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

Adrien Hirvoas, Clémentine Prieur, Elise Arnaud, Fabien Caleyron, Miguel Zuniga

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

Adrien Hirvoas[a], Clémentine Prieur[b], Elise Arnaud[b], Fabien Caleyron[c], Miguel Munoz Zuniga[c]

[a] *adrien.hirvoas@ifpen.fr, IFP Energies Nouvelles — Lyon site*
Rond-point de l'échangeur de Solaize BP 3, 69360 Solaize, France
[b] *LJK — Laboratoire Jean Kuntzmann, Université Grenoble Alpes*
[c] *IFP Energies Nouvelles*

Abstract

A framework to perform quantification and reduction of uncertainties in a wind turbine numerical model using global sensitivity analysis and 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 modelling, 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. Hereafter, global sensitivity analysis is also used in the context of identifiability.

Keywords: Inference, uncertainty quantification, model calibration, nonlinear Kalman filters, Global Sensitivity Analysis

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 turbines 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]. In this context, uncertainty quantification and reduction methods have been developed [2].

However, the conventional methods used in uncertainty quantification are not suitable in the present industrial context because of the stochastic nature of the external solicitation and the time consuming behavior of the simulator [3, 4]. Formally, the function $\mathcal{M}$ representing the time-consuming numerical model is defined as:

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

where, $\mathbf{x} = (x_1, \ldots, x_p) \in \mathcal{P} \subseteq \mathbb{R}^p$ are the model input parameters, V is a stochastic process modelling the external solicitation, $\mathbf{y} = (y_1, \ldots, y_m) \in \mathbb{R}^m$ is the vector of discretized functional output; e.g., generated power, structural accelerations or loads.

We aim at investigating a complete framework to quantify and reduce the input parameter uncertainties involved in a finite element wind turbine model. Our main contribution is twofold.

Firstly, the framework allows to quantifying the sources of uncertainties affecting the fatigue behavior of the structural components of the wind turbine. 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) [5]:

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

Global Sensitivity Analysis (GSA) methods have been developed to quantify the uncertainty in QoI, see Eq. (2), 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 [6], 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. However, classical GP model-based GSA cannot tackle the inherent randomness from stochastic simulation. We propose to model the mean of the QoI $\mathbb{E}_V[g \circ \mathcal{M}(\mathbf{X}, V)]$ with a GP model with heteroscedastic nugget effect. Then, this surrogate model is used to perform a sensitivity analysis based on classical MC estimation procedure [7].

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, 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 carry out the calibration using a recursive Bayesian inference approach based on an Ensemble Kalman Filter (EnKF) [8].

However, such problems can be solved assuming that several conditions of well-posedness and identifiability are achieved. These conditions have been summarized by Hadamard [9]. As highlighted in [10], 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 performed 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 [11].

Lastly 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 [12]. Among the large number of available approaches, variance-based sensitivity analysis introduced by [13] 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

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

taken into account when estimating sensitivity measures under interest. More precisely, let us use the formalism introduced in Eq. (2) to the model in hand such that:

$$\mathbf{y} = \mathcal{M}(\mathbf{x}, V) \tag{3}$$

where $\mathcal{M}$ is the wind turbine numerical model, $\mathbf{x} = (x_1, \ldots, x_p) \in \mathcal{P} \subseteq \mathbb{R}^p$ are the model input parameters, V is a stochastic process modelling the external solicitation and $\mathbf{y} = (y_1, \ldots, y_m) \in \mathbb{R}^m$ is the output vector.

We are interested in measuring the sensitivity of a QoI $g(\mathbf{y}) \in \mathbb{R}$, see Eq. 2, 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 stochasticity propagated from the model itself through V, which is assumed to be independent of $\mathbf{X}$. There exists two approaches to deal with this stochasticity in a GSA framework. The first is interested in the full probability distribution of the QoI while the other one is only concerned with the QoI averaged over the inherent randomness of the physical system. The latter is the one considered in this work. We are therefore interested in the sensitivity of the deterministic function f defined as:

$$Z = f(\mathbf{X}) = \mathbb{E}_V[g \circ \mathcal{M}(\mathbf{X}, V)]\,. \tag{4}$$

The total variance of Z 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 [14]). Each part of variance corresponds to the contribution of each set of parameters on the variance of the output Z. By considering the ratio of each part of variance to the total variance of Z, we obtain a measure of importance for each set of input parameters that is called the Sobol' index [13]. More precisely, assuming $Var_{\mathbf{X}}(Z) < +\infty$, $Var_{\mathbf{X}}(Z) \neq 0$, we define for any $\mathbf{u} \subset \{1, \ldots, p\}$ the closed Sobol' index of order $r = card(\mathbf{u})$ associated to the set of inputs $\mathbf{X_u} = \{X_j\}_{j \in \mathbf{u}}$ as:

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

where $-\mathbf{u}$ is the complement of $\mathbf{u}$ in $\{1, \ldots, p\}$. This index quantifies the main effect of all the variables within $\mathbf{X_u}$, including interactions, on Z. 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:

$$S^T_{\mathbf{u}} = 1 - S_{\mathbf{-u}} = 1 - \frac{Var_{\mathbf{X_{-u}}}(\mathbb{E}_{\mathbf{X_u}}[Z|\mathbf{X_{-u}}])}{Var_{\mathbf{X}}(Z)} \tag{6}$$

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 Z.

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 estimations under reasonable time constraints. To handle this issue, one may rely on a metamodel to compute cheap approximated 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 [15]. 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 [16, 17] 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 $\mathbf{C}$. Then , for any $\mathbf{x} \in \mathcal{P}$, $f(\mathbf{x})$ is approximated by the conditional Gaussian process $\{Z_n(\mathbf{x}),\ \mathbf{x} \in \mathcal{P}\} := \{[Z(\mathbf{x})|Z(\mathbf{D}) = \mathbf{z}],\ \mathbf{x} \in \mathcal{P}\}$, where $\mathbf{z} = (z_1, \dots z_n)$ are evaluations of f on n points $\mathbf{D} = \{\mathbf{d}_1, \dots, \mathbf{d}_n\}$, $\mathbf{d}_i \in \mathcal{P}$. The design $\{(\mathbf{d}_1, z_1), \dots, (\mathbf{d}_n, z_n)\}$ is called the learning sample. In the following, $\mathbf{D}$ is chosen as a Latin Hypercube sample to guarantee a good exploration of our numerical model. We then get the ordinary Kriging equations:

$$Z_n(\mathbf{x}) \sim \mathcal{N}(m_{OK}(\mathbf{x}), s^2_{OK}(\mathbf{x}))\,, \tag{7}$$

with

$$m_{OK}(\mathbf{x}) = \mu(\mathbf{x}) + \mathbf{c}(\mathbf{x})^T(\mathbf{C})^{-1}(\mathbf{z} - \boldsymbol{\mu})\,, \tag{8}$$

$$s^2_{OK}(\mathbf{x}) = C(\mathbf{x}, \mathbf{x}) - \mathbf{c}(\mathbf{x})^T(\mathbf{C})^{-1}\mathbf{c}(\mathbf{x})\,. \tag{9}$$

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

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 $\tilde{z}_i = f(\mathbf{d}_i) + \varepsilon_i$. Where, $f(\mathbf{d}_i)$ is defined as an empirical mean computed from a K-sample of $g \circ \mathcal{M}(\mathbf{d}_i, V)$ and ε_i is a centered noise whose variance is $\tau_i^2 = \frac{1}{K}\left(\frac{1}{K-1}(\sum_{j=1}^K(\mathcal{M}(\mathbf{d}_i, V = v_j) - \frac{1}{K}\sum_{j=1}^K \mathcal{M}(\mathbf{d}_i, V = v_j))^2\right)$. 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 $\mathrm{diag}(\tau_1^2, \dots, \tau_n^2)$. Then, provided that the process Z and the Gaussian measurement errors ε_i are stochastically independent, the process Z is still Gaussian conditionally on the noisy observations $\tilde{z}_i$ and its conditional mean and variance functions are given by the following slightly modified kriging equations:

$$\tilde{Z}_n(\mathbf{x}) \sim \mathcal{N}(m_{NK}(\mathbf{x}), s^2_{NK}(\mathbf{x})) \tag{10}$$

with

$$m_{NK}(\mathbf{x}) = \mu(\mathbf{x}) + \mathbf{c}(\mathbf{x})^T(\mathbf{C} + \boldsymbol{\Delta})^{-1}(\tilde{\mathbf{z}} - \boldsymbol{\mu})\,, \tag{11}$$

$$s^2_{NK}(\mathbf{x}) = C(\mathbf{x}, \mathbf{x}) - \mathbf{c}(\mathbf{x})^T(\mathbf{C} + \boldsymbol{\Delta})^{-1}\mathbf{c}(\mathbf{x})\,. \tag{12}$$

Kriging based Sobol' indices - Following [15] and [18] the idea is to substitute Z with $\tilde{Z}_n$ in Eq. (5):

$$\tilde{S}_{\mathbf{u}} = \frac{Var_{\mathbf{X_u}}(\mathbb{E}_{\mathbf{X_{-u}}}[\,\tilde{Z}_n(\mathbf{X}) \mid \mathbf{X_u}])}{Var_{\mathbf{X}}(\tilde{Z}_n(\mathbf{X}))}\,. \tag{13}$$

As $\tilde{Z}_n$ 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 $\mathbf{u} \subset \{1, \dots, p\}$, $\mathbf{x}_{i,\mathbf{u}} = \{x_{i,j}\}$, $j \in \mathbf{u}$. 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 (13) is as follows:

$$\widehat{\tilde{S}}_{\mathbf{u}} = \frac{\frac{1}{s}\sum_{i=1}^{s} \tilde{Z}_n(\mathbf{x}_i)\ \tilde{Z}_n(\mathbf{x}_{i,\mathbf{u}} : \mathbf{x}'_{i,-\mathbf{u}}) - \frac{1}{s}\sum_{i=1}^{s} \tilde{Z}_n(\mathbf{x}_i)\ \frac{1}{s}\sum_{i=1}^{s} \tilde{Z}_n(\mathbf{x}_{i,\mathbf{u}} : \mathbf{x}'_{i,-\mathbf{u}})}{\frac{1}{s}\sum_{i=1}^{s} \tilde{Z}_n(\mathbf{x}_i)^2 - (\frac{1}{s}\sum_{i=1}^{s} \tilde{Z}_n(\mathbf{x}_{i,\mathbf{u}} : \mathbf{x}'_{i,-\mathbf{u}}))^2}\,. \tag{14}$$

Confidence intervals can be obtained via a bootstrap method, as described in Algorithm 1 of [15]. 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 [19, 20].

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. The problematic 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 [21].
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 $\{k+1\}_{k=0\ldots T-1}$ as:

$$\left\{ \begin{array}{rcl} \mathbf{x}_{(k+1)} & = & \mathbf{x}_{(k)} + \delta_{(k)} \\ \mathbf{y}_{(k+1)} & = & \mathcal{M}(\mathbf{x}_{(k+1)}, V = v_{(k+1)}) + \boldsymbol{\epsilon}_{(k+1)} \end{array} \right. \tag{15}$$

where $\mathbf{x}_{(k+1)}$ is the input parameter vector, $v_{(k+1)}$ is a known realization of the stochastic external excitation at $k+1$, the Gaussian white noises $\delta_{(k)} \sim \mathcal{N}(0, \mathbf{Q}_{(k)})$ and $\boldsymbol{\epsilon}_{(k+1)} \sim \mathcal{N}(0, \mathbf{R}_{(k)})$ are respectively an artificial dynamic noise and a combination of the model and observation errors. For the sake of readability, $\mathbf{y}$ 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. (15) using the known input solicitation and the available measurements. Among all available filtering methods, the Kalman Filter (KF) [22] 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) [8] 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 an efficient stratified sampling technique instead of a conventional Monte Carlo method to generate the initial ensemble of parameter estimates. The developed technique is based on a Latin Hypercube Sampling (LHS) coupled with a geometrical criteria maximizing the minimum distance between the design points, see [23].
In the field of inverse problems, this inference method is referred to as Ensemble Kalman Inversion (EKI). The EnKF formulation used in [24] is adopted in this paper. At iteration k, we have an ensemble of size N_{ens}, of forecast parameter estimates $\mathbf{x}^f_{(k)} = [\mathbf{x}^{f(1)}_{(k)}, \ldots, \mathbf{x}^{f(N_{ens})}_{(k)}] \in \mathbb{R}^{p \times N_{ens}}$ where the superscript $.^{f(i)}$ denotes the i-th forecast member of the ensemble. The mean of the forecast

members of the ensemble is given by:

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

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

$$\mathbf{P}^f_{(k)} = \frac{1}{N_{ens}-1} \sum_{i=1}^{N_{ens}} (\mathbf{x}^{f(i)}_{(k)} - \bar{\mathbf{x}}^f_{(k)})(\mathbf{x}^{f(i)}_{(k)} - \bar{\mathbf{x}}^f_{(k)})^T \cdot \tag{17}$$

As said previously the structure of the EnKF is the same as the one of the Kalman filter [25]. Thus, we need to compute the Kalman Gain, referred as $\mathbf{K}_{(k)}$ and defined by:

$$\mathbf{K}_{(k)} = \mathbf{P}^f_{(k)} \mathbf{M}^T \left(\mathbf{M} \mathbf{P}^f_{(k)} \mathbf{M}^T + \mathbf{R}_{(k)} \right)^{-1} , \tag{18}$$

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 [26].
Nevertheless, for most applications this condition of linear (or linearized) observation operator cannot be applied. In that context, as proposed in [27], 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:

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

and,

$$\frac{1}{N_{ens}-1} \sum_{i=1}^{N_{ens}} \left(\mathcal{M}(\mathbf{x}^{f(i)}_{(k)}, V = v_{(k)}) - \mathcal{M}(\bar{\mathbf{x}}^f_{(k)}, V = v_{(k)}) \right) \left(\mathcal{M}(\mathbf{x}^{f(i)}_{(k)}, V = v_{(k)}) - \mathcal{M}(\bar{\mathbf{x}}^f_{(k)}, V = v_{(k)}) \right)^T \tag{20}$$

It has been argued, in [28], that Eq. (19) and Eq. (20) are good approximations if the following equations are verified:

$$\begin{aligned} \mathcal{M}(\bar{\mathbf{x}}^f_{(k)}, V = v_{(k)}) = \overline{\mathcal{M}(\mathbf{x}^f_{(k)}, V = v_{(k)})} = \frac{1}{N_{ens}} \sum_{i=1}^{N_{ens}} \mathcal{M}(\mathbf{x}^{f(i)}_{(k)}, V = v_{(k)}) \, , \\ \mathbf{x}^{f(i)}_{(k)} - \bar{\mathbf{x}}^f_{(k)} = \xi_i \text{ and } \|\xi_i\| \text{ is small for } i = 1 \dots N_{ens} \cdot \end{aligned} \tag{21}$$

The difference between the deterministic EnKF and the stochastic version presented hereafter is that the observations are now treated as random variables. 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:

$$\mathbf{y}^{(i)}_{(k)} = \mathbf{y}_{(k)} + \mathbf{e}^{o(i)}_{(k)} \, , \text{ with } \mathbf{e}^{o(i)}_{(k)} \sim \mathcal{N}(0, \mathbf{R}_{(k)}), i = 1 \dots N_{ens} \cdot \tag{22}$$

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 [29]. 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:

$$\mathbf{x}_{(k)}^{a(i)} = \mathbf{x}_{(k)}^{f(i)} + \mathbf{K}_{(k)} \left(\mathbf{y}_{(k)}^{(i)} - \mathcal{M}(\mathbf{x}_{(k)}^{f(i)}, V = v_{(k)}) \right) \cdot \tag{23}$$

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, see Eq. (24). The presented method is fully described in Algo. 1.

$$\mathbf{x}_{(k+1)}^{f} = \mathbf{x}_{(k)}^{a} + \delta_{(k)} \,, \text{ with } \delta_{(k)} \sim \mathcal{N}(0, \mathbf{Q}_{(k)}) \cdot \tag{24}$$

Data:
number of members in the ensemble N_{ens};
prior guess of the parameter vector $\mathbf{x}_b$;
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 $\{\mathbf{R}_{(k)}\}_{k=1,\dots,T}$;
artificial error covariance matrix $\{\mathbf{Q}_{(k)}\}_{k=0,\dots,T}$.

Initialisation step:

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

$$\mathbf{x}_0^{a(i)} = \mathbf{x}_b + \boldsymbol{\epsilon}^b \text{ with, } \boldsymbol{\epsilon}^b \sim \mathcal{N}(0, \mathbf{P}_b)$$

end

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

Forecast step:

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

$$\mathbf{x}_{(k)}^{f(i)} = \mathbf{x}_{(k-1)}^{a(i)}$$

end

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

Update step:

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

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

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

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

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

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

end

end

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

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 [30], Bladed [31], HAWCK2 [32] or Deeplines Wind [33].

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 [34]. 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 [3], 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. (2), the model for such stochastic simulation can be formulated as follows:

$$\mathbf{y} = \mathcal{M}(\mathbf{x}, V)\,, \tag{25}$$

where $\mathcal{M}$ is the Deeplines Wind model, $\mathbf{x} = (x_1, ..., x_p) \in \mathcal{P} \subseteq \mathbb{R}^p$ are the model input parameters, V is the stochastic Turbsim wind simulator, $\mathbf{y} = (y_1, ..., y_m) \in \mathbb{R}^m$ is the vector of discretized functional output.

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 previously. In this paper we will perform the study at an under-rated 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 auto-correlation studies of the outputs. This period permits to remove the effect of the ramp time period on the numerical model responses. Each numerical simulation lasts between 15 and 30 minutes according to the CPU performance.

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. 1 and Tab. 1. We

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

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

[4]International Electrotechnical Commission

can denote by g the function mapping the functional loads of the structure to the damage QoIs such as:

$$g(\mathbf{y}) := g \circ \mathcal{M}(\mathbf{x}, V) \tag{26}$$

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 as the considered irregular load history. The DEL is computed based on the Palgrem Miner's rule [35]:

$$DEL = \left(\frac{\sum_{i=1}^{N_c} S_i \cdot N_i^m}{N_{ref}} \right)^{\frac{1}{m}} . \tag{27}$$

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 [36].

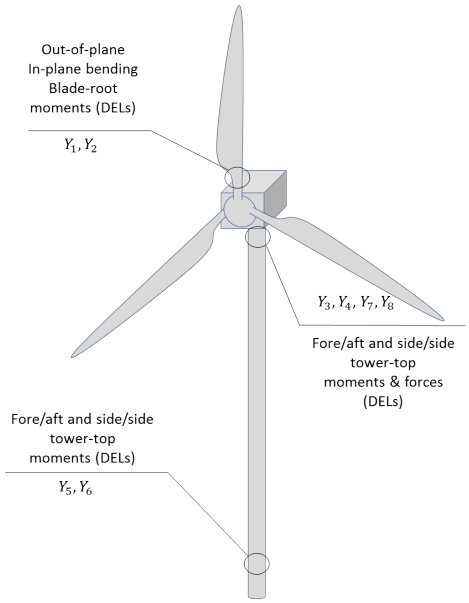


Figure 1: Recorded time series of loads at different locations of the wind turbine

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 of the input parameters is distinguishable from 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 location of the wind turbine, see Eq. (27). The number of stochastic process realizations used during the convergence study varies from 1 to 30.
Fig. 2 shows the convergence of the tower bottom fore-aft bending moment DEL on the number of realizations used for their averaging estimation. According to certification guidelines, see design load case 1.2 in [37], 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 realization to 10 for the GSA of the damage equivalent loads.

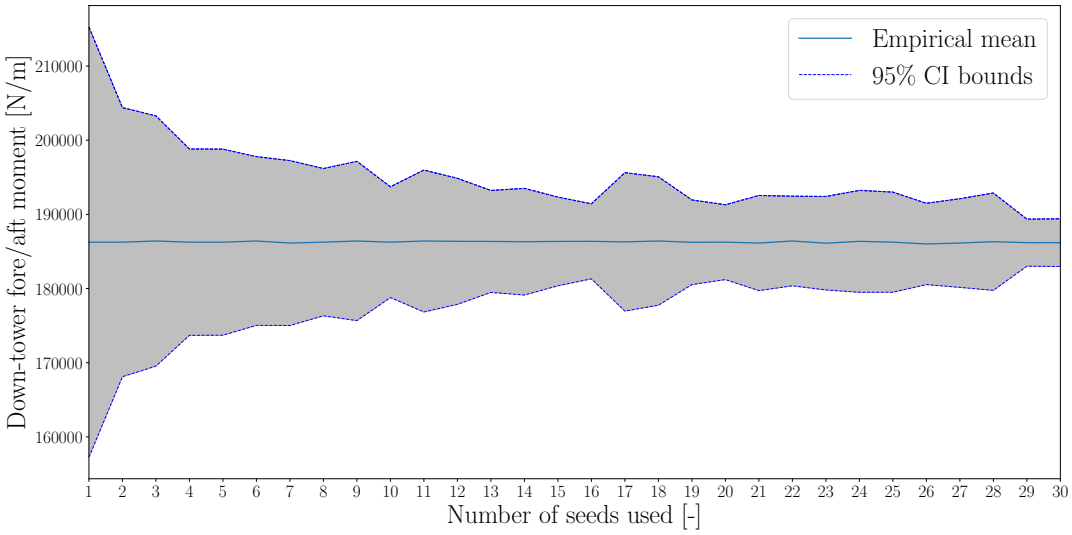


Figure 2: Convergence of the tower bottom fore-aft bending moment DEL as a function of the number of turbulent seeds used for its evaluation. 95% confidence interval around the estimated empirical mean is also represented (grey area).

The 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. All these parameters were considered independent of one another with Gaussian distributions.

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 $\pm 10\%$ of its reference value.

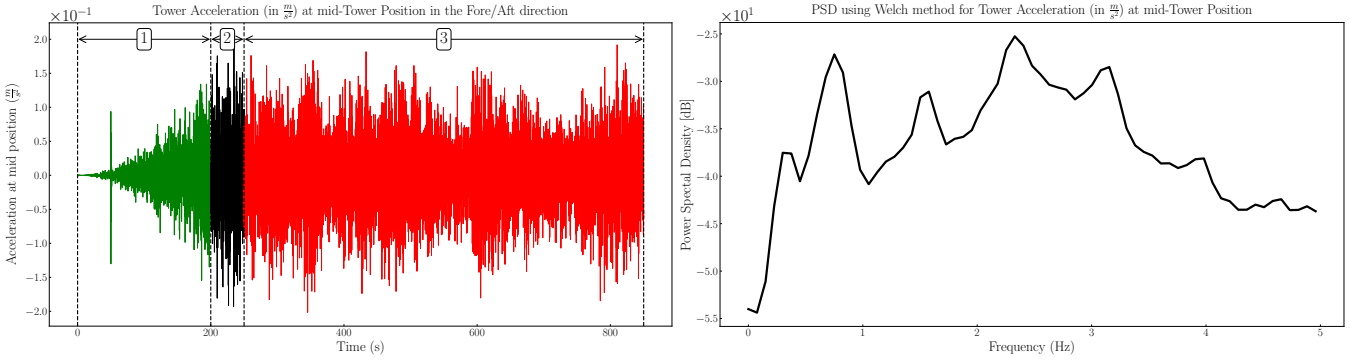
The uncertainties in blade structural properties have been represented using five parameters. The blade structural responses have led to the definition of the uncertainty range. Indeed, the frequency of the edge-wise (ES) and flap-wise (FS) mode were changed about 10% each from their nominal value. These 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	$\boldsymbol{\mu}$	$\boldsymbol{\sigma}$	REF
Nacelle Mass - N_{mass} [kg]	6.90e+04	2.30e+03	[38]
Nacelle center of mass - N_{CMx} [m]	1.00	3.33e-02	[39]
Tower thickness - e [%]	0	7	IFPEN $\pm 10\%$ 1 FA
Tower Rayleigh Damping - β_{TR} [-]	3.10e-02	9.93e-03	[40]
Inertial Nacelle - I_{zz} $[kg \cdot m^2]$	7.00e+05	2.33e+04	IFPEN $\pm 10\%$ μ
Drive-train Torsional stiffness - K_D $[\frac{N \cdot m^2}{rad}]$	4.45e+09	1.48e+08	[41]
Blade Flap wise Stiffness - α_{BF} $[N \cdot m^2]$	1.00	3.33e-02	IFPEN $\sim \pm 10\%$ 1 FS
Blade Edge wise Stiffness - α_{BE} $[N \cdot m^2]$	1.00	3.33e-02	IFPEN $\sim \pm 10\%$ 1 ES
Blade mass coefficient - α_{mass} [%]	1.00	1.67e-02	[38]
Blade Rayleigh Damping - β_{BR} [-]	5.39e-03	1.45e-03	[39]
Blade mass imbalance - η_B [%]	2.50	8.33e-01	[39]
Yaw misalignment - ω [°]	0	6.67	[42]
Individual pitch error - Ω [°]	0.10	3.33e-02	[43]

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 [44], see Fig. 3.



(a) (b)

Figure 3: 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 $\boxed{1}$, an oversight period $\boxed{2}$ and a dynamical period of interest $\boxed{3}$. On the right side (b): estimated PSD of the period of interest using Welch's method [44].

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, [45]. 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 [46]. 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 meta-modeling [15].

The results for the total Sobol' indices with their corresponding 95% confidence intervals are summarized in Fig. 4. A threshold of 2.5e-2 was chosen to display a separation between input parameters with high and low total Sobol' indices. Fig. 5 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.

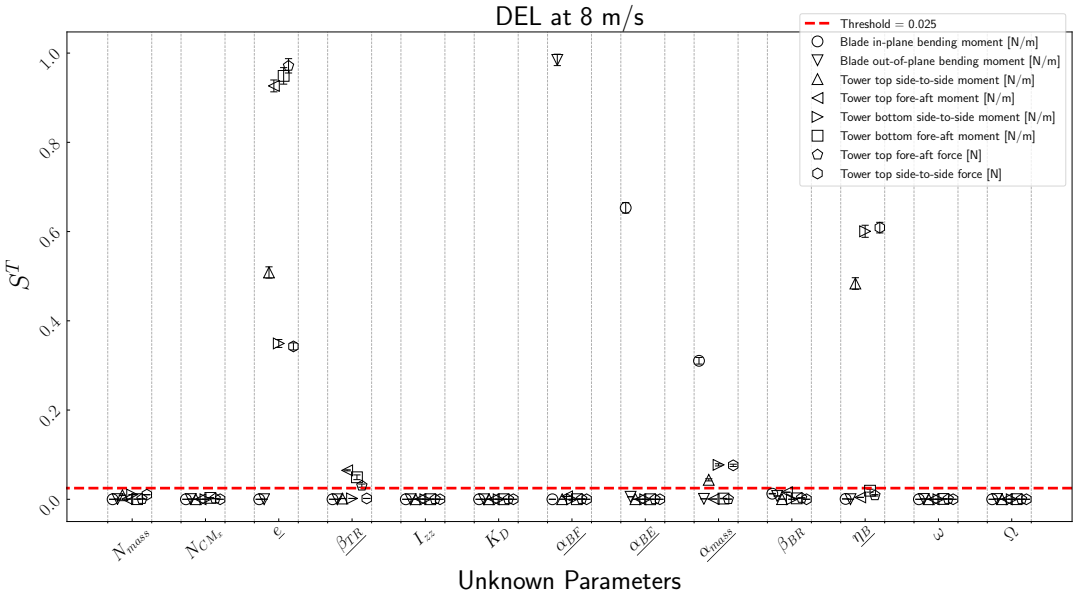


Figure 4: Estimation of total Sobol' indices (y-axis) with their 95% confidence interval corresponding to each of the 13 parameters (x-axis) for the different quantities of interest in fatigue. The dashed line corresponds to a threshold arbitrarily chosen to 2.5e-2. Confidence intervals (CI) are obtained by taking into account the uncertainties due to both the meta-model 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.

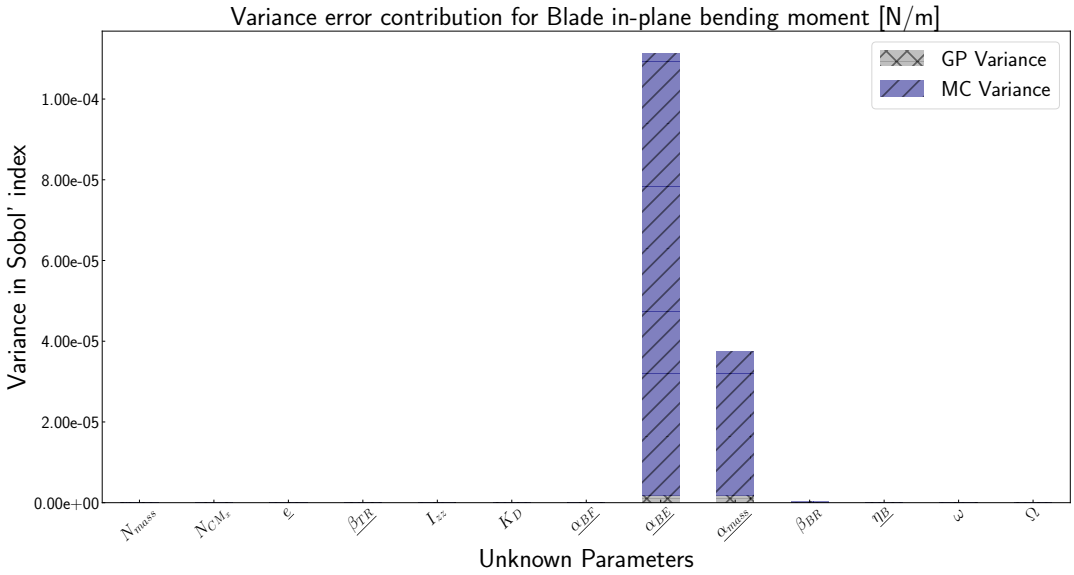


Figure 5: Splitting of the variance of total Sobol' index estimators (y-axis) corresponding to each of the 13 input parameters (x-axis)for the DEL of the out-of-plane bending blade-root moment. The number of samples for the conditional Gaussian process, in order to quantify the uncertainty of the Gaussian 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. 4, 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 life 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 [10] 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) [47]. 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 [11] , the proposed index synthesizes the influence of the parameter on the whole time series output.

In order to ensure that the variation of the input parameters is distinguishable from the realization of the stochastic 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 and total aggregated Sobol' indices obtained with the GP model built on the train set from the lastly mentioned stratified sampling. 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 indentifiable 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 $\boldsymbol{P}_b = \mathrm{diag}(7.00e-02,\, 9.93e-03,\, 3.33e-02,\, 3.33e-02,\, 1.67e-02,\, 8.33e-03)$.

For the sake of numerical time consumption of this test case, we use a number of inference iterations T equals to 20 and a number of members of the ensemble set at 100. Fig. 6 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

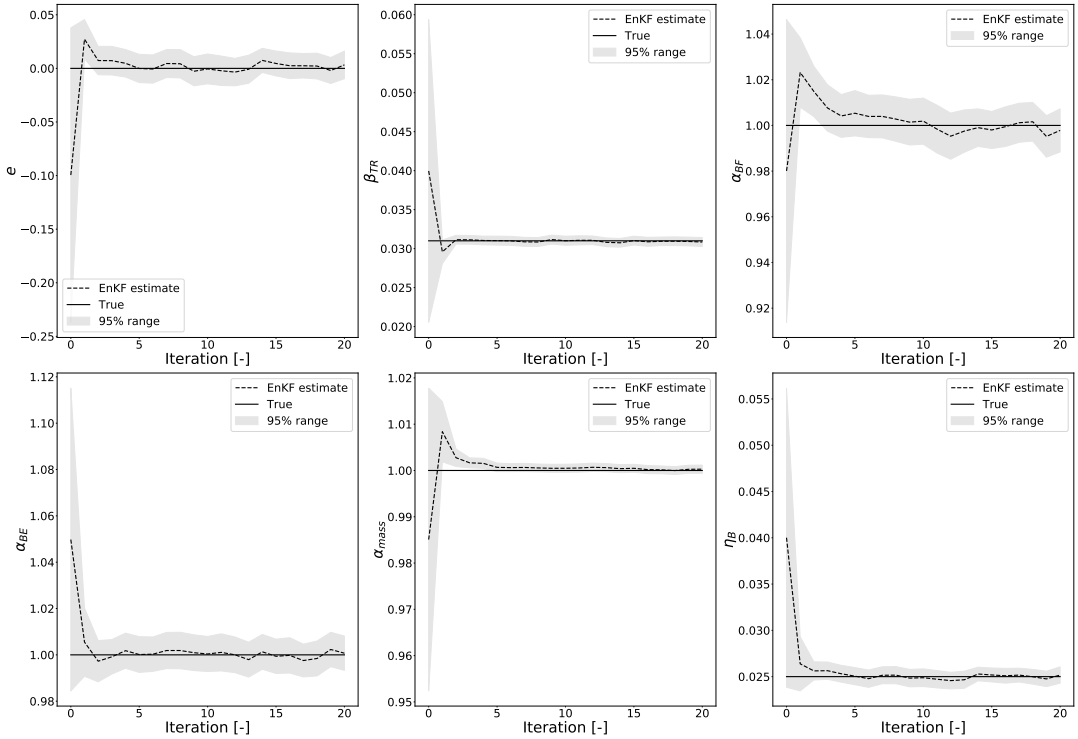


Figure 6: Iteration evolution of the a posteriori estimates of the input parameters. Results obtained by running EnKF presented in Section 2 with $N = 100$ members of the ensemble used for the estimation and considering pseudo-experimental numerical observations.

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 the identification of input parameters much harder to perform. 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 influent 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.

References

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

[2] R. C. Smith, *Uncertainty quantification: theory, implementation, and applications*, vol. 12. Siam, 2013.

[3] T. Perdrizet, J.-C. Gilloteaux, D. Teixeira, G. Ferrer, L. Piriou, D. Cadiou, J.-M. Heurtier, and C. Le Cunff, "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*, pp. V008T09A052–V008T09A052, Citeseer, 2013.

[4] B. J. Jonkman, "Turbsim user's guide: Version 1.50," tech. rep., National Renewable Energy Lab (NREL), Golden, CO, USA, 2009.

[5] G. Freebury and W. Musial, "Determining equivalent damage loading for full-scale wind turbine blade fatigue tests," in *2000 ASME Wind Energy Symposium*, p. 50, 01 2000.

[6] I. M. Sobol', "Sensitivity estimates for nonlinear mathematical models," *Mathematical modelling and computational experiments*, vol. 1, no. 4, pp. 407–414, 1993.

[7] L. Le Gratiet, S. Marelli, and B. Sudret, "Metamodel-Based Sensitivity Analysis: Polynomial Chaos Expansions and Gaussian Processes," in *Handbook of Uncertainty Quantification*, pp. 1–37, Springer International Publishing, 2017.

[8] G. Evensen, *Data assimilation: the ensemble Kalman filter*. Springer Science & Business Media, 2009.

[9] Hadamard, Jacques, "Sur les problèmes aux dérivées partielles et leur signification physique," *Princeton university bulletin*, pp. 49–52, 1902.

[10] S. Dobre, T. Bastogne, C. Profeta, M. Barberi-Heyob, and A. Richard, "Limits of variance-based sensitivity analysis for non-identifiability testing in high dimensional dynamic models," *Automatica*, vol. 48, no. 11, pp. 2740–2749, 2012.

[11] M. Lamboni, H. Monod, and D. Makowski, "Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models," *Reliability Engineering and System Safety*, vol. 96, no. 4, pp. 450–459, 2011.

[12] A. Saltelli, K. Chan, and E. M. Scott, eds., *Sensitivity analysis*. Wiley series in probability and statistics, New York and Chichester and Weinheim etc.: J. Wiley & sons, 2000.

[13] I. M. Sobol', "On sensitivity estimation for nonlinear mathematical models," *Matematicheskoe modelirovanie*, vol. 2, no. 1, pp. 112–118, 1990.

[14] W. Hoeffding, *A class of statistics with asymptotically normal distribution*. Springer, 1992.

[15] L. Le Gratiet, C. Cannamela, and B. Iooss, "A bayesian approach for global sensitivity analysis of (multifidelity) computer codes," *SIAM/ASA Journal on Uncertainty Quantification*, vol. 2, no. 1, pp. 336–363, 2014.

[16] D. G. Krige, M. Guarascio, and F. A. Camisani-Calzolari, "Early South African geostatistical techniques in today's perspective," *Geostatistics*, vol. 1, pp. 1–19, 1989.

[17] C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer School on Machine Learning*, pp. 63–71, Springer, 2003.

[18] A. Marrel, B. Iooss, B. Laurent, and O. Roustant, "Calculations of sobol indices for the gaussian process metamodel," *Reliability Engineering & System Safety*, vol. 94, no. 3, pp. 742 – 751, 2009.

[19] O. Roustant, D. Ginsbourger, and Y. Deville, "DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization," *Journal of Statistical Software*, vol. 51, no. 1, pp. 1–55, 2012.

[20] B. Iooss, A. Janon, G. Pujol, with contributions from Baptiste Broto, K. Boumhaout, S. D. Veiga, T. Delage, J. Fruth, L. Gilquin, J. Guillaume, L. Le Gratiet, P. Lemaitre, B. L. Nelson, F. Monari, R. Oomen, O. Rakovec, B. Ramos, O. Roustant, E. Song, J. Staum, R. Sueur, T. Touati, and F. Weber, *sensitivity: Global Sensitivity Analysis of Model Outputs*, 2019. R package version 1.16.0.

[21] N. B. Kovachki and A. M. Stuart, "Ensemble kalman inversion: a derivative-free technique for machine learning tasks," *Inverse Problems*, vol. 35, no. 9, p. 095005, 2019.

[22] R. E. Kalman *et al.*, "Contributions to the theory of optimal control," *Boletin de la Sociedad Matematica Mexicana*, vol. 5, no. 2, pp. 102–119, 1960.

[23] M. E. Johnson, L. M. Moore, and D. Ylvisaker, "Minimax and maximin distance designs," *Journal of Statistical Planning and Inference*, vol. 26, no. 2, pp. 131 – 148, 1990.

[24] C. Snyder and F. Zhang, "Assimilation of simulated doppler radar observations with an ensemble kalman filter.," *Monthly Weather Review*, vol. 131, no. 8, 2003.

[25] G. Welch and G. Bishop, "An introduction to the kalman filter," tech. rep., USA, 1995.

[26] R. E. Kopp and R. J. Orford, "Linear regression applied to system identification for adaptive control systems," *Aiaa Journal*, vol. 1, no. 10, pp. 2300–2306, 1963.

[27] P. L. Houtekamer and H. L. Mitchell, "A sequential ensemble kalman filter for atmospheric data assimilation," *Monthly Weather Review*, vol. 129, no. 1, pp. 123–137, 2001.

[28] J. T. Ambadan and Y. Tang, "Sigma-point kalman filter data assimilation methods for strongly nonlinear systems," *Journal of the Atmospheric Sciences*, vol. 66, no. 2, pp. 261–285, 2009.

[29] G. W. Lawson and J. A. Hansen, "Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth," *Monthly weather review*, vol. 132, no. 8, pp. 1966–1981, 2004.

[30] The National Renewable Energy Laboratory, "Openfast." `http://openfast.readthedocs.io`, 2018.

[31] DNV GL, "Bladed: Wind turbine design software," 2013.

[32] T. J. Larsen and A. M. Hansen, *How 2 HAWC2, the user's manual*, 2007.

[33] Principia, "Principia 2019 deeplines wind."

[34] C. Le Cunff, J.-M. Heurtier, L. Piriou, C. Berhault, T. Perdrizet, D. Teixeira, G. Ferrer, and J.-C. Gilloteaux, "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*, pp. V008T09A050–V008T09A050, Citeseer, 2013.

[35] H. J. Sutherland, "On the fatigue analysis of wind turbines," tech. rep., Sandia National Labs., Albuquerque, NM (US); Sandia National Labs , 1999.

[36] N. Cosack, *Fatigue load monitoring with standard wind turbine signals*. PhD thesis, 01 2011.

[37] "Loads and site conditions for wind turbines," standard, DNV GL, November 2016.

[38] D. Witcher, "Uncertainty Quantification Techniques in Wind Turbine," 2017.

[39] A. N. Robertson, K. Shaler, L. Sethuraman, and J. Jonkman, "Sensitivity of uncertainty in wind characteristics and wind turbine properties on wind turbine extreme and fatigue loads," *Wind Energy Science Discussions*, pp. 1–41, 2019.

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

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

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

[43] D. Simms, S. Schreck, M. Hand, and L. J. Fingersh, "Nrel unsteady aerodynamics experiment in the nasa-ames wind tunnel: a comparison of predictions to measurements," tech. rep., National Renewable Energy Lab., Golden, CO (US), 2001.

[44] P. Welch, "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*, vol. 15, no. 2, pp. 70–73, 1967.

[45] M. Stein, "Large sample properties of simulations using latin hypercube sampling," *Technometrics*, vol. 29, no. 2, pp. 143–151, 1987.

[46] M. J. Jansen, "Analysis of variance designs for model output," *Computer Physics Communications*, vol. 117, no. 1-2, pp. 35–43, 1999.

[47] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.