption of a constant innovations variance may hinder the representation of changing power in the vibrational response of the structure according to environmental and operational conditions. A solution for this limitation is to define a GP to represent the dependence of the innovations variance on the EOPs, in the same manner as already explained for the coefficients of the time-series model. Toward this end, it would be further necessary to modify the EM algorithm for the estimation of the parameters of the innovations variance GP.

The proposed GP time-series modeling method offers a promising tool for assimilation in damage diagnosis algorithms. For this purpose, a key element lies in formalizing the selection of the conditions used for training of the model, namely, specification of the range of environmental and operational conditions under which the GP time-series model is able to lead to a robust decision. An exploratory study on this issue can be found in Avendaño-Valencia and Chatzi (2017), and will be extended as future work.

AUTHOR CONTRIBUTIONS

LA-V has developed the mathematical techniques involved in this work and performed the simulations shown in the case studies. He has further carried out the major part of the writing and organization of the text. EC has collaborated on the method's development, writing, and proof-reading of the document and has cross-checked the mathematical techniques. JB has provided the vibration data of the Humber bridge and has further contributed to the scientific writing. KK contributed to the revised version of the work by providing the complete bridge vibration data used in the application example on the Humber bridge. In addition, he participated in the revision of the manuscript pre-prints and in the approval of the finalized version.

FUNDING

REFERENCES

Avendaño-Valencia, L., and Chatzi, E. (2017). "Sensitivity driven robust vibration-based damage diagnosis under uncertainty through hierarchical bayes time-series representations," in *Proceedings of X International Conference on Structural Dynamics, EURODYN 2017*, Rome, Italy.

Avendaño-Valencia, L. D., Chatzi, E. N., and Spiridonakos, M. D. (2015a). "Non-stationary random coefficient models for vibration-based shm in structures influenced by strong operational and environmental variability," in *International Workshop on Structural Health Monitoring 2015*, Stanford, CA, USA.

Avendaño-Valencia, L. D., Chatzi, E. N., and Spiridonakos, M. D. (2015b). "Surrogate modeling of nonstationary systems with uncertain properties," in *Proceedings of the European Safety and Reliability Conference ESREL, 2015*, Zürich, Switzerland.

Avendaño-Valencia, L. D., and Fassois, S. D. (2015). "On multiple-model linear-parameter-varying based shm for time-dependent structures under uncertainty," in *International Conference Surveillance 8* (France: Roanne Institute of Technology).

Avendaño-Valencia, L. D., and Fassois, S. D. (2017a). Damage/fault diagnosis in an operating wind turbine under uncertainty via a vibration response Gaussian mixture random coefficient model based framework. *Mech. Syst. Signal Process.* 91, 326–353. doi:10.1016/j.ymssp.2016.11.028

Avendaño-Valencia, L. D., and Fassois, S. D. (2017b). Gaussian mixture random coefficient model based framework for SHM in structures with time-dependent dynamics under uncertainty. *Mech. Syst. Signal Process.* 97, 59–83. doi:10.1016/j.ymssp.2017.04.016

Bair, E., Hastie, T., Paul, D., and Tibshirani, R. (2006). Prediction by supervised principal components. *J. Am. Stat. Assoc.* 101, 119–137. doi:10.1198/016214505000000628

Basu, P., Rudoy, D., and Wolfe, P. J. (2009). "A nonparametric test for stationarity based on local Fourier analysis," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei, 3005–3008.

Bogoevska, S., Spiridonakos, M., Chatzi, E. N., Dumova-Jovanoska, E., and Höffer, R. (2017). A data-driven diagnostic framework for wind turbine structures: a holistic approach. *Sensors* 17, 1–28. doi:10.3390/s17040720

Borgnat, P., Flandrin, P., Honeine, P., Richard, C., and Xiao, J. (2010). Testing stationarity with surrogates: a time-frequency approach. *IEEE Trans. Signal Process.* 58, 3459–3470. doi:10.1109/TSP.2010.2043971

Brownjohn, J., Boccionlone, M., Curami, A., Falco, M., and Zasso, A. (1994). Humber bridge full-scale measurement campaigns 1990-1991. *J Wind Eng. Ind. Aerodyn.* 52, 185–218. doi:10.1016/0167-6105(94)90047-7

Brownjohn, J., Magalhaes, F., Caetano, E., and Cunha, A. (2010). Ambient vibration re-testing and operational modal analysis of the Humber Bridge. *Eng. Struct.* 32, 2003–2018. doi:10.1016/j.engstruct.2010.02.034

Deraemaeker, A., Reynders, E., Roeck, G. D., and Kullaa, J. (2008). Vibration-based structural health monitoring using output-only measurements under changing environment. *Mech. Syst. Signal Process.* 22, 34–56. doi:10.1016/j.ymssp.2007.07.004

Diana, G., Cheli, F., Zasso, A., Collina, A., and Brownjohn, J. (1992). Suspension bridge parameter identification in full scale test. *J. Wind Eng. Ind. Aerodyn.* 41-44, 165–176. doi:10.1016/0167-6105(92)90404-X

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning, Data Mining, Inference, and Prediction. Springer Series in Statistics.* New York: Springer.

Jonkman, J. M., and Buhl, M. L. (2005). *FAST User's Guide. Technical Report.* Battelle, CO: National Renewable Energy Laboratory, U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy.

Jonkman, J. M., Butterfield, S., Musial, W., and Scott, G. (2009). *Definition of a 5-MW Reference Wind Turbine for Offshore System Development. Technical Report.* Battelle, CO: National Renewable Energy Laboratory, U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy.

Kay, S. (2008). A new nonstationarity detector. *IEEE Trans. Signal Process.* 56, 1440–1451. doi:10.1109/TSP.2007.909346

Kopsaftopoulos, F., Nardari, R., Li, Y.-H., and Chang, F.-K. (2018). A stochastic global identification framework for aerospace structures operating under varying flight states. *Mech. Syst. Signal Process.* 98, 425–447. doi:10.1016/j.ymssp.2017.05.001

Ljung, L. (1999). *System Identification: Theory for the User*, 2nd Edn. Prentice Hall PTR.

Poulimenos, A., and Fassois, S. D. (2006). Parametric time-domain methods for non-stationary random vibration modeling and analysis: a critical survey and comparison. *Mech. Syst. Signal Process.* 20, 763–816. doi:10.1016/j.ymssp.2005.10.003

Rahbari, R., Niu, J., Brownjohn, J., and Koo, K.-Y. (2015). Structural identification of Humber Bridge for performance prognosis. *Smart Struct. Syst.* 15, 665–682. doi:10.12989/sss.2015.15.3.665

Rasmussen, C., and Williams, C. (2006). *Gaussian Processes for Machine Learning*. MIT Press.

Sakellariou, J. S., and Fassois, S. D. (2016). Functionally pooled models for the global identification of stochastic systems under different pseudo-static operating conditions. *Mech. Syst. Signal Process.* 72-73, 785–807. doi:10.1016/j.ymssp.2015.10.018

Shumway, R. H., and Stoffer, D. S. (2011). *Time Series Analysis and Its Applications*. New York: Springer.

Sohn, H. (2007). Effects of environmental and operational variability on structural health monitoring. *Philos. Trans. A Math. Phys. Eng. Sci.* 365, 539–560. doi:10.1098/rsta.2006.1935

Spiridonakos, M. D., and Chatzi, E. N. (2014). "Polynomial chaos expansion models for SHM under environmental variability," in *Proceedings of the 9th International Conference on Structural Dynamics (EURODYN 2014)*, Porto, Portugal, 2393–2398.

Spiridonakos, M. D., and Chatzi, E. N. (2015). Metamodeling of dynamic nonlinear structural systems through polynomial chaos NARX models. *Comput. Struct.* 157, 99–113. doi:10.1016/j.compstruc.2015.05.002

Spiridonakos, M. D., Chatzi, E. N., and Sudret, B. (2016). Polynomial Chaos expansion models for the monitoring of structures under operational variability. *ASCE-ASME J. Risk Uncertain. Eng. Syst. A Civ. Eng.* 2, B4016003. doi:10.1061/AJRUA6.0000872

Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory*. New York: Springer.

Worden, K., Sohn, H., and Farrar, C. (2002). Novelty detection in a changing environment: regression and interpolation approaches. *J. Sound Vibr.* 258, 741–761. doi:10.1006/jsvi.2002.5148

Yan, A.-M., Kerschen, G., Boe, P. D., and Golinval, J.-C. (2005). Structural damage diagnosis under varying environmental conditions. Part I: a linear analysis. *Mech. Syst. Signal Process.* 19, 847–864. doi:10.1016/j.ymssp.2004.12.003

APPENDIX

A. Demonstrations

A.1. Demonstration of the PDF in Equation (10)

This section aims to demonstrate the PDF of the dynamic response vector $\boldsymbol{y} = \begin{bmatrix} y[1] & \cdots & y[N] \end{bmatrix}^T$ given the parameter vector $\boldsymbol{\theta}$ and the regression matrix $\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\phi}(\boldsymbol{z}[1]) & \cdots & \boldsymbol{\phi}(\boldsymbol{z}[N]) \end{bmatrix}^T$. For that purpose, consider the PDF:

$$p(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{\Phi}) = \prod_{t=1}^{N} p(y[t] \mid \boldsymbol{\theta}, \boldsymbol{\phi}(\boldsymbol{z}[t])) \tag{A1}$$

substituting equation (8) and displaying the Gaussian distribution as an exponential function, yields:

$$p(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{\Phi}) = \prod_{t=1}^{N} \mathcal{N}_{y[t]}(y[t|t-1], \sigma_w^2)$$

$$= \prod_{t=1}^{N} (2\pi\sigma_w^2)^{-1/2} \cdot \exp\left(-\frac{1}{2\sigma_w^2}(y[t] - \boldsymbol{\phi}^T(\boldsymbol{z}[t]) \cdot \boldsymbol{\theta})^2\right) \tag{A2}$$

applying the product operator, yields:

$$p(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{\Phi})$$

$$= (2\pi\sigma_w^2)^{-N/2} \cdot \exp\left(-\frac{1}{2\sigma_w^2}\sum_{t=1}^{N}(y[t] - \boldsymbol{\phi}^T(\boldsymbol{z}[t]) \cdot \boldsymbol{\theta})\right)$$

$$= (2\pi)^{-N/2} \cdot |\sigma_w^2 \cdot \boldsymbol{I}_N|^{-1/2} \cdot \exp\left(-\frac{1}{2\sigma_w^2}\sum_{t=1}^{N}(y[t] - \boldsymbol{\phi}^T(\boldsymbol{z}[t]) \cdot \boldsymbol{\theta})^2\right) \tag{A3}$$

Then, operating in the sum inside of the exponential function, leads to:

$$\sum_{t=1}^{N}(y[t] - \boldsymbol{\phi}^T(\boldsymbol{z}[t]) \cdot \boldsymbol{\theta})^2$$

$$= \sum_{t=1}^{N}(y[t] - \boldsymbol{\phi}^T(\boldsymbol{z}[t]) \cdot \boldsymbol{\theta}) \cdot (y[t] - \boldsymbol{\phi}^T(\boldsymbol{z}[t]) \cdot \boldsymbol{\theta})$$

$$= \begin{bmatrix} y[1] - \boldsymbol{\phi}^T(\boldsymbol{z}[1]) \cdot \boldsymbol{\theta} \\ y[2] - \boldsymbol{\phi}^T(\boldsymbol{z}[2]) \cdot \boldsymbol{\theta} \\ \vdots \\ y[N] - \boldsymbol{\phi}^T(\boldsymbol{z}[N]) \cdot \boldsymbol{\theta} \end{bmatrix}^T \cdot \begin{bmatrix} y[1] - \boldsymbol{\phi}^T(\boldsymbol{z}[1]) \cdot \boldsymbol{\theta} \\ y[2] - \boldsymbol{\phi}^T(\boldsymbol{z}[2]) \cdot \boldsymbol{\theta} \\ \vdots \\ y[N] - \boldsymbol{\phi}^T(\boldsymbol{z}[N]) \cdot \boldsymbol{\theta} \end{bmatrix}$$

$$= \left(\begin{bmatrix} y[1] \\ y[2] \\ \vdots \\ y[N] \end{bmatrix} - \begin{bmatrix} \boldsymbol{\phi}^T(\boldsymbol{z}[1]) \\ \boldsymbol{\phi}^T(\boldsymbol{z}[2]) \\ \vdots \\ \boldsymbol{\phi}^T(\boldsymbol{z}[N]) \end{bmatrix} \cdot \boldsymbol{\theta}\right)^T \cdot \left(\begin{bmatrix} y[1] \\ y[2] \\ \vdots \\ y[N] \end{bmatrix} - \begin{bmatrix} \boldsymbol{\phi}^T(\boldsymbol{z}[1]) \\ \boldsymbol{\phi}^T(\boldsymbol{z}[2]) \\ \vdots \\ \boldsymbol{\phi}^T(\boldsymbol{z}[N]) \end{bmatrix} \cdot \boldsymbol{\theta}\right) \tag{A4}$$

and thus:

$$\sum_{t=1}^{N}(y[t] - \boldsymbol{\phi}^T(\boldsymbol{z}[t]) \cdot \boldsymbol{\theta})^2 = \left(\boldsymbol{y} - \boldsymbol{\phi}^T \cdot \boldsymbol{\theta}\right)^T \cdot \left(\boldsymbol{y} - \boldsymbol{\Phi}^T \cdot \boldsymbol{\theta}\right) \tag{A5}$$

Then, after putting everything together, the following result is obtained:

$$p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\Phi}) = (2\pi)^{-N/2} \cdot |\sigma_w^2 \cdot \boldsymbol{I}_N|^{-1/2}$$

$$\cdot \exp\left(-\frac{1}{2\sigma_w^2}\left(\boldsymbol{y} - \boldsymbol{\Phi}^T \cdot \boldsymbol{\theta}\right)^T \cdot \left(\boldsymbol{y} - \boldsymbol{\Phi}^T \cdot \boldsymbol{\theta}\right)\right)$$

$$= \mathcal{N}_y(\boldsymbol{\Phi}^T \cdot \boldsymbol{\theta}, \sigma_w^2 \cdot \boldsymbol{I}_N) \tag{A6}$$