Quantification and reduction of uncertainties in a wind turbine numerical model based on a global sensitivity analysis and a recursive Bayesian inference approach

Adrien Hirvoas, Clémentine Prieur, Élise Arnaud, Fabien Caleyron, Miguel Munoz Zuniga

RESEACH ARTICLE

Quantification and reduction of uncertainties in a wind turbine numerical model based on a global sensitivity analysis and a recursive Bayesian inference approach.

Adrien Hirvoas[1] | **Clémentine Prieur**[2] | **Elise Arnaud**[2] | **Fabien Caleyron**[1] | **Miguel Munoz Zuniga**[1]

[1]IFP Energies Nouvelles

[2]LJK — Laboratoire Jean Kuntzmann

Correspondence
IFP Energies Nouvelles — Lyon site,
Rond-point de l'échangeur de Solaize BP 3,
69360 Solaize, France
Email: adrien.hirvoas@ifpen.fr

Funding information
This research was funded by IFP Energies nouvelles in the scope of its framework agreement with the French Institute for Research in Computer Science and Automation (INRIA).

A framework to perform quantification and reduction of uncertainties in a wind turbine numerical model using a global sensitivity analysis and a recursive Bayesian inference method is developed in this paper. We explain how a prior probability distribution on the model parameters is transformed into a posterior probability distribution, by incorporating a physical model and real field noisy observations. Nevertheless, these approaches suffer from the so-called curse of dimensionality. In order to reduce the dimension, Sobol' indices approach for global sensitivity analysis, in the context of wind turbine modeling, is presented. A major issue arising for such inverse problems is identifiability, i.e. whether the observations are sufficient to unambiguously determine the input parameters that generated the observations. Global sensitivity analysis is also used in the context of identifiability.

KEYWORDS
Inference, model calibration, nonlinear Kalman filters, GSA, UQ

Abbreviations: UQ, Uncertainty Quantification; GSA, Global Sensitivity Analysis; EnKF, Ensemble Kalman Filter.

Introduction and Scope

In the current profound worldwide energy transition, wind power generation is developing rapidly. As a consequence, wind turbines monitoring, performance optimization and lifetime assessment are becoming major issues. In the context of digitalization of the industry, the exploitation of collected data can be optimized by combination with wind turbine numerical models. Such numerical models can be complex and costly as they involve nonlinear dynamic equations with different physics as well as stochastic loading from the wind. Moreover, some input parameters of the models can be poorly or badly known as the structure ages over time and defaults can appear. Consequently, model predictions are affected by these uncertainties. Characterization and reduction of these uncertainties is important for decision making [1]. It is the case in wind energy applications where uncertainties are omnipresent both in external conditions and in the models used during design process. In this context, uncertainty quantification and reduction methods have been developed [2]. As mentioned by Hart *et al.* [3], , even the concept of sensitivity is delicate when dealing with stochastic models, as the one in our industrial application whose stochasticity is due to the stochastic nature of the wind external solicitation. Note also that the models used in wind energy applications are often time consuming [4, 5]. Therefore most of the commonly used methodologies for uncertainty quantification are inappropriate in our setting. An historical strategy for uncertainty quantification in wind energy fields was to take into account uncertainties by employing Monte-Carlo methods [6, 7]. Nevertheless, for high-fidelity numerical models, such uncertainty quantification approaches based on Monte-Carlo methods become cumbersome due to the computational cost. Advanced methods such as polynomial chaos expansion, stochastic collocation or Gaussian process have been developed to alleviate this computational issue, see [8, 9, 10].

In the present work, we aim at investigating a complete framework to quantify and reduce the input parameter uncertainties involved in a finite element wind turbine model. Such methods to reduce uncertainties involved in models used in wind energy are currently investigated [11]. In the literature, Van Buren *et al.* have developed a framework to quantify and reduce such uncertainties based on ANOVA decomposition and Bayesian inference [12]. However, our approach is innovative due to the fact that we are dealing with a high-fidelity wind turbine numerical model and by the recursive aspect of our inference procedure. Recently, similar research work on recursive inference with a low-fidelity wind turbine numerical model has been led by Branlard *et al.*, see [13]. They develop a digital twin concept in order to estimate turbine states; e.g. wind speed, torque; based on the Kalman filter. Our main contribution is twofold.

Firstly, the framework allows quantifying the sources of uncertainties affecting the fatigue behavior of the structural components of the wind turbine. To perform such analysis, we usually use an aero-servo-elastic software fed by model parameters and wind solicitation. Each wind field is computed using a stochastic turbulent wind simulator. The obtained outputs of the simulations are different time-series describing the behavior of the wind turbine, which are reduced to some quantities of interest (QoIs), see Fig. 1. The function $\mathcal{M}$ representing the time-consuming numerical model is defined as:

$$\mathbf{y} = \mathcal{M}(\mathbf{x}, V)\,, \tag{1}$$

where, $\mathbf{x} = (x_1, ..., x_p) \in \mathcal{P} \subseteq \mathbb{R}^p$ are the model input parameters, V is a stochastic process modeling the external wind solicitation, $\mathbf{y} = (y_1, ..., y_m) \in \mathbb{R}^m$ is the vector of discretized functional output; e.g., generated power, structural accelerations or loads. Let g be the function mapping the functional loads of the structure in $\mathbf{y}$ to the damage quantity of interest (QoI), such as the damage-equivalent load (DEL) [14]:

$$g(\mathbf{y}) := g \circ \mathcal{M}(\mathbf{x}, V)\,. \tag{2}$$

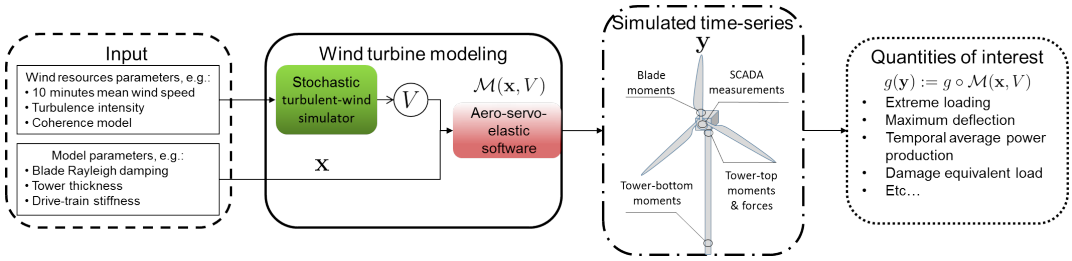


FIGURE 1 General sketch for wind turbine modeling.

Global Sensitivity Analysis (GSA) methods have been developed to quantify the uncertainty in QoI w.r.t. the input parameters, their individual contributions, or the contribution of their interactions. We propose a variance-based GSA methodology, relied on the so-called Sobol' indices [15], for stochastic computer simulations. Such techniques, which often refer to the probabilistic framework and Monte-Carlo (MC) methods, require a lot of calls to the numerical model. The uncertain input parameters are modeled by independent random variables gathered into a random vector and characterized by their probability density function. Variance-based SA for time consuming deterministic computer models has been mainly performed by approximating the model by a mathematical function, a.k.a. a surrogate model. Among the different surrogate models, we focus on Gaussian process (GP), a.k.a. Kriging, which is characterized by its mean and covariance functions. One advantage of the GP model is to provide both a prediction of the numerical model and the associated uncertainty. Such surrogate modeling assumes that prior beliefs about the numerical code can be modeled by a Gaussian process. Nowadays, in industrial applications, numerical models are often run at different levels of complexity and then a hierarchy of simulations is available. In this context, if several resolutions of simulation are obtained, multi-fidelity GP regression has been proposed to predict the output of a costly numerical model, see [16, 17]. In particular, Parussini and Perdikaris have proposed a recursive formulation of the approach combined with other approximations to enhance the computational efficiency, see [18, 19, 20]. These multi-fidelity formulations of the GP regression are computing time efficient in terms of number of model evaluations in comparison to the simple GP regression. However in the present industrial application, we did not implement multi-fidelity formulations of GP regression as we could only run a high-resolution version of the stochastic simulator $\mathcal{M}$. In order to take into account the inherent randomness from wind turbine simulation, our approach consists in focusing on the mean behavior of the high-resolution runs of the stochastic simulator described by the deterministic model $\mathbf{x} \mapsto \mathbb{E}_V[g \circ \mathcal{M}(\mathbf{x}, V)]$. GP regression is then used to reduce the numerical costs. More precisely, noisy evaluations of the conditional expectation are computed via Monte-Carlo and then filtered using heteroskedastic GP modeling. Lastly, variance-based sensitivity indices are computed by running the GP based surrogate in a so-called pick-freeze estimation procedure [21].

Secondly, after identification of the less influential input parameters on the fatigue behavior of the wind turbine, we propose a Bayesian inference framework to carry out a model calibration procedure based on in situ-measurements. It uses measurements $\mathbf{y}^{mes}$ to update some prior probability density functions about the unknown input parameters $\mathbf{X} \sim p(\mathbf{x})$ and yields some posterior probability density functions, through the Bayes' theorem $p(\mathbf{x}|\mathbf{y}^{mes}) \propto p(\mathbf{y}^{mes}|\mathbf{x})p(\mathbf{x})$ [1]. Numerous batch techniques have been developed to solve such Bayesian problems. Nevertheless, recent decades have been marked by a simultaneous development of sensor technologies and Internet of Things capabilities. Thus,

[1] Random variables are written in upper case roman letters and particular realizations of a random variable are written in corresponding lower case letters.

our research efforts have been directed towards inference techniques where the data are sequentially processed when new observations become available. In this context, the model parameter inference can be carried out using sequential Bayesian techniques. In geosciences, these techniques are called data-assimilation methods. We perform the calibration using a recursive Bayesian inference approach based on an Ensemble Kalman Filter (EnKF) [22].

However, such recursive inverse problems can be solved assuming that several conditions of well-posedness and identifiability are achieved. These conditions have been summarized by Hadamard [23]. As highlighted in [24], a relationship between the non-identifiability of input parameters and the GSA can be established. Indeed, input parameters with null total sensitivity indices on the measured outputs imply their non-identifiability. Therefore, for the purpose of identifiability a second GSA is conducted on the calibration parameters. However, due to the functional behavior of the measurements, we propose to first reduce their dimensionality through principal component analysis. Then, a GP is fitted to the different principal components and used to compute an aggregated Sobol' index for each model parameter [25].

Last but not least, the proposed framework has been applied to an industrial wind turbine numerical model. The developed recursive inference procedure has shown promising results in the industrial inversion problem.

The paper is organized as follows. The GSA methodology in the context of stochastic time-consuming numerical model is introduced in Section 1. In Section 2, we explore how the EnKF can be employed in model calibration problems. Section 3 is devoted to present the wind turbine numerical model, its uncertain input parameters and the selected output quantities used for quantifying and reducing the uncertainties. In Section 4, a wind turbine numerical case study is used to illustrate the proposed framework, and its performance in calibrating parameters with noisy pseudo-experimental output data.

1 | KRIGING BASED GLOBAL SENSITIVITY ANALYSIS

The aim of sensitivity analysis is to quantify the relative influence of input parameters on some QoI produced from the model outputs of the numerical model. In the context of model calibration, conducting such an analysis can help to identify which input parameters should be properly estimated. One may distinguish two categories of methods: local and global. While local sensitivity analysis considers small perturbations of the inputs around some nominal values, global sensitivity analysis (GSA) varies the inputs on their whole variation range [26]. Among the large number of available approaches, variance-based sensitivity analysis introduced by [27] proposes to measure the sensitivity by computing the so-called Sobol' indices. When no analytical formulae of these indices are available, one way to perform their estimation is to rely on Monte-Carlo (MC) techniques, which require a huge number of model evaluations. In the context of costly numerical codes as, e.g., the wind turbine numerical model under interest, the use of a cheap metamodel in place of the true costly model is thus crucial. In addition to being computationally expensive, the numerical model we are dealing with is stochastic. This means that from a same set of input parameters, the output can have different values depending on the wind conditions. This specificity has to be carefully taken into account when estimating sensitivity measures under interest. More precisely, let us use the formalism introduced in Eq. (1) and (2) to the model in hand. We are interested in measuring the sensitivity of a QoI $g(\mathbf{y}) \in \mathbb{R}$ with respect to the input $\mathbf{x}$. In the context of GSA, each input is now considered as a random variable X_j with its uncertainty modelled by a probability distribution, such as $\mathbf{X} = (X_1, \ldots, X_p)$. These one-dimensional probability distributions reflect the practitioner's belief in the uncertainty on the parameter values and the X_j are assumed to be independent from each other. Then, the QoI $g(\mathbf{Y}) := g \circ \mathcal{M}(\mathbf{X}, V)$ is a random variable itself.

The randomness of the QoI has two sources: the randomness from the parameters $\mathbf{X}$, and the one due to the stochas-

ticity propagated from the model itself through V, which is assumed to be independent of $\mathbf{X}$. There exists at least two approaches to deal with this stochasticity in a GSA framework. The first one considers the full probability distribution of the QoI while the other one is only concerned with quantitative measures of the probability distribution, e.g. quantile or expectation [28]. The latter is the one considered in this work by investigating the QoI averaged over the inherent randomness of the physical system. We are therefore interested in the sensitivity of the deterministic function f defined as:

$$f(\mathbf{X}) = \mathbb{E}_V\left[g \circ \mathcal{M}(\mathbf{X}, V)\right]. \tag{3}$$

The total variance of $f(\mathbf{X})$ can be split into different parts of variance under the assumption that the input parameters are independent (this is the so-called Hoeffding decomposition, see [29]). Each part of variance corresponds to the contribution of each set of parameters on the variance of the output $f(\mathbf{X})$. By considering the ratio of each part of variance to the total variance, we obtain a measure of importance for each set of input parameters that is called the Sobol' index [27]. In the literature, functional analysis of variance (FANOVA) has been widely used to quantify the sensitivity of a model output to input variables (see Appendix A in [30] and [31]). Let us denote $\mathbf{u}$ a subset of $\{1, ..., p\}$, $-\mathbf{u}$ its complement, and $|\mathbf{u}|$ its cardinality. Assuming $Var_{\mathbf{X}}(f(\mathbf{X})) < +\infty$, $Var_{\mathbf{X}}(f(\mathbf{X})) \neq 0$, we define for any $\mathbf{u}$ the closed Sobol' index of order $r = |\mathbf{u}|$ associated to the set of inputs $\mathbf{X_u} = \{X_j\}_{j\in\mathbf{u}}$ as:

$$S_{\mathbf{u}} = \frac{Var_{\mathbf{X_u}}(\mathbb{E}_{\mathbf{X_{-u}}}[f(\mathbf{X})|\mathbf{X_u}])}{Var_{\mathbf{X}}(f(\mathbf{X}))}. \tag{4}$$

This index quantifies the main effect of all the variables within $\mathbf{X_u}$, including interactions, on $f(\mathbf{X})$. The most influential sets of input parameters can then be identified as the sets of input parameters with the largest Sobol' indices. Total Sobol' indices can also been defined as:

$$S_{\mathbf{u}}^T = 1 - S_{-\mathbf{u}} = 1 - \frac{Var_{\mathbf{X_{-u}}}(\mathbb{E}_{\mathbf{X_u}}[f(\mathbf{X})|\mathbf{X_{-u}}])}{Var_{\mathbf{X}}(f(\mathbf{X}))} \tag{5}$$

This index quantifies the effect of $\mathbf{X_u}$ plus the effect of all interactions between variables in $\mathbf{X_u}$ and variables in $\mathbf{X_{-u}}$ on $f(\mathbf{X})$.

A general approach to estimate Sobol' indices is based on Monte-Carlo and requires an important number of evaluations of f. The high computational cost of the wind turbine model prevents from performing such estimation in reasonable time. It is the reason why we would like to rely on a metamodel to compute cheap evaluations of the initial costly computer code. In this study, we chose to approximate the true numerical code by a Gaussian process, a.k.a. Kriging metamodel, in order to apply a Kriging based sensitivity analysis, e.g., in [32]. We firstly present the noiseless framework and then we detail the case where we are facing to heterogeneously noisy evaluations of the function f.

Ordinary Kriging - First, a Gaussian process regression model is built to surrogate the function f. The principle of Kriging based metamodeling [33] is to consider that our deterministic model f can be considered as a realization of a Gaussian process $\{Z(\mathbf{x}),\ \mathbf{x} \in \mathcal{P}\}$ with mean function μ and covariance kernel C. Such covariance kernel (a.k.a. covariance function, kernel function, or kernel), is a positive-definite function of two distinct inputs $\mathbf{x}, \mathbf{x'}$ allowing to define the prior covariance between any two values of the function of interest. Many kernels can be used, each one corresponding to a different set of prior assumptions made about the function of interest [34, 35, 36]. Each kernel is defined by a number of parameters which specify the shape of the covariance function. These parameters, a.k.a. hyper-parameters, can be either estimated by minimizing a loss function with a Leave-One-Out Cross-Validation procedure or maximizing a likelihood function [37]. In our study, we use the last mentioned approach to estimate these

hyper-parameters of a 5/2 Matérn kernel.

For any $\mathbf{x} \in \mathcal{P}$, $f(\mathbf{x})$ is approximated by the conditional Gaussian process $\{Z_n(\mathbf{x}),\ \mathbf{x} \in \mathcal{P}\} := \{[Z(\mathbf{x})|Z(\mathbf{X}^n) = \mathbf{z}],\ \mathbf{x} \in \mathcal{P}\}$, where $\mathbf{z} = \{z_1, \dots z_n\}$ are evaluations of f on n points $\mathbf{X}^n = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, $\mathbf{x}_i \in \mathcal{P}$. The design $\{(\mathbf{x}_1, z_1), \dots, (\mathbf{x}_n, z_n)\}$ is called the learning sample. In the following, $\mathbf{X}^n$ is chosen as a Latin Hypercube sample [38] to guarantee a good exploration of our numerical model. We then get the ordinary Kriging equations:

$$Z_n(\mathbf{x}) \sim \mathcal{N}(m_{OK}(\mathbf{x}), s^2_{OK}(\mathbf{x}))\,, \tag{6}$$

with

$$m_{OK}(\mathbf{x}) = \mu(\mathbf{x}) + \mathbf{c}(\mathbf{x})^T \mathbf{C}^{-1}(\mathbf{z} - \boldsymbol{\mu})\,, \tag{7}$$

$$s^2_{OK}(\mathbf{x}) = C(\mathbf{x}, \mathbf{x}) - \mathbf{c}(\mathbf{x})^T \mathbf{C}^{-1} \mathbf{c}(\mathbf{x})\,. \tag{8}$$

We denote by $\boldsymbol{\mu} = \mu(\mathbf{X}^n)$ the vector of trend values on $\mathbf{X}^n$, by $\mathbf{C} = (C(\mathbf{x}_i, \mathbf{x}_j))_{1 \leq i,j \leq n}$ the covariance matrix of $Z(\mathbf{X}^n)$, and by $\mathbf{c}(\mathbf{x}) = (C(\mathbf{x}, \mathbf{x}_i))_{1 \leq i \leq n}$ the vector of covariances between $Z(\mathbf{x})$ and $Z(\mathbf{X}^n)$.

Noisy Kriging - In our context, exact evaluations of f can not be obtained directly. We have, for each $i = 1, \dots, n$, a noisy evaluation $\tilde{z}_i = f(\mathbf{x}_i) + \varepsilon_i$, where $\tilde{z}_i$ is defined as an empirical mean computed from a K-sample of $g \circ \mathcal{M}(\mathbf{x}_i, V)$, and ε_i is a centered noise whose variance is $\tau_i^2 = \frac{1}{K}\left(\frac{1}{K-1}(\sum_{j=1}^K (\mathcal{M}(\mathbf{x}_i, V = v_j) - \frac{1}{K}\sum_{j=1}^K \mathcal{M}(\mathbf{x}_i, V = v_j))^2\right)$. We then consider, as a first approximation, that the vector $(\varepsilon_1, \dots, \varepsilon_n)$ is a centered Gaussian random vector with diagonal covariance matrix $\mathrm{diag}(\tau_1^2, \dots, \tau_n^2)$ denoted by $\boldsymbol{\Delta}$. Then, provided that the process Z and the Gaussian measurement errors $\{\varepsilon_i\}_{1 \leq i \leq n}$ are stochastically independent, the process conditionally on the noisy observations $\{\tilde{Z}_n(\mathbf{x}),\ \mathbf{x} \in \mathcal{P}\} := \{[Z(\mathbf{x})|\tilde{Z}(\mathbf{X}^n) = \{\tilde{z}_i\}_{1 \leq i \leq n}],\ \mathbf{x} \in \mathcal{P}\}$ is still Gaussian, and its conditional mean and variance functions are given by the following slightly modified kriging equations:

$$\tilde{Z}_n(\mathbf{x}) \sim \mathcal{N}(m_{NK}(\mathbf{x}), s^2_{NK}(\mathbf{x}))\,, \tag{9}$$

with

$$m_{NK}(\mathbf{x}) = \mu(\mathbf{x}) + \mathbf{c}(\mathbf{x})^T (\mathbf{C} + \boldsymbol{\Delta})^{-1} (\tilde{\mathbf{z}} - \boldsymbol{\mu})\,, \tag{10}$$

$$s^2_{NK}(\mathbf{x}) = C(\mathbf{x}, \mathbf{x}) - \mathbf{c}(\mathbf{x})^T (\mathbf{C} + \boldsymbol{\Delta})^{-1} \mathbf{c}(\mathbf{x})\,. \tag{11}$$

Kriging based Sobol' indices - Following [32] and [39] the idea is to substitute f with $\tilde{Z}_n$ in Eq. (4):

$$\tilde{S}_{\mathbf{u}} = \frac{Var_{\mathbf{X_u}}(\mathbb{E}_{\mathbf{X_{-u}}}[\,\tilde{Z}_n(\mathbf{X}) \mid \mathbf{X_u}])}{Var_{\mathbf{X}}(\tilde{Z}_n(\mathbf{X}))}\,. \tag{12}$$

As Z_n is a random process, the resulting indices are also random. These indices are estimated via Monte-Carlo samples from two designs of experiments, using a pick-freeze procedure. A design is a point set $\mathbf{P} = \{\mathbf{x}_i\}_{i=1}^s$ in which each point is obtained by sampling s times each input variable $X_j \in \mathcal{P}$, $j = 1, \dots, p$. Each row of the design is a point $\mathbf{x}_i$ in $\mathcal{P}$, the j-th column of the design refers to a sample of X_j and for $\mathbf{u} \subseteq \{1, \dots, p\}$, $\mathbf{x}_{i,\mathbf{u}} = \{x_{i,j}\}$, $j \in \mathbf{u}$. Given two points $\mathbf{x}$ and $\mathbf{x}'$, the hybrid point $(\mathbf{x_u} : \mathbf{x}'_{-\mathbf{u}}) \in \mathcal{P}$ is defined as x_j if $j \in \mathbf{u}$ and x'_j if $j \notin \mathbf{u}$. Consider $\mathbf{P} = \{\mathbf{x}_i\}_{i=1}^s$ and $\mathbf{P'} = \{\mathbf{x'}_i\}_{i=1}^s$ two designs sampled from the distribution of the input random vector $\mathbf{X}$. One way to estimate the quantity in Eq. (12)

has been proposed by Homma and Saltelli [40] and further studied in Janon *et al.*, see Lemma 1 in [41]. They propose the following estimator:

$$\widehat{S}_{\mathbf{u}} = \frac{\frac{1}{s}\sum_{i=1}^{s} \tilde{Z}_n(\mathbf{x}_i)\ \tilde{Z}_n(\mathbf{x}_{i,\mathbf{u}} : \mathbf{x}'_{i,-\mathbf{u}}) - \frac{1}{s}\sum_{i=1}^{s} \tilde{Z}_n(\mathbf{x}_i)\ \frac{1}{s}\sum_{i=1}^{s} \tilde{Z}_n(\mathbf{x}_{i,\mathbf{u}} : \mathbf{x}'_{i,-\mathbf{u}})}{\frac{1}{s}\sum_{i=1}^{s} \tilde{Z}_n(\mathbf{x}_i)^2 - (\frac{1}{s}\sum_{i=1}^{s} \tilde{Z}_n(\mathbf{x}_{i,\mathbf{u}} : \mathbf{x}'_{i,-\mathbf{u}}))^2} \,. \tag{13}$$

Confidence intervals can be obtained via a bootstrap method, as described in Algorithm 1 of [32]. These intervals integrate two sources of uncertainty, the first one is related to the metamodel approximation, and the second one is related to the Monte-Carlo integration.

Each step of the procedure was implemented in **R** using the *km* and *sobolGP* functions from respectively the *DiceKriging* and *Sensitivity* packages, see [42, 43].

2 | BAYESIAN INFERENCE FOR ONLINE PARAMETER IDENTIFICATION

In our context, we suppose that data are collected sequentially and we seek to refine our choice of parameters in the numerical model at each iteration. This problem can be seen as a supervised learning problem that we aim to solve sequentially as each pair of data points $\{v_{(k)}, \mathbf{y}_{(k)}\}$ is obtained at the iteration k [44].

In the recursive Bayesian parameter estimation framework, developed in this paper, the unknown time-invariant input parameter vector $\mathbf{x}$ is modeled as a discrete Markov chain, the evolution of which is governed by a random walk process. In our context, the dynamic evolution of the input parameter and the measurement modelisation with the simulator responses can be formulated at iteration $\{k\}_{k=1\dots T}$ as:

$$\left\{ \begin{array}{rcl} \mathbf{x}_{(k)} & = & \mathbf{x}_{(k-1)} + \delta_{(k)} \\ \mathbf{y}_{(k)} & = & \mathcal{M}(\mathbf{x}_{(k)}, V = v_{(k)}) + \boldsymbol{\epsilon}_{(k)} \end{array} \right. , \tag{14}$$

where $\mathbf{x}_{(k)}$ is the input parameter vector, $v_{(k)}$ is a known realization of the stochastic external excitation at k, the Gaussian noises $\delta_{(k)} \sim \mathcal{N}(0, \mathbf{Q}_{(k)})$ and $\boldsymbol{\epsilon}_{(k)} \sim \mathcal{N}(0, \mathbf{R}_{(k)})$ are respectively an artificial dynamic noise and a combination of the model and observation errors. For the sake of readability, $\mathbf{y} \in \mathbb{R}^m$ will represent the vector gathering the measured responses obtained on the structure of interest.

Filtering techniques, a type of data assimilation, can be used to sequentially estimate the parameter vector in Eq. (14) using the known input solicitation and the available measurements. Among all available filtering methods, the Kalman Filter (KF) [45] has been widely applied when dealing with a linear system with Gaussian error sources. In this paper due to the non-linearity in our numerical model, the Ensemble Kalman filter (EnKF) [22] is used to perform parameters estimation. The EnKF is a sequential Monte-Carlo method that provides an alternative to the traditional KF. The method works on an ensemble of parameter estimates transforming them from the prior distribution into the posterior one. We propose to use a Latin Hypercube sampling technique coupled with a geometrical criteria maximizing the minimum distance between the design points instead of a conventional Monte-Carlo method to generate the initial ensemble of parameter estimates, see [46].

In the field of inverse problems, this inference method is referred to as Ensemble Kalman Inversion (EKI). The EnKF formulation used in [47] is adopted in this paper. At iteration k, we have an ensemble of size N_{ens}, of forecast parameter estimates $\mathbf{x}^f_{(k)} = [\mathbf{x}^{f(1)}_{(k)}, \dots, \mathbf{x}^{f(N_{ens})}_{(k)}] \in \mathbb{R}^{p \times N_{ens}}$ where the superscript $\cdot^{f(i)}$ denotes the i-th forecast member of

the ensemble. The mean of the forecast members of the ensemble is given by:

$$\bar{\mathbf{x}}^f_{(k)} = \frac{1}{N_{ens}} \sum_{i=1}^{N_{ens}} \mathbf{x}^{f(i)}_{(k)} \cdot \tag{15}$$

The error covariance matrix for the forecast estimate in the KF can be empirically estimated as:

$$\mathbf{P}^f_{(k)} = \frac{1}{N_{ens}-1} \sum_{i=1}^{N_{ens}} (\mathbf{x}^{f(i)}_{(k)} - \bar{\mathbf{x}}^f_{(k)})(\mathbf{x}^{f(i)}_{(k)} - \bar{\mathbf{x}}^f_{(k)})^T \cdot \tag{16}$$

As said previously the structure of the EnKF is the same as the one of the Kalman filter [48]. Thus, we need to compute the Kalman Gain, referred as $\mathbf{K}_{(k)}$ and defined by:

$$\mathbf{K}_{(k)} = \mathbf{P}^f_{(k)}\mathbf{M}^T \left(\mathbf{M}\mathbf{P}^f_{(k)}\mathbf{M}^T + \mathbf{R}_{(k)}\right)^{-1}, \tag{17}$$

where the observation operator, denoted by $\mathbf{M} \in \mathbb{R}^{m \times N_{ens}}$, is linear or has been linearized from the function $\mathcal{M}$, see [49].

Nevertheless, for most applications this condition of linear (or linearized) observation operator cannot be applied. In that context, as proposed in [50], we can replace the terms $\mathbf{P}^f_{(k)}\mathbf{M}^T$ and $\mathbf{M}\mathbf{P}^f_{(k)}\mathbf{M}^T$ of the Kalman Gain equation by the following ones:

$$\frac{1}{N_{ens}-1} \sum_{i=1}^{N_{ens}} \left(\mathbf{x}^{f(i)}_{(k)} - \bar{\mathbf{x}}^f_{(k)}\right) \left(\mathcal{M}(\mathbf{x}^{f(i)}_{(k)}, V = v_{(k)}) - \mathcal{M}(\bar{\mathbf{x}}^f_{(k)}, V = v_{(k)})\right)^T \tag{18}$$

and,

$$\frac{1}{N_{ens}-1} \sum_{i=1}^{N_{ens}} \left(\mathcal{M}(\mathbf{x}^{f(i)}_{(k)}, V = v_{(k)}) - \mathcal{M}(\bar{\mathbf{x}}^f_{(k)}, V = v_{(k)})\right) \left(\mathcal{M}(\mathbf{x}^{f(i)}_{(k)}, V = v_{(k)}) - \mathcal{M}(\bar{\mathbf{x}}^f_{(k)}, V = v_{(k)})\right)^T \cdot \tag{19}$$

It has been argued, in [51], that Eq. (18) and Eq. (19) are good approximations if the following hypothesis are verified:

$$\mathcal{M}(\bar{\mathbf{x}}^f_{(k)}, V = v_{(k)}) = \overline{\mathcal{M}(\mathbf{x}^f_{(k)}, V = v_{(k)})} = \frac{1}{N_{ens}} \sum_{i=1}^{N_{ens}} \mathcal{M}(\mathbf{x}^{f(i)}_{(k)}, V = v_{(k)}),$$
$$\mathbf{x}^{f(i)}_{(k)} - \bar{\mathbf{x}}^f_{(k)} = \xi_i \text{ and } \|\xi_i\| \text{ is small for } i = 1 \dots N_{ens} \cdot$$

This version of EnKF treats the observations as random variables [22]. Indeed, an ensemble of observations of the same size N_{ens} is generated by adding noise terms to the observation set $\mathbf{y}_{(k)}$ such that:

$$\mathbf{y}^{(i)}_{(k)} = \mathbf{y}_{(k)} + \mathbf{e}^{o(i)}_{(k)}, \text{ with } \mathbf{e}^{o(i)}_{(k)} \sim \mathcal{N}(0, \mathbf{R}_{(k)}), i = 1 \dots N_{ens} \cdot \tag{20}$$

The noise terms are sampled from the distribution of the error covariance matrix $\mathbf{R}_{(k)}$. The stochastic EnKF has been showed to have the advantage to "re-Gaussianize" the ensemble distribution thanks to the observation perturbations [52]. Maintaining Gaussianity has a positive impact on analysis quality of the ensemble filter by maintaining the correct forecast error covariance. Most of the time, the measurement observational error covariance matrix is diagonal

according to the assumption of independent observations. Using the presented approximation, the computation of the Kalman gain $\mathbf{K}_{(k)}$ can be done. We can independently update the ensemble members using:

$$\mathbf{x}^{a(i)}_{(k)} = \mathbf{x}^{f(i)}_{(k)} + \mathbf{K}_{(k)} \left(\mathbf{y}^{(i)}_{(k)} - \mathcal{M}(\mathbf{x}^{f(i)}_{(k)}, V = v_{(k)}) \right) \cdot \tag{21}$$

where the superscript $\cdot^{a(i)}$ denotes the i-th updated member of the ensemble. The last step of the EnKF method is the forecast step of the ensemble parameters at $k+1$ and involves an ensemble of N_{ens} updated parameters for iteration k, such as:

$$\mathbf{x}^{f}_{(k+1)} = \mathbf{x}^{a}_{(k)} + \delta_{(k)}\,, \text{ with } \delta_{(k)} \sim \mathcal{N}(0, \mathbf{Q}_{(k)}) \cdot \tag{22}$$

The presented method is fully described in Algo. 1.

Algorithm 1: Ensemble Kalman Filter for parameter inference, a.k.a. Ensemble Kalman Inversion.

Data:

number of members in the ensemble N_{ens};

prior guess of the parameter vector $\mathbf{x}_b$ and prior parameter covariance matrix $\mathbf{P}_b$;

some measurements $\{\mathbf{y}_{(k)}\}_{k=1,\dots,T}$ and known realization of the external solicitation $\{v_{(k)}\}_{k=1,\dots,T}$;

error covariance matrix $\{\mathbf{R}_{(k)}\}_{k=1,\dots,T}$ and artificial error covariance matrix $\{\mathbf{Q}_{(k)}\}_{k=0,\dots,T}$.

Initialisation step:

for $i = 1$ *to* N_{ens} **do**

$$\mathbf{x}^{a(i)}_{(0)} = \mathbf{x}_b + \epsilon^b \text{ with, } \epsilon^b \sim \mathcal{N}(0, \mathbf{P}_b)$$

for $k = 1$ *to* T **do**

Forecast step:

$$\mathbf{x}^{f}_{(k)} = \mathbf{x}^{a}_{(k-1)} + \delta_{(k-1)}\,, \text{ with } \delta_{(k-1)} \sim \mathcal{N}(0, \mathbf{Q}_{(k-1)})$$

$$\bar{\mathbf{x}}^{f}_{(k)} = \frac{1}{N_{ens}} \sum_{i=1}^{N_{ens}} \mathbf{x}^{f(i)}_{(k)} \text{ and } \mathbf{P}^{f}_{(k)} = \frac{1}{N_{ens}-1} \sum_{i=1}^{N_{ens}} (\mathbf{x}^{f(i)}_{(k)} - \bar{\mathbf{x}}^{f}_{(k)})(\mathbf{x}^{f(i)}_{(k)} - \bar{\mathbf{x}}^{f}_{(k)})^T$$

Update step:

$$\mathbf{P}^{f}_{(k)}\mathbf{M}^T = \frac{1}{N_{ens}-1} \sum_{i=1}^{N_{ens}} \left(\mathbf{x}^{f(i)}_{(k)} - \bar{\mathbf{x}}^{f}_{(k)} \right) \left(\mathcal{M}(\mathbf{x}^{f(i)}_{(k)}, V = v_{(k)}) - \mathcal{M}(\bar{\mathbf{x}}^{f}_{(k)}, V = v_{(k)}) \right)^T$$

$$\mathbf{M}\mathbf{P}^{f}_{(k)}\mathbf{M}^T = \frac{1}{N_{ens}-1} \sum_{i=1}^{N_{ens}} \left(\mathcal{M}(\mathbf{x}^{f(i)}_{(k)}, V = v_{(k)}) - \mathcal{M}(\bar{\mathbf{x}}^{f}_{(k)}, V = v_{(k)}) \right) \left(\mathcal{M}(\mathbf{x}^{f(i)}_{(k)}, V = v_{(k)}) - \mathcal{M}(\bar{\mathbf{x}}^{f}_{(k)}, V = v_{(k)}) \right)^T$$

$$\mathbf{K}_{(k)} = \mathbf{P}^{f}_{(k)}\mathbf{M}^T (\mathbf{M}\mathbf{P}^{f}_{(k)}\mathbf{M}^T + \mathbf{R}_{(k)})^{-1}$$

for $i = 1$ *to* N_{ens} **do**

$$\mathbf{y}^{(i)}_{(k)} = \mathbf{y}_{(k)} + \mathbf{e}^{o(i)}_{(k)} \text{ with } \mathbf{e}^{o(i)}_{(k)} \sim \mathcal{N}(0, \mathbf{R}_{(k)})$$

$$\mathbf{x}^{a(i)}_{(k)} = \mathbf{x}^{f(i)}_{(k)} + \mathbf{K}_{(k)} \left(\mathbf{y}^{(i)}_{(k)} - \mathcal{M}(\mathbf{x}^{f(i)}_{(k)}, V = v_{(k)}) \right)$$

3 | DESCRIPTION OF THE WIND TURBINE NUMERICAL MODEL

Dynamic analysis of wind turbines involves strong interactions between the turbines' aerodynamics, the control system and the structural mechanics. The main solicitations are the environmental conditions and the rotating machinery during operating term. In order to model and simulate the nonlinear response of wind turbine structures under such solicitations, various servo-aero-elastic software have been developed, such as OpenFAST [53], Bladed [54], HAWCK2 [55] or Deeplines Wind™[56].

In our study, a simulator of a Senvion MM82 wind turbine has been developed using Deeplines Wind™software from technical specifications. This software is a fully coupled simulation tool taking into account the aerodynamics of the aero-generator, the elasticity of the structural wind turbine components (mast, blades and drive-train systems) and the control system [57]. The software architecture developed by IFPEN[2] and Principia[3] is fully modular with different dynamic libraries (DLL) called by the solver. The integration in time is performed with an implicit Newmark integration scheme. The developed simulator includes a nonlinear beam finite element formulation to model the structural components. The aerodynamic loads acting on the turbine rotor are dynamically computed by employing the blade element momentum (BEM) theory for Horizontal Axis Wind Turbine (HAWT). A Deeplines Wind™software validation, based on code comparisons [5], has shown accurate results in various conditions.

Wind turbine simulation consists of two stages: first the generation of the input turbulent wind field and then the fully coupled servo-aero-elastic simulation. The generation of the input stochastic process is done by using a simulator called Turbsim [4]. This simulator has some deterministic inputs such as the turbulence intensity, the mean wind speed, the mean flow angles, the spectrum and the spatial correlation model. In our model, we have used an IEC[4] Kaimal turbulence spectrum with an exponential spatial coherence. Nevertheless, these deterministic values cannot uniquely determine a stochastic wind field and a pseudo-random number generator have to be used in order to create random phases for the velocity time. Then structural responses are time computed with a multi-physics numerical code such as Deeplines Wind™, following the formalism introduced in Eq. (1). A wind turbine structure can encounter a variety of operating conditions. Each of the operating conditions, modelled by the stochastic process V, is parameterized by measurable deterministic quantities mentioned in Fig. 1. In this paper we will perform the study at an underrated average wind speed of 8 m/s corresponding to the most common operating regime of the considered turbine. All computed responses are based on 10-min *effective* simulations of the MM82 Senvion wind turbine. By *effective*, we mean that the transient start-up behavior is removed from the analysis. The transient start-up behavior can be decomposed in a ramp time wind and an oversight periods. The oversight period has been set according to autocorrelation studies of the outputs. This period permits to remove the effect of the ramp time period on the numerical model responses. A numerical simulation lasts 15 minutes on an Intel Xeon Scalable Gold 6140 processor operating at 2.3 GHz.

From the structural time responses computed by the Deeplines Wind™model, we obtain some QoIs describing the fatigue behavior of the wind turbine. They are obtained by post-treating the resulting time series of internal loads at different locations of the analysed design, see Fig. 2 and Tab. 1. In our study, we denote by g the function mapping the functional loads of the structure to the damage QoIs, see Eq. (2). Each fatigue QoI has been estimated by using the damage-equivalent load (DEL). The DEL is computed for a set of parameter values and different realizations of the stochastic process V. It is defined as the regular load amplitude that would create in N_{ref} cycles the same fatigue

[2] see `https://www.ifpenergiesnouvelles.fr/`

[3] see `http://www.principia-group.com/`

[4] International Electrotechnical Commission

as the considered irregular load history. The DEL is computed based on the Palgrem Miner's rule [58]:

$$DEL = \left(\frac{\sum_{i=1}^{N_c} S_i \cdot N_i^m}{N_{ref}} \right)^{\frac{1}{m}} . \tag{23}$$

where $i = 1, \ldots, N_c$ corresponds to each range bin, S_i is the cycle range value and N_i is the number of cycles for the i-th bin. The exponent m is the negative inverse slope of the cyclic stress against the cycles to failure curve (S-N curve) and N_{ref} is the reference number of cycles usually set to an arbitrary value. The cycles in an irregular load history are computed using the Rainflow counting method [59].

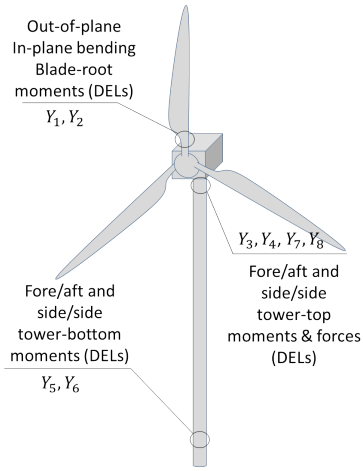


FIGURE 2 Recorded time series of loads at different locations of the wind turbine

Load's position	m	Type of load	
Blade root moments	10	Out-of-plane bending	In-plane bending
Tower top moment	3	Fore-aft bending	Side-to-side bending
Tower bottom moment	3	Fore-aft bending	Side-to-side bending
Tower top force	3	Fore-aft force	Side-to-side force

TABLE 1 Fatigue damage equivalent loads used for the global sensitivity analyses with their corresponding Wöhler's exponent, i.e., the negative inverse slope of the S-N curve.

In order to ensure that the variation induced by the input parameters is distinguishable from the one induced by the realization of the stochastic process V, multiple wind realizations have to be generated. A convergence study has been led in order to determine the number of realizations needed to encompass the variation of the selected realization of the stochastic process. The QoIs analyzed in this study are the DEL at different locations of the wind turbine, see Eq. (23). The number of stochastic process realizations used during the convergence study varies from 1 to 30.

Fig. 3 shows the convergence of the tower bottom fore-aft bending moment DEL on the number of realizations used for their averaging estimation. We assume a 10-minutes mean hub-height wind speed of 8 m/s and a wind fluctuation following the Kaimal IEC. The IEC 61400-1 standard has three turbulence categories: A, B and C, see [60]. We have decided to use the A class corresponding to the highest turbulence intensity, i.e. the ratio of standard deviation of fluctuating wind velocity to the mean wind speed is around 24 %. A time series respecting the 10-minutes statistics, which will be different at each generation, is created thanks to the simulator [4]. According to certification guidelines, see design load case 1.2 in [61], the fatigue analysis has to be led with 6 wind realizations of 10 min time period. Nevertheless, as we can underline with the last mentioned figure, the empirical mean computed from these 6 realizations does not seem to be a reliable estimator. In other words, 6 simulations are not sufficient for QoIs' statistics to converge. With our industrial numerical model, a compromise has been made to balance the quality of the empirical estimator and the computing time goal by fixing the number of realizations to 10 for the GSA of the

damage equivalent loads.

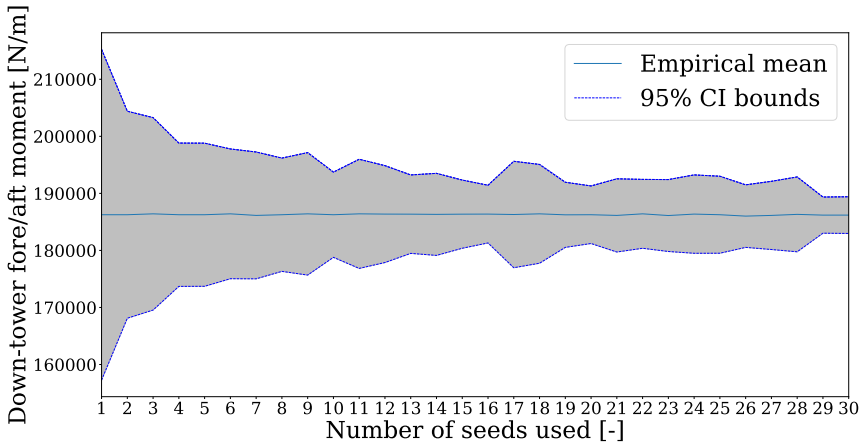


FIGURE 3 Convergence of the tower bottom fore-aft bending moment DEL as a function of the number of turbulent seeds used for its evaluation. 95% confidence interval around the estimated empirical mean is also represented (grey area). The wind turbulence intensity is set around 24 %, see [60].

The aim of this paper is to identify and reduce the structural sources of uncertainty on the fatigue QoIs. We will focus our study on some wind turbine properties represented by 13 parameters gathered in the vector $\mathbf{x}$. A literature review has been done to specify the uncertainty in the parameter values. Based on expert knowledge, all these parameters were considered independent of one another with Gaussian distributions due to their physical properties.

Here follows a description of the 13 considered parameters, see Tab. 2. For the support structural properties, six parameters have been considered, including nacelle mass, nacelle center of mass, tower Rayleigh damping, inertial nacelle and drive-train torsion stiffness. The tower thickness has been changed by uniformly scaling the distributed tower thickness. The boundary values have been set by changing the first fore-aft tower frequency mode by ±10% of its reference value.

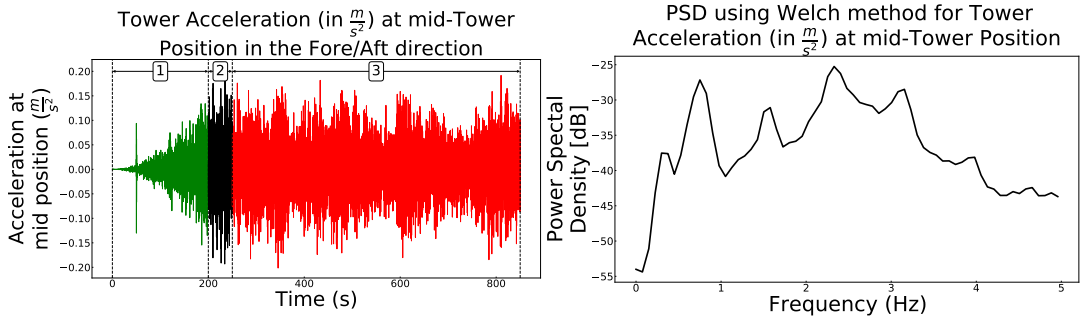
The uncertainties in blade structural properties have been represented using five parameters. The blade structural responses have led to the definition of the uncertainty range. Indeed, the frequency of the edge-wise (ES) and flap-wise (FS) mode were changed about 10% each from their nominal value. These modifications of the mode were done by uniformly scaling the associated stiffness and the distributed blade mass of all blades. Blade imbalance effects have been also included by applying a different change value to each blade. One blade is modified to be a value that is higher than the nominal value, and another one modified to a lower value. The third blade remains unchanged at the nominal value.

For the aerodynamic properties, we have considered the wind turbine yaw misalignment by changing the yaw angle of the turbine. For the individual blade pitch error, a constant offset angle is applied to two of the blades, respectively above and below nominal value.

TABLE 2 Structural properties - uncertainties affecting the input parameters of the wind turbine model.

Input parameter	μ	σ	REF
Nacelle Mass - N_{mass} [kg]	6.90e+04	2.30e+03	[62]
Nacelle center of mass - N_{CMx} [m]	1.00	3.33e-02	[63]
Tower thickness - e [%]	0	7	IFPEN ±10% 1 FA
Tower Rayleigh Damping - β_{TR} [-]	3.10e-02	9.93e-03	[64]
Inertial Nacelle - I_{zz} $[kg \cdot m^2]$	7.00e+05	2.33e+04	IFPEN ±10% μ
Drive-train Torsional stiffness - K_D $[\frac{N \cdot m^2}{rad}]$	4.45e+09	1.48e+08	[65]
Blade Flap wise Stiffness - α_{BF} $[N \cdot m^2]$	1.00	3.33e-02	IFPEN ~ ±10% 1 FS
Blade Edge wise Stiffness - α_{BE} $[N \cdot m^2]$	1.00	3.33e-02	IFPEN ~ ±10% 1 ES
Blade mass coefficient - α_{mass} [%]	1.00	1.67e-02	[62]
Blade Rayleigh Damping - β_{BR} [-]	5.39e-03	1.45e-03	[63]
Blade mass imbalance - η_B [%]	2.50	8.33e-01	[63]
Yaw misalignment - ω [°]	0	6.67	[66]
Individual pitch error - Ω [°]	0.10	3.33e-02	[67]

After an appropriate sensitivity analysis leading to the identification of the less influential input parameters on the fatigue QoIs, we can fix their value to a nominal one without affecting the fatigue behavior of the structure. Then, the uncertainties of the other parameters has to be reduced by employing an appropriate inference method based on in-situ measurements. In this context, let us consider a wind turbine instrumented with accelerometers. We assume that bi-axial measuring devices are located at mid and top tower height position. From these sensors, we can record four functional acceleration time series. Then, the power spectral density (PSD) of each measured acceleration time series is computed using Welch's method [68], see Fig. 4.



(a) (b)

FIGURE 4 On the left side (a): simulated acceleration in $\frac{m}{s^2}$ of the wind turbine tower in the fore-aft direction obtained at the accelerometer device located at mid-tower decomposed in a ramp time wind 1, an oversight period 2 and a dynamical period of interest 3. On the right side (b): estimated PSD of the period of interest using Welch's method [68].

4 | ILLUSTRATION OF THE PROPOSED FRAMEWORK ON THE WIND TURBINE NUMERICAL MODEL

GSA of the fatigue QoIs - The Deeplines Wind™numerical model presented in Section 3 is used with 13 uncertain input parameters, each one having its associated probability distribution, see Tab.2. The total Sobol' indices associated to these parameters for each DEL have been estimated using the heteroscedastic noisy GP model-based Sobol' index procedure as described in Section 1. A Latin Hypercube Sampling (LHS) of size 500 has been used to emulate the numerical model. Then, an augmented LHS of size 150 has been generated to determine the accuracy of the surrogate models. To apply this approach 6,500 forward wind turbine numerical simulations were submitted on the 206 TFlops supercomputer of IFPEN.

The function *sobolGP* performs a kriging-based GSA by taking into account the complete conditional predictive distribution of the surrogate model. The function estimates total Sobol' indices thanks to the Jansen estimators, see [69]. Jansen Sobol' index estimators are accurate for large and small total indices. Moreover, by taking into account the complete conditional predictive distribution of the trained surrogate model in the estimation procedure of the total Sobol' indices, we can evaluate the uncertainty due to Monte-Carlo estimation, but also due to metamodeling [32].

The results for the total Sobol' indices with their corresponding 95% confidence intervals are summarized in Fig. 5. A threshold of 2.5e-2 was chosen to display a separation between input parameters with high and low total Sobol' indices. Fig. 6 represents the different sources of uncertainty for the estimation of total Sobol' indices, obtained thanks to the function *sobolGP* for the DEL of the out-of-plane bending blade-root moment.

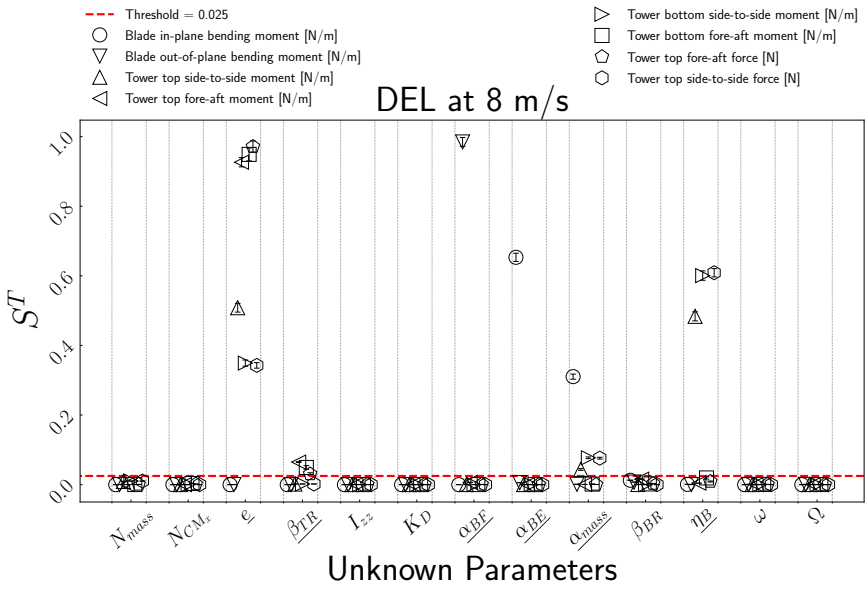


FIGURE 5 Estimation of total Sobol' indices (y-axis) with their 95% confidence interval corresponding to each of the 13 parameters (x-axis) for the different quantities of interest in fatigue. The dashed line corresponds to a threshold arbitrarily chosen to 2.5e-2. Confidence intervals (CI) are obtained by taking into account the uncertainties due to both the metamodel and the Monte-Carlo estimation. The number of samples for the conditional Gaussian process, in order to quantify the uncertainty of the Kriging approximation, was set to 100. The uncertainty due to Monte-Carlo integration was computed with a bootstrap procedure with a sample size of 100.

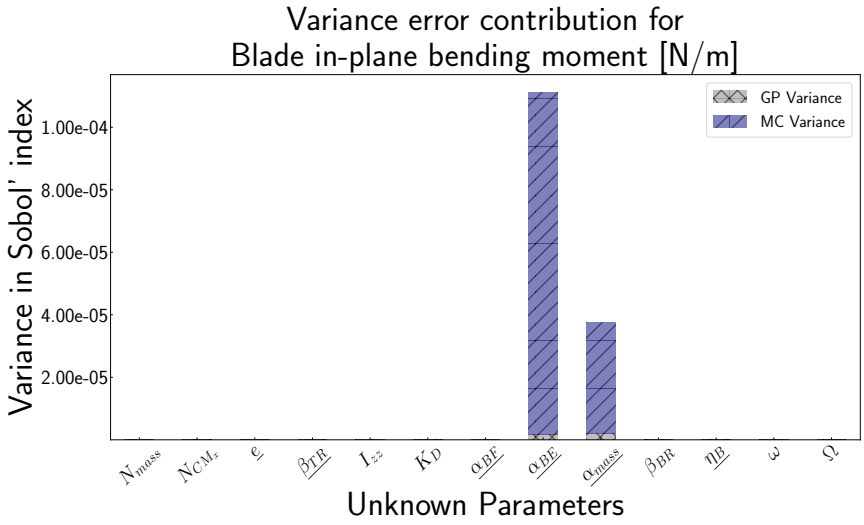


FIGURE 6 Splitting of the variance of total Sobol' index estimators (y-axis) corresponding to each of the 13 input parameters (x-axis)for the DEL of the out-of-plane bending blade-root moment. The number of samples for the conditional Gaussian process, in order to quantify the uncertainty of the Gaussian process approximation, was set to 100. The uncertainty due to Monte-Carlo integration was computed with a bootstrap procedure of 100.

By considering all the total Sobol' indices obtained for the different fatigue QoIs presented in Fig. 5, we can notice that only 6 parameters have indices values greater than the threshold. Consequently, we can fix other parameters to any specific value in the range of variability without affecting the fatigue QoIs. During the recursive Bayesian estimation, these non-influential parameters will be arbitrarily set at their mean value described in Tab. 2. By employing this method, we will reduce the parameter dimension in the inference procedure without affecting the fatigue QoIs which are crucial for assessing the lifespan of wind turbines.

Identifiability study - It is possible that the considered experimental measurement settings do not offer enough information for the identification of some input parameters. In this context, another interesting property of the GSA underlined in Proposition 4.2 in [24] is that nullity of the total sensitivity index for a specific input parameter implies its non-identifiability from the output under consideration. Thus, a GSA led on the measured outputs will determine which parameters cannot be inferred, although it is not a sufficient condition for identifiability.

In our industrial application study, the measured outputs, obtained from the accelerometers, are expressed as discretized time series. We are interested in their response in the frequency-domain by using the power spectral density (PSD). Discretized PSD series involve a substantial dimensionality and a high degree of redundancy. To bypass this issue, we have firstly focused our study on an orthogonal decomposition, of the different discretized PSDs, in order to reduce their dimensionality. This orthogonal decomposition will be performed by a data-driven method called Principal Component Analysis (PCA) [70]. PCA allows the functional output expansion in a new reduced space spanned by the most significant directions in term of variance of the discretized functional output.

In our study, a method based on PCA and GSA with GP model is used to compute an aggregated Sobol' index for each input parameter of the model. As described in [25] , the proposed index synthesizes the influence of the parameter on the whole time series output.

In our study to ensure that the variation of the input parameters is distinguishable from the realization of the stochas-

tic process V, 10 wind realizations have been used in this GSA. A new LHS of size 300 with a geometrical criteria maximizing the minimum distance between the design points has been used to emulate the numerical model. In Tab. 3, we summarize the total aggregated Sobol' indices obtained with the GP model built on the trained set from the lastly mentioned LHS. In this analysis, parameters with total Sobol' index values under a threshold set at 1e-02 will be considered as non-identifiable from the measured output. We can conclude that none of the significant input parameters can be considered a-priori as non-identifiable.

TABLE 3 Total Sobol' indices for each output used during the recursive inference procedure described in details in Section 2. Estimated total Sobol' indices higher than the arbitrary threshold are underlined.

Measured outputs	e [%]	β_{TR} [−]	α_{BF} [%]	α_{BE} [%]	α_{mass} [%]	η_B [%]
Tower middle fore-aft acceleration's PSD	2.44e-01	7.64e-01	6.07e-03	1.37e-04	3.59e-04	4.46e-03
Tower middle side-to-side acceleration's PSD	3.84e-01	3.95e-01	1.38e-01	6.73e-04	8.60e-02	7.17e-03
Tower top fore-aft acceleration's PSD	1.21e-01	6.70e-01	2.09e-03	5.91e-02	6.76e-02	1.05e-01
Tower top side to side acceleration's PSD	6.56e-02	6.24e-01	1.36e-03	9.67e-02	1.39e-01	9.52e-02

Recursive Bayesian inference study - The 6 input parameters having an influential effect on the fatigue behavior of the structure and potentially identifiable are considered during the inference procedure. These unknown input parameters define the model parameter vector to be estimated, i.e., $\mathbf{x} = [e, \beta_{TR}, \alpha_{BF}, \alpha_{BE}, \alpha_{mass}, \eta_B]^T$.

In this section we focus on pseudo-experimental numerical tests in order to validate the inference procedure. These tests consist in running direct numerical analyses considering known values of input parameters, and then adding a Gaussian noise of known variance to the observed outputs. The dynamic response of the wind turbine structure is simulated using the mean values of the unknown model parameters described in Tab. 2. The noisy pseudo-experimental outputs used to recursively update the wind turbine model are the PSD of the acceleration time series obtained for side to side and fore-aft at the two different tower positions.

Algo. 1 is used to recursively estimate the expected values of the unknown input parameters at each iteration step. The output measurement noise covariance matrix $\boldsymbol{R}_k$ is assumed to be diagonal, i.e., cross-correlations between the noise components are disregarded. Usually, the amplitudes of the measurement noise can be estimated based on the characteristics of the used sensors. Nevertheless, in our pseudo-experimental numerical case, these amplitudes have been arbitrarily chosen. Indeed, to mimic real-life applications, noise is incorporated in the simulated data by considering a noise covariance matrix such as the obtained standard deviation is equivalent to a 1% signal-to-noise ratio.

For the initialization of the Bayesian estimation procedure, $\mathbf{x}_b = [-0.10, 4.00e-02, 0.98, 1.05, 9.85e-01, 0.04]^T$ is assumed to be the initial prior of the value of the input parameters. The initial error covariance matrix of the input parameters, denoted by $\boldsymbol{P}_b$, is also assumed to be diagonal. In other words, the initial prior of the input parameters are assumed to be statistically independent. Diagonal elements of $\boldsymbol{P}_b$ represent the practitioner's belief on the input parameters uncertainties, such as $\boldsymbol{P}_b = \mathrm{diag}(7.00e-02, 9.93e-03, 3.33e-02, 3.33e-02, 1.67e-02, 8.33e-03)$.

The number of inference iterations T and of the number of members N were respectively fixed at $T = 20$ and $N = 100$. This choice was mainly guide by the maximal simulation budget allocated to our study and by the fact that modest

ensemble size is a reasonable practice as observed in industrial setups [22, 71, 72]. Fig. 7 shows the results of the identification. It can be observed that the considered input parameters are well recovered. Tab. 4 reports the final a posteriori estimate of the input parameters.

TABLE 4 Target, initial prior and final a posteriori estimates of the input parameters of the wind turbine numerical model.

	e [%]	β_{TR} [−]	α_{BF} [%]	α_{BE} [%]	α_{mass} [%]	η_B [%]
Target	0	3.10e-02	1	1	1	2.50e-02
Prior estimates	-1.00e-01	4.00e-02	9.80e-01	1.05	9.85e-01	4.00e-02
A posteriori estimates (T=20)	3.26e-03	3.08e-02	9.98e-01	1.00e+00	1.00e+00	2.52e-02

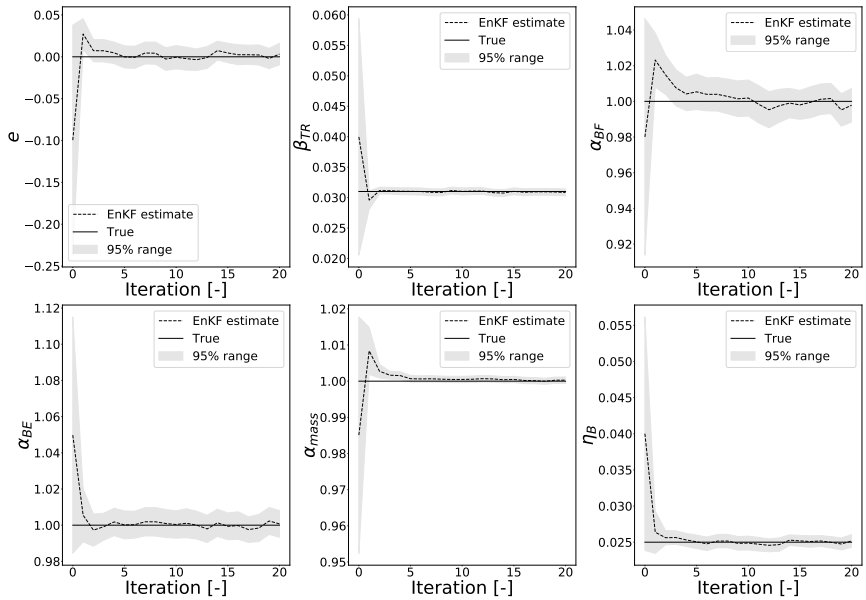


FIGURE 7 Iteration evolution of the a posteriori estimates of the input parameters. Results obtained by running EnKF presented in Section 2 with $N = 100$ members of the ensemble used for the estimation and considering pseudo-experimental numerical observations.

Robustness analysis - To test the effectiveness of the proposed EnKF procedure, different noise levels affecting the synthetic data have been considered. We have chosen different structures of noise covariance matrices such as the obtained standard deviations affecting the measurements are respectively equivalent to 3% and 5% signal-to-noise ratios. The performed analysis have highlighted that the incorporation of higher noise leads to a harder identification of input parameters. The estimation of these parameters is less reliable because their identifiability property becomes weaker. The issue of identifiability is a crucial aspect due to the presence of noise in real-life applications. However, the proposed recursive Bayesian inference method have the ability to give confidence intervals on the inferred parameters due to its probabilistic framework.

Conclusion

This paper presents a framework to quantify and reduce the uncertainties from the input parameters of a wind turbine numerical model.

The contributions of this paper are twofold. First, we have proposed a global sensitivity analysis based on Sobol' indices using a Gaussian process model with heteroscedastic nugget effect to quantify uncertainties of a stochastic and time-consuming wind turbine numerical model. The procedure we present is efficient to balance the inherent uncertainty of the stochastic numerical model and the one from the input parameters. More precisely the GSA has been performed on the fatigue quantities of interest and showed that only a restricted number of unknown parameters happens to influence these responses. Since fatigue quantities of interest are a crucial wind turbine design and life criteria, these influential input parameters have to be properly estimated in order to give confidence in fatigue analysis results.

Consequently, the second objective of this paper was to develop a recursive inference procedure to properly determine these parameters based on available measurements. But first was addressed the question of parameter non-identifiability by employing a global sensibility study on the measured outputs. As previously stated, performing such sensitivity analysis is not a sufficient condition for identifiability. Finally for the inference task, this paper demonstrates the applicability and computational efficiency of the ensemble Kalman filter (EnKF) for this type of challenging problem. The EnKF-based approach has been integrated into the commercial software Deeplines Wind™. The proposed methodology was verified using numerically simulated response data. For future work, the recursive Bayesian estimation procedure will be further tested by incorporating other measured output data.

Acknowledgements

The author would like to thank the framework agreement between IFP Energies Nouvelles (IFPEN) and the French Institute for Research in Computer Science and Automation (INRIA) which has funded the present work. Moreover, the author gratefully thanks all researchers who directly or indirectly contributed to the achievement of the work mentioned in this paper.

Conflict of interest

The authors declare that they have no conflict of interest.

Data availability statement

Data available on request from the authors.

References

[1] De Rocquigny E, Devictor N, Tarantola S. Uncertainty in industrial practice: a guide to quantitative uncertainty management. John Wiley & Sons; 2008.

[2] Smith RC. Uncertainty quantification: theory, implementation, and applications, vol. 12. Siam; 2013.

[3] Hart JL, Alexanderian A, Gremaud PA. Efficient computation of Sobol'indices for stochastic models. SIAM Journal on Scientific Computing 2017;39(4):A1514–A1530.

[4] Jonkman BJ. TurbSim user's guide: Version 1.50. Golden, CO, USA: National Renewable Energy Lab (NREL); 2009.

[5] Perdrizet T, Gilloteaux JC, Teixeira D, Ferrer G, Piriou L, Cadiou D, et al. Fully Coupled Floating Wind Turbine Simulator Based on Nonlinear Finite Element Method: Part II Validation Results. In: ASME 2013 32nd International Conference on Ocean, Offshore and Arctic Engineering Citeseer; 2013. p. V008T09A052–V008T09A052.

[6] Kwon SD. Uncertainty analysis of wind energy potential assessment. Applied Energy 2010;87(3):856–865.

[7] Jin T, Tian Z. Uncertainty analysis for wind energy production with dynamic power curves. In: 2010 IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems IEEE; 2010. p. 745–750.

[8] Petrone G, de Nicola C, Quagliarella D, Witteveen J, Iaccarino G. Wind turbine performance analysis under uncertainty. In: 49th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition; 2011. p. 544.

[9] Wang Y, Réthoré PE, van der Laan P, Murcia Leon JP, Liu Y, Li L. Multi-fidelity wake modelling based on Co-Kriging method. In: Journal of Physics: Conference Series (Online), vol. 753 IOP Publishing; 2016. p. 032065.

[10] Murcia JP, Réthoré PE, Dimitrov N, Natarajan A, Sørensen JD, Graf P, et al. Uncertainty propagation through an aeroelastic wind turbine model using polynomial surrogates. Renewable Energy 2018;119:910–922.

[11] Sørensen JD, Toft HS. Probabilistic design of wind turbines. Energies 2010;3(2):241–257.

[12] Van Buren KL, Mollineaux MG, Hemez FM, Atamturktur S. Simulating the dynamics of wind turbine blades: part II, model validation and uncertainty quantification. Wind Energy 2013;16(5):741–758.

[13] Branlard E, Jonkman J, Dana S, Doubrawa P. A digital twin based on OpenFAST linearizations for real-time load and fatigue estimation of land-based turbines. In: Journal of Physics: Conference Series, vol. 1618 IOP Publishing; 2020. p. 022030.

[14] Freebury G, Musial W. Determining Equivalent Damage Loading for Full-Scale Wind Turbine Blade Fatigue Tests. In: 2000 ASME Wind Energy Symposium; 2000. p. 50.

[15] Sobol' IM. Sensitivity estimates for nonlinear mathematical models. Mathematical modelling and computational experiments 1993;1(4):407–414.

[16] Forrester AI, Sóbester A, Keane AJ. Multi-fidelity optimization via surrogate modelling. Proceedings of the royal society a: mathematical, physical and engineering sciences 2007;463(2088):3251–3269.

[17] Forrester AI, Keane AJ. Recent advances in surrogate-based optimization. Progress in aerospace sciences 2009;45(1-3):50–79.

[18] Perdikaris P, Venturi D, Royset JO, Karniadakis GE. Multi-fidelity modelling via recursive co-kriging and Gaussian–Markov random fields. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 2015;471(2179):20150018.

[19] Perdikaris P, Venturi D, Karniadakis GE. Multifidelity information fusion algorithms for high-dimensional systems and massive data sets. SIAM Journal on Scientific Computing 2016;38(4):B521–B538.

[20] Parussini L, Venturi D, Perdikaris P, Karniadakis GE. Multi-fidelity Gaussian process regression for prediction of random fields. Journal of Computational Physics 2017;336:36–50.

[21] Le Gratiet L, Marelli S, Sudret B. Metamodel-Based Sensitivity Analysis: Polynomial Chaos Expansions and Gaussian Processes. In: Handbook of Uncertainty Quantification Springer International Publishing; 2017.p. 1–37.

[22] Evensen G. Data assimilation: the ensemble Kalman filter. Springer Science & Business Media; 2009.

[23] Hadamard, Jacques. Sur les problèmes aux dérivées partielles et leur signification physique. Princeton university bulletin 1902;p. 49–52.

[24] Dobre S, Bastogne T, Profeta C, Barberi-Heyob M, Richard A. Limits of variance-based sensitivity analysis for non-identifiability testing in high dimensional dynamic models. Automatica 2012;48(11):2740–2749.

[25] Lamboni M, Monod H, Makowski D. Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. Reliability Engineering and System Safety 2011;96(4):450–459.

[26] Saltelli A, Chan K, Scott EM, editors. Sensitivity analysis. Wiley series in probability and statistics, New York and Chichester and Weinheim etc.: J. Wiley & sons; 2000.

[27] Sobol' IM. On sensitivity estimation for nonlinear mathematical models. Matematicheskoe modelirovanie 1990;2(1):112–118.

[28] Etoré P, Prieur C, Pham DK, Li L. Global sensitivity analysis for models described by stochastic differential equations; 2018, https://hal.archives-ouvertes.fr/hal-01926919, working paper or preprint.

[29] Hoeffding W. A class of statistics with asymptotically normal distribution. Springer; 1992.

[30] Owen AB. Monte Carlo theory, methods and examples; 2013.

[31] Saltelli A. Making best use of model evaluations to compute sensitivity indices. Computer physics communications 2002;145(2):280–297.

[32] Le Gratiet L, Cannamela C, Iooss B. A Bayesian approach for global sensitivity analysis of (multifidelity) computer codes. SIAM/ASA Journal on Uncertainty Quantification 2014;2(1):336–363.

[33] Krige DG, Guarascio M, Camisani-Calzolari FA. Early South African geostatistical techniques in today's perspective. Geostatistics 1989;1:1–19.

[34] Stein ML. Interpolation of spatial data: some theory for kriging. Springer Science & Business Media; 2012.

[35] Rasmussen CE. Gaussian processes in machine learning. In: Summer School on Machine Learning Springer; 2003. p. 63–71.

[36] Duvenaud D, The Kernel cookbook: Advice on covariance functions. URl: http://www. cs. toronto. edu/˜ duvenaud/cookbook; 2014.

[37] Bachoc F. Cross validation and maximum likelihood estimations of hyper-parameters of Gaussian processes with model misspecification. Computational Statistics & Data Analysis 2013;66:55–69.

[38] McKay MD, Beckman RJ, Conover WJ. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 1979;21(2):239–245.

[39] Marrel A, Iooss B, Laurent B, Roustant O. Calculations of Sobol indices for the Gaussian process metamodel. Reliability Engineering & System Safety 2009;94(3):742 – 751.

[40] Homma T, Saltelli A. Importance measures in global sensitivity analysis of nonlinear models. Reliability Engineering & System Safety 1996;52(1):1–17.

[41] Janon A. Analyse de sensibilité et réduction de dimension. Application à l'océanographie. PhD thesis; 2012.

[42] Roustant O, Ginsbourger D, Deville Y. DiceKriging, DiceOptim: Two R Packages for the Analysis of Computer Experiments by Kriging-Based Metamodeling and Optimization. Journal of Statistical Software 2012;51(1):1–55. http://www.jstatsoft.org/v51/i01/.

[43] Iooss B, Janon A, Pujol G, with contributions from Baptiste Broto, Boumhaout K, Veiga SD, et al. sensitivity: Global Sensitivity Analysis of Model Outputs; 2019, https://CRAN.R-project.org/package=sensitivity, r package version 1.16.0.

[44] Kovachki NB, Stuart AM. Ensemble Kalman inversion: a derivative-free technique for machine learning tasks. Inverse Problems 2019;35(9):095005.

[45] Kalman RE, et al. Contributions to the theory of optimal control. Boletin de la Sociedad Matematica Mexicana 1960;5(2):102–119.

[46] Johnson ME, Moore LM, Ylvisaker D. Minimax and maximin distance designs. Journal of Statistical Planning and Inference 1990;26(2):131 – 148. http://www.sciencedirect.com/science/article/pii/037837589090122B.

[47] Snyder C, Zhang F. Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. Monthly Weather Review 2003;131(8).

[48] Welch G, Bishop G. An Introduction to the Kalman Filter. USA; 1995.

[49] Kopp RE, Orford RJ. Linear regression applied to system identification for adaptive control systems. Aiaa Journal 1963;1(10):2300–2306.

[50] Houtekamer PL, Mitchell HL. A sequential ensemble Kalman filter for atmospheric data assimilation. Monthly Weather Review 2001;129(1):123–137.

[51] Ambadan JT, Tang Y. Sigma-point Kalman filter data assimilation methods for strongly nonlinear systems. Journal of the Atmospheric Sciences 2009;66(2):261–285.

[52] Lawson GW, Hansen JA. Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth. Monthly weather review 2004;132(8):1966–1981.

[53] The National Renewable Energy Laboratory, OpenFAST; 2018. http://openfast.readthedocs.io.

[54] DNV GL, Bladed: Wind Turbine Design Software; 2013.

[55] Larsen TJ, Hansen AM. How 2 HAWC2, the user's manual; 2007.

[56] Principia, Principia 2019 Deeplines Wind;. http://www.principia-group.com/blog/product/deeplines-wind/.

[57] Le Cunff C, Heurtier JM, Piriou L, Berhault C, Perdrizet T, Teixeira D, et al. Fully Coupled Floating Wind Turbine Simulator Based on Nonlinear Finite Element Method: Part I Methodology. In: ASME 2013 32nd International Conference on Ocean, Offshore and Arctic Engineering Citeseer; 2013. p. V008T09A050–V008T09A050.

[58] Sutherland HJ. On the fatigue analysis of wind turbines. Sandia National Labs., Albuquerque, NM (US); Sandia National Labs ; 1999.

[59] Cosack N. Fatigue load monitoring with standard wind turbine signals. PhD thesis; 2011.

[60] Commission IE, et al. IEC 61400-1. Wind Turbines–Part 2005;1.

[61] Loads and site conditions for wind turbines. DNV GL; 2016.

[62] Witcher D, Uncertainty Quantification Techniques in Wind Turbine. DNV GL; 2017.

[63] Robertson AN, Shaler K, Sethuraman L, Jonkman J. Sensitivity of Uncertainty in Wind Characteristics and Wind Turbine Properties on Wind Turbine Extreme and Fatigue Loads. Wind Energy Science Discussions 2019;p. 1–41. https://www.wind-energ-sci-discuss.net/wes-2019-2/.

[64] Koukoura C. Validated Loads Prediction Models for Offshore Wind Turbines for Enhanced Component Reliability. PhD thesis, Technical University of Denmark; 2014.

[65] Holierhoek J, Korterink H, van de Pieterman R, Rademakers L, Lekou D. Recommended Practices for Measuring in Situ the 'Loads' on Drive Train, Pitch System and Yaw System. Energy Research Center of the Netherlands (ECN) 2010;.

[66] Quick J, Annoni J, King R, Dykes K, Fleming P, Ning A. Optimization under uncertainty for wake steering strategies. In: Journal of Physics: Conference Series, vol. 854 IOP Publishing; 2017. p. 012036.

[67] Simms D, Schreck S, Hand M, Fingersh LJ. NREL unsteady aerodynamics experiment in the NASA-Ames wind tunnel: a comparison of predictions to measurements. National Renewable Energy Lab., Golden, CO (US); 2001.

[68] Welch P. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Transactions on audio and electroacoustics 1967;15(2):70–73.

[69] Jansen MJ. Analysis of variance designs for model output. Computer Physics Communications 1999;117(1-2):35–43.

[70] Wold S, Esbensen K, Geladi P. Principal component analysis. Chemometrics and intelligent laboratory systems 1987;2(1-3):37–52.

[71] Eknes M, Evensen G. An ensemble Kalman filter with a 1-D marine ecosystem model. Journal of Marine Systems 2002;36(1-2):75–100.

[72] Houtekamer PL, Mitchell HL, Pellerin G, Buehner M, Charron M, Spacek L, et al. Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. Monthly weather review 2005;133(3):604–620.