



**HAL**  
open science

# 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, Élise Arnaud, Fabien Caleyron, Miguel Munoz Zuniga

## ► To cite this version:

Adrien Hirvoas, Clémentine Prieur, Élise Arnaud, Fabien Caleyron, Miguel Munoz Zuniga. Quantification and reduction of uncertainties in a wind turbine numerical model based on global sensitivity analysis and recursive Bayesian inference approach. 2020. hal-03279947v1

**HAL Id: hal-03279947**

**<https://inria.hal.science/hal-03279947v1>**

Preprint submitted on 22 Jun 2020 (v1), last revised 6 Jul 2021 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Adrien Hirvoas<sup>a</sup>, Clémentine Prieur<sup>b</sup>, Elise Arnaud<sup>b</sup>, Fabien Caleyron<sup>c</sup>, Miguel Munoz Zuniga<sup>c</sup>

<sup>a</sup>*adrien.hirvoas@ifpen.fr, IFP Energies Nouvelles — Lyon site*

*Rond-point de l'échangeur de Solaize BP 3, 69360 Solaize, France*

<sup>b</sup>*LJK — Laboratoire Jean Kuntzmann, Université Grenoble Alpes*

<sup>c</sup>*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

---

## 1 Introduction and Scope

2 In the current profound worldwide energy transition, wind power generation is developing rapidly.  
3 As a consequence, wind turbines monitoring, performance optimization and lifetime assessment are  
4 becoming major issues. In the context of digitalization of the industry, the exploitation of collected  
5 data can be optimized by combination with wind turbines numerical models. Such numerical models  
6 can be complex and costly as they involve nonlinear dynamic equations with different physics as  
7 well as stochastic loading from the wind. Moreover, some input parameters of the models can be  
8 poorly or badly known as the structure ages over time and defaults can appear. Consequently,  
9 model predictions are affected by these uncertainties. Characterization and reduction of these  
10 uncertainties is important for decision making [1]. In this context, uncertainty quantification and  
11 reduction methods have been developed [2].

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

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

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

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

21 Firstly, the framework allows to quantifying the sources of uncertainties affecting the fatigue behav-  
22 ior of the structural components of the wind turbine. Let  $g$  be the function mapping the functional  
23 loads of the structure in  $\mathbf{y}$  to the damage quantity of interest (QoI), such as the damage-equivalent  
24 load (DEL) [5]:

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

25 Global Sensitivity Analysis (GSA) methods have been developed to quantify the uncertainty in  
26 QoI, see Eq. (2), w.r.t. the input parameters, their individual contributions, or the contribution of  
27 their interactions. We propose a variance-based GSA methodology, relied on the so-called Sobol’  
28 indices [6], for stochastic computer simulations. Such techniques, which often refer to the proba-  
29 bilistic framework and Monte Carlo (MC) methods, require a lot of calls to the numerical model.  
30 The uncertain input parameters are modeled by independent random variables gathered into a ran-  
31 dom vector and characterized by their probability density function. Variance-based SA for time  
32 consuming deterministic computer models has been mainly performed by approximating the model  
33 by a mathematical function, a.k.a a surrogate model. Among the different surrogate models, we  
34 focus on Gaussian process (GP), a.k.a. Kriging, which is characterized by its mean and covariance  
35 functions. One advantage of the GP model is to provide both a prediction of the numerical model  
36 and the associated uncertainty. However, classical GP model-based GSA cannot tackle the inherent  
37 randomness from stochastic simulation. We propose to model the mean of the QoI  $\mathbb{E}_V[g \circ \mathcal{M}(\mathbf{X}, V)]$   
38 with a GP model with heteroscedastic nugget effect. Then, this surrogate model is used to perform  
39 a sensitivity analysis based on classical MC estimation procedure [7].

40 Secondly, after identification of the less influential input parameters on the fatigue behavior of the  
41 wind turbine, we propose a Bayesian inference framework to carry out a model calibration procedure  
42 based on in situ-measurements. It uses measurements  $\mathbf{y}^{mes}$  to update some prior probability density  
43 functions about the unknown input parameters  $\mathbf{X} \sim p(\mathbf{x})$  and yields some posterior probability

44 density functions, through the Bayes' theorem  $p(\mathbf{x}|\mathbf{y}^{mes}) \propto p(\mathbf{y}^{mes}|\mathbf{x})p(\mathbf{x})$ <sup>1</sup>. Numerous batch  
45 techniques have been developed to solve such Bayesian problems. Nevertheless, recent decades  
46 have been marked by a simultaneous development of sensor technologies and Internet of Things  
47 capabilities. Thus, our research efforts have been directed towards inference techniques where the  
48 data are sequentially processed when new observations become available. In this context, the model  
49 parameter inference can be carried out using sequential Bayesian techniques. In geosciences, these  
50 techniques are called data-assimilation methods. We carry out the calibration using a recursive  
51 Bayesian inference approach based on an Ensemble Kalman Filter (EnKF) [8].

52 However, such problems can be solved assuming that several conditions of well-posedness and identi-  
53 fifiability are achieved. These conditions have been summarized by Hadamard [9]. As highlighted  
54 in [10], a relationship between the non-identifiability of input parameters and the GSA can be estab-  
55 lished. Indeed, input parameters with null total sensitivity indices on the measured outputs imply  
56 their non-identifiability. Therefore, for the purpose of identifiability a second GSA is performed  
57 on the calibration parameters. However, due to the functional behavior of the measurements, we  
58 propose to first reduce their dimensionality through Principal Component Analysis. Then, a GP  
59 is fitted to the different principal components and used to compute an aggregated Sobol' index for  
60 each model parameter [11].

61 Lastly but not least, the proposed framework has been applied to an industrial wind turbine nu-  
62 merical model. The developed recursive inference procedure has shown promising results in the  
63 industrial inversion problem.

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

## 71 1. Kriging based global sensitivity analysis

72 The aim of sensitivity analysis is to quantify the relative influence of input parameters on some  
73 QoI produced from the model outputs of the numerical model. In the context of model calibration,  
74 conducting such an analysis can help to identify which input parameters should be properly esti-  
75 mated. One may distinguish two categories of methods: local and global. While local sensitivity  
76 analysis considers small perturbations of the inputs around some nominal values, global sensitivity  
77 analysis (GSA) varies the inputs on their whole variation range [12]. Among the large number of  
78 available approaches, variance-based sensitivity analysis introduced by [13] proposes to measure the  
79 sensitivity by computing the so-called Sobol' indices. When no analytical formulae of these indices  
80 are available, one way to perform their estimation is to rely on Monte-Carlo (MC) techniques, which  
81 require a huge number of model evaluations. In the context of costly numerical codes as, e.g., the  
82 wind turbine numerical model under interest, the use of a cheap metamodel in place of the true  
83 costly model is thus crucial. In addition to being computationally expensive, the numerical model  
84 we are dealing with is stochastic. This means that from a same set of input parameters, the output  
85 can have different values depending on the wind conditions. This specificity has to be carefully

---

<sup>1</sup>Random variables are written in upper case roman letters and particular realizations of a random variable are written in corresponding lower case letters.

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

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

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

91 We are interested in measuring the sensitivity of a QoI  $g(\mathbf{y}) \in \mathbb{R}$ , see Eq. 2, with respect to the  
 92 input  $\mathbf{x}$ . In the context of GSA, each input is now considered as a random variable  $X_j$  with its un-  
 93 certainty modelled by a probability distribution, such as  $\mathbf{X} = (X_1, \dots, X_p)$ . These one-dimensional  
 94 probability distributions reflect the practitioner's belief in the uncertainty on the parameter values  
 95 and the  $X_j$  are assumed to be independent from each other. Then, the QoI  $g(\mathbf{Y}) := g \circ \mathcal{M}(\mathbf{X}, V)$   
 96 is a random variable itself.

97 The randomness of the QoI has two sources: the randomness from the parameters  $\mathbf{X}$ , and the  
 98 one due to the stochasticity propagated from the model itself through  $V$ , which is assumed to be  
 99 independent of  $\mathbf{X}$ . There exists two approaches to deal with this stochasticity in a GSA framework.  
 100 The first is interested in the full probability distribution of the QoI while the other one is only  
 101 concerned with the QoI averaged over the inherent randomness of the physical system. The latter  
 102 is the one considered in this work. We are therefore interested in the sensitivity of the deterministic  
 103 function  $f$  defined as:

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

104 The total variance of  $Z$  can be split into different parts of variance under the assumption that the  
 105 input parameters are independent (this is the so-called Hoeffding decomposition, see [14]). Each  
 106 part of variance corresponds to the contribution of each set of parameters on the variance of the  
 107 output  $Z$ . By considering the ratio of each part of variance to the total variance of  $Z$ , we obtain a  
 108 measure of importance for each set of input parameters that is called the Sobol' index [13]. More  
 109 precisely, assuming  $Var_{\mathbf{X}}(Z) < +\infty$ ,  $Var_{\mathbf{X}}(Z) \neq 0$ , we define for any  $\mathbf{u} \subset \{1, \dots, p\}$  the closed  
 110 Sobol' index of order  $r = card(\mathbf{u})$  associated to the set of inputs  $\mathbf{X}_{\mathbf{u}} = \{X_j\}_{j \in \mathbf{u}}$  as:

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

111 where  $-\mathbf{u}$  is the complement of  $\mathbf{u}$  in  $\{1, \dots, p\}$ . This index quantifies the main effect of all the  
 112 variables within  $\mathbf{X}_{\mathbf{u}}$ , including interactions, on  $Z$ . The most influential sets of input parameters  
 113 can then be identified as the sets of input parameters with the largest Sobol' indices. Total Sobol'  
 114 indices can also be defined as:

$$S_{\mathbf{u}}^T = 1 - S_{-\mathbf{u}} = 1 - \frac{Var_{\mathbf{X}_{-\mathbf{u}}}(\mathbb{E}_{\mathbf{X}_{\mathbf{u}}}[Z|\mathbf{X}_{-\mathbf{u}}])}{Var_{\mathbf{X}}(Z)} \quad (6)$$

115 This index quantifies the effect of  $\mathbf{X}_{\mathbf{u}}$  plus the effect of all interactions between variables in  $\mathbf{X}_{\mathbf{u}}$  and  
 116 variables in  $\mathbf{X}_{-\mathbf{u}}$  on  $Z$ .

117 A general approach to estimate Sobol' indices is based on Monte-Carlo and requires an important  
 118 number of evaluations of  $f$ . The high computational cost of the wind turbine model prevents from  
 119 performing such estimations under reasonable time constraints. To handle this issue, one may rely  
 120 on a metamodel to compute cheap approximated evaluations of the initial costly computer code. In  
 121 this study, we chose to approximate the true numerical code by a Gaussian process, a.k.a. Kriging  
 122 metamodel, in order to apply a Kriging based sensitivity analysis, e.g., in [15]. We firstly present

123 the noiseless framework and then we detail the case where we are facing to heterogeneously noisy  
 124 evaluations of the function  $f$ .

125 **Ordinary Kriging** - First, a Gaussian process regression model is built to surrogate the function  
 126  $f$ . The principle of Kriging based metamodeling [16, 17] is to consider that our deterministic model  
 127  $f$  can be considered as a realization of a Gaussian process  $\{Z(\mathbf{x}), \mathbf{x} \in \mathcal{P}\}$  with mean function  $\mu$   
 128 and covariance kernel  $\mathbf{C}$ . Then, for any  $\mathbf{x} \in \mathcal{P}$ ,  $f(\mathbf{x})$  is approximated by the conditional Gaussian  
 129 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  
 130  $f$  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  
 131 sample. In the following,  $\mathbf{D}$  is chosen as a Latin Hypercube sample to guarantee a good exploration  
 132 of our numerical model. We then get the ordinary Kriging equations:

$$Z_n(\mathbf{x}) \sim \mathcal{N}(m_{OK}(\mathbf{x}), s_{OK}^2(\mathbf{x})), \quad (7)$$

with

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

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

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

135 **Noisy Kriging** - In our context, exact evaluations of  $f$  can not be obtained directly. We have,  
 136 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 empiri-  
 137 cal mean computed from a  $K$ -sample of  $g \circ \mathcal{M}(\mathbf{d}_i, V)$  and  $\varepsilon_i$  is a centered noise whose variance  
 138 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  
 139 approximation, that the vector  $(\varepsilon_1, \dots, \varepsilon_n)$  is a centered Gaussian random vector with diagonal co-  
 140 variance matrix  $\text{diag}(\tau_1^2, \dots, \tau_n^2)$ . Then, provided that the process  $Z$  and the Gaussian measurement  
 141 errors  $\varepsilon_i$  are stochastically independent, the process  $Z$  is still Gaussian conditionally on the noisy  
 142 observations  $\tilde{z}_i$  and its conditional mean and variance functions are given by the following slightly  
 143 modified kriging equations:

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

with

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

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

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

$$\tilde{S}_{\mathbf{u}} = \frac{\text{Var}_{\mathbf{X}_{\mathbf{u}}}(\mathbb{E}_{\mathbf{X}_{-\mathbf{u}}}[\tilde{Z}_n(\mathbf{X}) | \mathbf{X}_{\mathbf{u}}])}{\text{Var}_{\mathbf{X}}(\tilde{Z}_n(\mathbf{X}))}. \quad (13)$$

146 As  $\tilde{Z}_n$  is a random process, the resulting indices are also random. These indices are estimated via  
 147 Monte-Carlo samples from two designs of experiments, using a pick-freeze procedure. A design is  
 148 a point set  $\mathbf{P} = \{\mathbf{x}_i\}_{i=1}^s$  in which each point is obtained by sampling  $s$  times each input variable  
 149  $X_j \in \mathcal{P}$ ,  $j = 1, \dots, p$ . Each row of the design is a point  $\mathbf{x}_i$  in  $\mathcal{P}$ , the  $j$ -th column of the design  
 150 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  $\mathbf{x}'$ ,  
 151 the hybrid point  $(\mathbf{x}_{\mathbf{u}} : \mathbf{x}'_{-\mathbf{u}}) \in \mathcal{P}$  is defined as  $x_j$  if  $j \in \mathbf{u}$  and  $x'_j$  if  $j \notin \mathbf{u}$ . Consider  $\mathbf{P} = \{\mathbf{x}_i\}_{i=1}^s$

152 and  $\mathbf{P}' = \{\mathbf{x}'_i\}_{i=1}^s$  two designs sampled from the distribution of the input random vector  $\mathbf{X}$ . One  
 153 way to estimate the quantity in (13) is as follows:

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

154 Confidence intervals can be obtained via a bootstrap method, as described in Algorithm 1 of [15].  
 155 These intervals integrate two sources of uncertainty, the first one is related to the metamodel  
 156 approximation, and the second one is related to the Monte-Carlo integration.  
 157 Each step of the procedure was implemented in  $\mathbf{R}$  using the *km* and *sobolGP* functions from  
 158 respectively the *DiceKriging* and *Sensitivity* packages, see [19, 20].

## 159 2. Bayesian inference for online parameter identification

160 In our context, we suppose that data are collected sequentially and we seek to refine our choice of  
 161 parameters in the numerical model at each iteration. The problematic can be seen as a supervised  
 162 learning problem that we aim to solve sequentially as each pair of data points  $\{v_{(k)}, \mathbf{y}_{(k)}\}$  is obtained  
 163 at the iteration  $k$  [21].

164 In the recursive Bayesian parameter estimation framework, developed in this paper, the unknown  
 165 time-invariant input parameter vector  $\mathbf{x}$  is modeled as a discrete Markov chain, the evolution of  
 166 which is governed by a random walk process. In our context, the dynamic evolution of the input  
 167 parameter and the measurement modelisation with the simulator responses can be formulated at  
 168 iteration  $\{k+1\}_{k=0\dots T-1}$  as:

$$\begin{cases} \mathbf{x}_{(k+1)} &= \mathbf{x}_{(k)} + \delta_{(k)} \\ \mathbf{y}_{(k+1)} &= \mathcal{M}(\mathbf{x}_{(k+1)}, V = v_{(k+1)}) + \boldsymbol{\epsilon}_{(k+1)} \end{cases} \quad (15)$$

169 where  $\mathbf{x}_{(k+1)}$  is the input parameter vector,  $v_{(k+1)}$  is a known realization of the stochastic external  
 170 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  
 171 respectively an artificial dynamic noise and a combination of the model and observation errors. For  
 172 the sake of readability,  $\mathbf{y}$  will represent the vector gathering the measured responses obtained on  
 173 the structure of interest.

174 Filtering techniques, a type of data assimilation, can be used to sequentially estimate the parameter  
 175 vector in Eq. (15) using the known input solicitation and the available measurements. Among all  
 176 available filtering methods, the Kalman Filter (KF) [22] has been widely applied when dealing with  
 177 a linear system with Gaussian error sources. In this paper due to the non-linearity in our numerical  
 178 model, the Ensemble Kalman filter (EnKF) [8] is used to perform parameters estimation. The  
 179 EnKF is a sequential Monte Carlo method that provides an alternative to the traditional KF. The  
 180 method works on an ensemble of parameter estimates transforming them from the prior distribution  
 181 into the posterior one. We propose to use an efficient stratified sampling technique instead of a  
 182 conventional Monte Carlo method to generate the initial ensemble of parameter estimates. The  
 183 developed technique is based on a Latin Hypercube Sampling (LHS) coupled with a geometrical  
 184 criteria maximizing the minimum distance between the design points, see [23].

185 In the field of inverse problems, this inference method is referred to as Ensemble Kalman Inversion  
 186 (EKI). The EnKF formulation used in [24] is adopted in this paper. At iteration  $k$ , we have an  
 187 ensemble of size  $N_{ens}$ , of forecast parameter estimates  $\mathbf{x}_{(k)}^f = [\mathbf{x}_{(k)}^{f(1)}, \dots, \mathbf{x}_{(k)}^{f(N_{ens})}] \in \mathbb{R}^{p \times N_{ens}}$  where  
 188 the superscript  $\cdot^{f(i)}$  denotes the  $i$ -th forecast member of the ensemble. The mean of the forecast

189 members of the ensemble is given by:

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

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

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

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

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

193 where the observation operator, denoted by  $\mathbf{M} \in \mathbb{R}^{m \times N_{ens}}$ , is linear or has been linearized from the  
194 function  $\mathcal{M}$ , see [26].

195 Nevertheless, for most applications this condition of linear (or linearized) observation operator  
196 cannot be applied. In that context, as proposed in [27], we can replace the terms  $\mathbf{P}_{(k)}^f \mathbf{M}^T$  and  
197  $\mathbf{M} \mathbf{P}_{(k)}^f \mathbf{M}^T$  of the Kalman Gain equation by the following ones:

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

198 and,

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

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

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

199 The difference between the deterministic EnKF and the stochastic version presented hereafter is  
200 that the observations are now treated as random variables. Indeed, an ensemble of observations of  
201 the same size  $N_{ens}$  is generated by adding noise terms to the observation set  $\mathbf{y}_{(k)}$  such that:

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

202 The noise terms are sampled from the distribution of the error covariance matrix  $\mathbf{R}_{(k)}$ . The stochas-  
203 tic EnKF has been showed to have the advantage to “re-Gaussianize” the ensemble distribution  
204 thanks to the observation perturbations [29]. Maintaining Gaussianity has a positive impact on  
205 analysis quality of the ensemble filter by maintaining the correct forecast error covariance. Most



206 of the time, the measurement observational error covariance matrix is diagonal according to the  
 207 assumption of independent observations. Using the presented approximation, the computation of  
 208 the Kalman gain  $\mathbf{K}_{(k)}$  can be done. We can independently update the ensemble members using:

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

209 where the superscript  $\cdot^{a(i)}$  denotes the  $i$ -th updated member of the ensemble. The last step of the  
 210 EnKF method is the forecast step of the ensemble parameters at  $k + 1$  and involves an ensemble of  
 211  $N_{ens}$  updated parameters for iteration  $k$ , see Eq. (24). The presented method is fully described in  
 212 Algo. 1.

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

**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.

### 213 3. Description of the wind-turbine numerical model

214 Dynamic analysis of wind turbines involves strong interactions between the turbines' aerodynamics,  
 215 the control system and the structural mechanics. The main solicitations are the environmental

216 conditions and the rotating machinery during operating term. In order to model and simulate the  
217 nonlinear response of wind turbine structures under such solicitations, various servo-aero-elastic  
218 software have been developed, such as OpenFAST [30], Bladed [31], HAWCK2 [32] or Deeplines  
219 Wind [33].

220 In our study, a simulator of a Senvion MM82 wind turbine has been developed using Deeplines  
221 Wind software from technical specifications. This software is a fully coupled simulation tool taking  
222 into account the aerodynamics of the aero-generator, the elasticity of the structural wind turbine  
223 components (mast, blades and drive-train systems) and the control system [34]. The software archi-  
224 tecture developed by IFPEN<sup>2</sup> and Principia<sup>3</sup> is fully modular with different dynamic libraries (DLL)  
225 called by the solver. The integration in time is performed with an implicit Newmark integration  
226 scheme. The developed simulator includes a nonlinear beam finite element formulation to model  
227 the structural components. The aerodynamic loads acting on the turbine rotor are dynamically  
228 computed by employing the blade element momentum (BEM) theory for Horizontal Axis Wind  
229 Turbine (HAWT). A Deeplines Wind software validation, based on code comparisons [3], has shown  
230 accurate results in various conditions.

231 Wind turbine simulation consists of two stages: first the generation of the input turbulent wind  
232 field and then the fully coupled servo-aero-elastic simulation. The generation of the input stochastic  
233 process is done by using a simulator called Turbsim [4]. This simulator has some deterministic inputs  
234 such as the turbulence intensity, the mean wind speed, the mean flow angles, the spectrum and  
235 the spatial correlation model. In our model, we have used an IEC<sup>4</sup> Kaimal turbulence spectrum  
236 with an exponential spatial coherence. Nevertheless, these deterministic values cannot uniquely  
237 determine a stochastic wind field and a pseudo-random number generator have to be used in order  
238 to create random phases for the velocity time. Then structural responses are time computed with  
239 a multi-physics numerical code such as Deeplines Wind. Following the formalism introduced in  
240 Eq. (2), the model for such stochastic simulation can be formulated as follows:

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

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

244 A wind turbine structure can encounter a variety of operating conditions. Each of the operating  
245 conditions, modelled by the stochastic process  $V$ , is parameterized by measurable deterministic  
246 quantities mentioned previously. In this paper we will perform the study at an under-rated average  
247 wind speed of 8 m/s corresponding to the most common operating regime of the considered turbine.  
248 All computed responses are based on 10-min *effective* simulations of the MM82 Senvion wind  
249 turbine. By *effective*, we mean that the transient start-up behavior is removed from the analysis.  
250 The transient start-up behavior can be decomposed in a ramp time wind and an oversight periods.  
251 The oversight period has been set according to auto-correlation studies of the outputs. This period  
252 permits to remove the effect of the ramp time period on the numerical model responses. Each  
253 numerical simulation lasts between 15 and 30 minutes according to the CPU performance.

254 From the structural time responses computed by the Deeplines Wind model, we obtain some QoIs  
255 describing the fatigue behavior of the wind turbine. They are obtained by post-treating the resulting  
256 time series of internal loads at different locations of the analysed design, see Fig. 1 and Tab. 1. We

---

<sup>2</sup>see <https://www.ifpenergiesnouvelles.fr/>

<sup>3</sup>see <http://www.principia-group.com/>

<sup>4</sup>International Electrotechnical Commission

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

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

259 Each fatigue QoI has been estimated by using the damage-equivalent load (DEL). The DEL is  
 260 computed for a set of parameter values and different realizations of the stochastic process  $V$ . It  
 261 is defined as the regular load amplitude that would create in  $N_{ref}$  cycles the same fatigue as the  
 262 considered irregular load history. The DEL is computed based on the Palmgren Miner's rule [35]:

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

263 where  $i = 1, \dots, N_c$  corresponds to each range bin,  $S_i$  is the cycle range value and  $N_i$  is the number  
 264 of cycles for the  $i$ -th bin. The exponent  $m$  is the negative inverse slope of the cyclic stress against  
 265 the cycles to failure curve (S-N curve) and  $N_{ref}$  is the reference number of cycles usually set to an  
 266 arbitrary value. The cycles in an irregular load history are computed using the Rainflow counting  
 267 method [36].

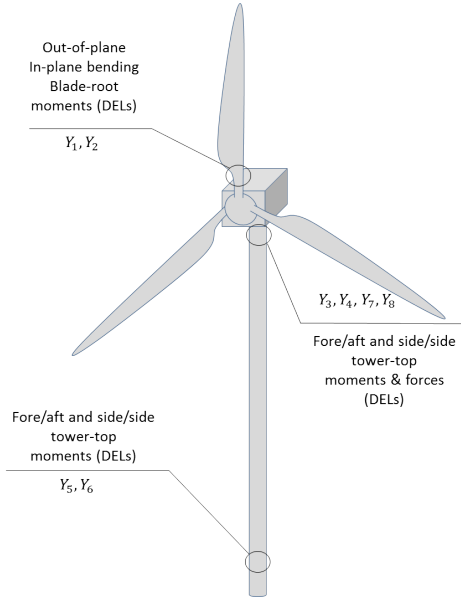


Figure 1: Recorded time series of loads at different locations of the wind turbine

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.

268 In order to ensure that the variation of the input parameters is distinguishable from the realization  
 269 of the stochastic process  $V$ , multiple wind realizations have to be generated. A convergence study  
 270 has been led in order to determine the number of realizations needed to encompass the variation of  
 271 the selected realization of the stochastic process. The QoIs analyzed in this study are the DEL at  
 272 different location of the wind turbine, see Eq. (27). The number of stochastic process realizations  
 273 used during the convergence study varies from 1 to 30.

274 Fig. 2 shows the convergence of the tower bottom fore-aft bending moment DEL on the number  
 275 of realizations used for their averaging estimation. According to certification guidelines, see design  
 276 load case 1.2 in [37], the fatigue analysis has to be led with 6 wind realizations of 10 min time period.  
 277 Nevertheless, as we can underline with the last mentioned figure, the empirical mean computed from  
 278 these 6 realizations does not seem to be a reliable estimator. In other words, 6 simulations are not

279 sufficient for QoIs' statistics to converge. With our industrial numerical model, a compromise has  
 280 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.

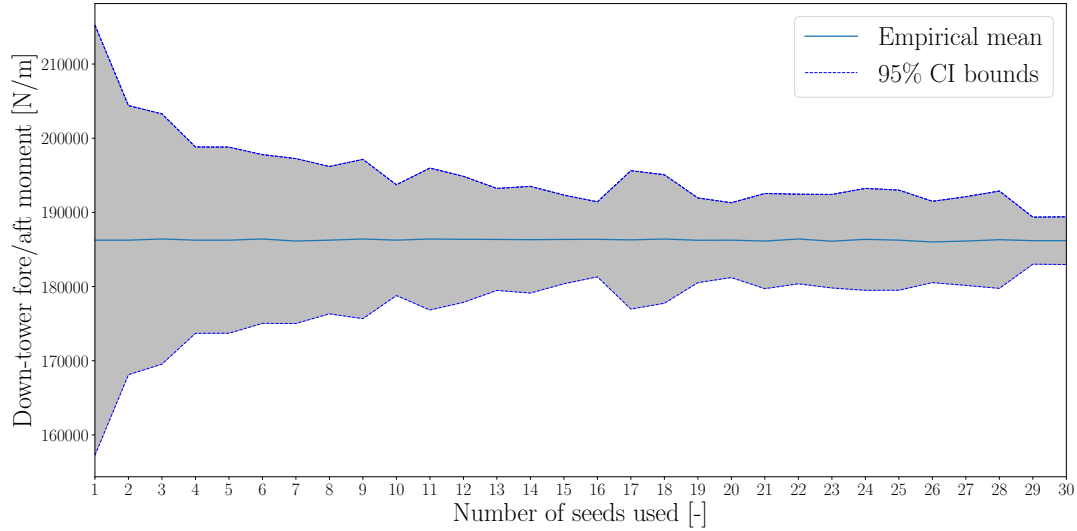


Figure 2: Convergence of the tower bottom fore-aft bending moment DEL as a function of the number of turbulent seeds used for its evaluation. 95% confidence interval around the estimated empirical mean is also represented (grey area).

281 The aim of this paper is to identify and reduce the structural sources of uncertainty on the fatigue  
 282 QoIs. We will focus our study on some wind turbine properties represented by 13 parameters  
 283 gathered in the vector  $\mathbf{x}$ . A literature review has been done to specify the uncertainty in the  
 284 parameter values. All these parameters were considered independent of one another with Gaussian  
 285 distributions.  
 286

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

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

300 For the aerodynamic properties, we have considered the wind turbine yaw misalignment by changing  
 301 the yaw angle of the turbine. For the individual blade pitch error, a constant offset angle is applied  
 302 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	$\mu$	$\sigma$	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 - $\beta_{TR}$ [-]	3.10e-02	9.93e-03	[40]
Inertial Nacelle - $I_{zz}$ [ $kg \cdot m^2$ ]	7.00e+05	2.33e+04	IFPEN $\pm 10\%$ $\mu$
Drive-train Torsional stiffness - $K_D$ [ $\frac{N \cdot m^2}{rad}$ ]	4.45e+09	1.48e+08	[41]
Blade Flap wise Stiffness - $\alpha_{BF}$ [ $N \cdot m^2$ ]	1.00	3.33e-02	IFPEN $\sim \pm 10\%$ 1 FS
Blade Edge wise Stiffness - $\alpha_{BE}$ [ $N \cdot m^2$ ]	1.00	3.33e-02	IFPEN $\sim \pm 10\%$ 1 ES
Blade mass coefficient - $\alpha_{mass}$ [%]	1.00	1.67e-02	[38]
Blade Rayleigh Damping - $\beta_{BR}$ [-]	5.39e-03	1.45e-03	[39]
Blade mass imbalance - $\eta_B$ [%]	2.50	8.33e-01	[39]
Yaw misalignment - $\omega$ [ $^\circ$ ]	0	6.67	[42]
Individual pitch error - $\Omega$ [ $^\circ$ ]	0.10	3.33e-02	[43]

303 After an appropriate sensitivity analysis leading to the identification of the less influential input  
 304 parameters on the fatigue QoIs, we can fix their value to a nominal one without affecting the fatigue  
 305 behavior of the structure. Then, the uncertainties of the other parameters has to be reduced by  
 306 employing an appropriate inference method based on in-situ measurements. In this context, let  
 307 us consider a wind turbine instrumented with accelerometers. We assume that bi-axial measuring  
 308 devices are located at mid and top tower height position. From these sensors, we can record four  
 309 functional acceleration time series. Then, the power spectral density (PSD) of each measured  
 310 acceleration time series is computed using Welch's method [44], see Fig. 3.

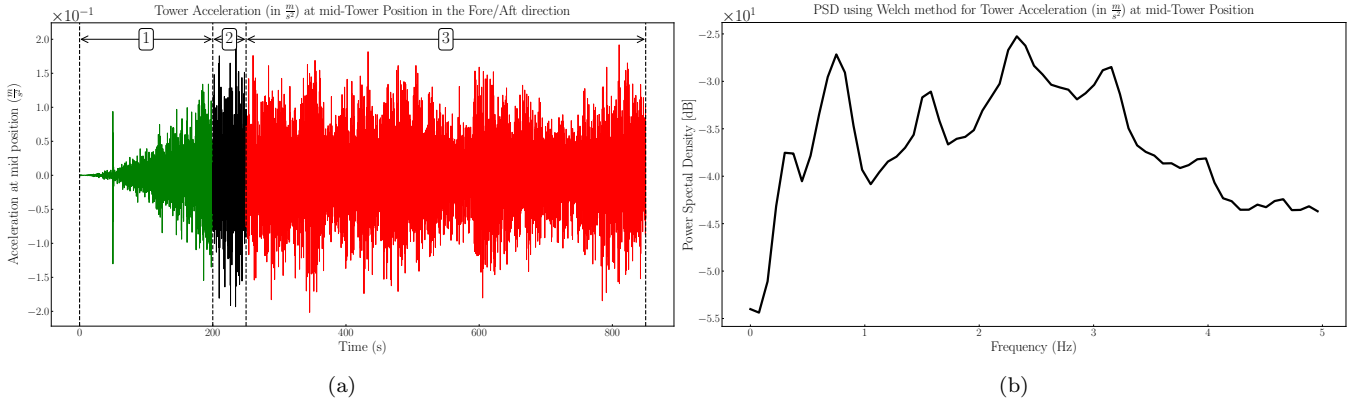


Figure 3: On the left side (a): simulated acceleration in  $\frac{m}{s^2}$  of the wind turbine tower in the fore-aft direction obtained at the accelerometer device located at mid-tower decomposed in a ramp time wind [1], an oversight period [2] and a dynamical period of interest [3]. On the right side (b): estimated PSD of the period of interest using Welch's method [44].

#### 311 4. Illustration of the proposed framework on the wind-turbine numerical model

312 **GSA of the fatigue QoIs** - The Deeplines Wind numerical model presented in Section 3 is used  
 313 with 13 uncertain input parameters, each one having its associated probability distribution, see  
 314 Tab.2. The total Sobol' indices associated to these parameters for each DEL have been estimated  
 315 using the heteroscedastic noisy GP model-based Sobol' index procedure as described in Section 1. A

316 Latin Hypercube Sampling (LHS) of size 500 has been used to emulate the numerical model. Then,  
 317 an augmented LHS of size 150 has been generated to determine the accuracy of the surrogate models,  
 318 [45]. To apply this approach 6,500 forward wind turbine numerical simulations were submitted on  
 319 the 206 TFlops supercomputer of IFPEN.

320 The function *sobolGP* performs a kriging-based GSA by taking into account the complete condi-  
 321 tional predictive distribution of the surrogate model. The function estimates total Sobol' indices  
 322 thanks to the Jansen estimators, see [46]. Jansen Sobol' index estimators are accurate for large and  
 323 small total indices. Moreover, by taking into account the complete conditional predictive distribu-  
 324 tion of the trained surrogate model in the estimation procedure of the total Sobol' indices, we can  
 325 evaluate the uncertainty due to Monte-Carlo estimation, but also due to meta-modeling [15].

326 The results for the total Sobol' indices with their corresponding 95% confidence intervals are sum-  
 327 marized in Fig. 4. A threshold of  $2.5e-2$  was chosen to display a separation between input param-  
 328 eters with high and low total Sobol' indices. Fig. 5 represents the different sources of uncertainty for  
 329 the estimation of total Sobol' indices, obtained thanks to the function *sobolGP* for the DEL of the  
 330 out-of-plane bending blade-root moment.

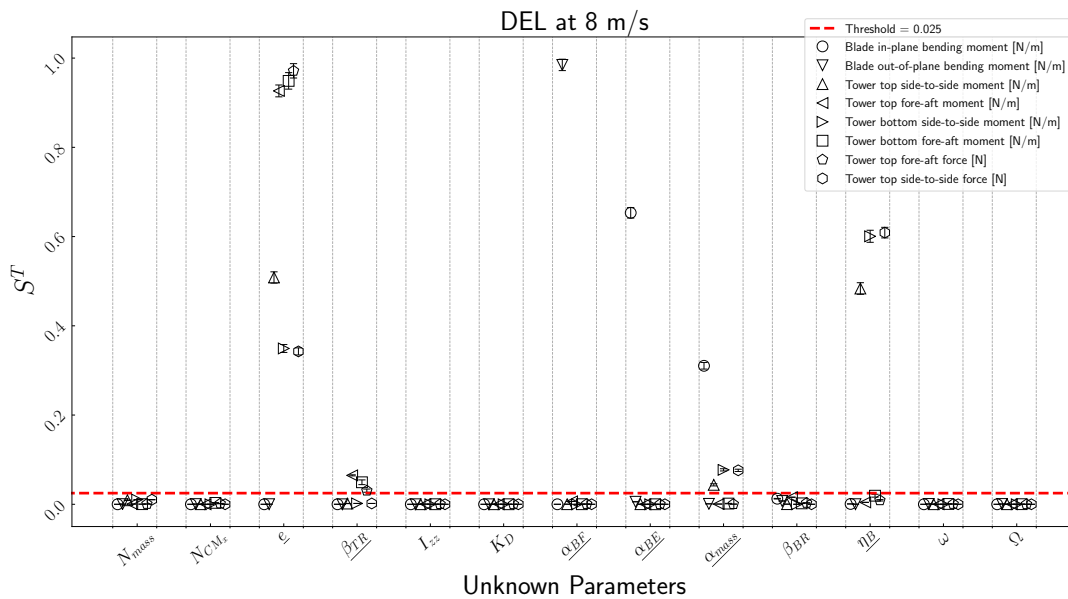


Figure 4: Estimation of total Sobol' indices (y-axis) with their 95% confidence interval corresponding to each of the 13 parameters (x-axis) for the different quantities of interest in fatigue. The dashed line corresponds to a threshold arbitrarily chosen to  $2.5e-2$ . Confidence intervals (CI) are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

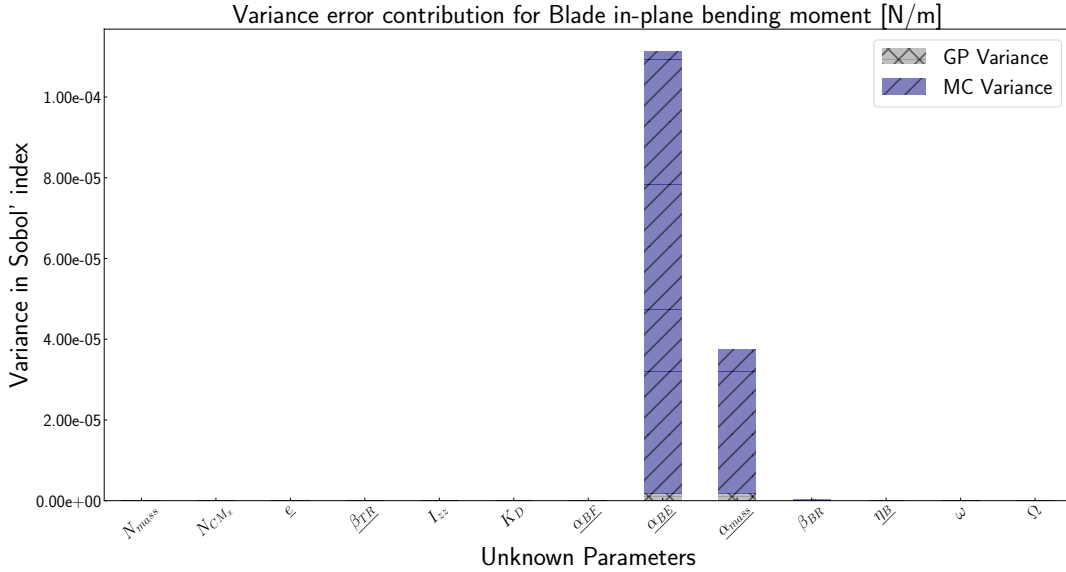


Figure 5: Splitting of the variance of total Sobol’ index estimators (y-axis) corresponding to each of the 13 input parameters (x-axis) for the DEL of the out-of-plane bending blade-root moment. The number of samples for the conditional Gaussian process, in order to quantify the uncertainty of the Gaussian process approximation, was set to 100. The uncertainty due to Monte-Carlo integration was computed with a bootstrap procedure of 100.

331 By considering all the total Sobol’ indices obtained for the different fatigue QoIs presented in Fig. 4,  
 332 we can notice that only 6 parameters have indices values greater than the threshold. Consequently,  
 333 we can fix other parameters to any specific value in the range of variability without affecting the  
 334 fatigue QoIs. During the recursive Bayesian estimation, these non-influential parameters will be  
 335 arbitrarily set at their mean value described in Tab. 2. By employing this method, we will reduce  
 336 the parameter dimension in the inference procedure without affecting the fatigue QoIs which are  
 337 crucial for life of wind turbines.

338 **Identifiability study** - It is possible that the considered experimental measurement settings  
 339 do not offer enough information for the identification of some input parameters. In this context,  
 340 another interesting property of the GSA underlined in Proposition 4.2 in [10] is that nullity of the  
 341 total sensitivity index for a specific input parameter implies its non-identifiability from the output  
 342 under consideration. Thus, a GSA led on the measured outputs will determine which parameters  
 343 cannot be inferred, although it is not a sufficient condition for identifiability.

344 In our industrial application study, the measured outputs, obtained from the accelerometers, are  
 345 expressed as discretized time series. We are interested in their response in the frequency-domain by  
 346 using the power spectral density (PSD). Discretized PSD series involve a substantial dimensionality  
 347 and a high degree of redundancy. To bypass this issue, we have firstly focused our study on an  
 348 orthogonal decomposition, of the different discretized PSDs, in order to reduce their dimensionality.  
 349 This orthogonal decomposition will be performed by a data-driven method called Principal Component  
 350 Analysis (PCA) [47]. PCA allows the functional output expansion in a new reduced space  
 351 spanned by the most significant directions in term of variance of the discretized functional output.  
 352 In our study, a method based on PCA and GSA with GP model is used to compute an aggregated  
 353 Sobol’ index for each input parameter of the model. As described in [11], the proposed index  
 354 synthesizes the influence of the parameter on the whole time series output.

355 In order to ensure that the variation of the input parameters is distinguishable from the realization  
 356 of the stochastic process  $V$ , 10 wind realizations have been used in this GSA. A new LHS of size  
 357 300 with a geometrical criteria maximizing the minimum distance between the design points has

358 been used to emulate the numerical model. In Tab. 3, we summarize the total and total aggregated  
 359 Sobol’ indices obtained with the GP model built on the train set from the lastly mentioned stratified  
 360 sampling. In this analysis, parameters with total Sobol’ index values under a threshold set at 1e-02  
 361 will be considered as non-identifiable from the measured output. We can conclude that none of the  
 362 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	$e$ [%]	$\beta_{TR}$ [-]	$\alpha_{BF}$ [%]	$\alpha_{BE}$ [%]	$\alpha_{mass}$ [%]	$\eta_B$ [%]
Tower middle fore-aft acceleration’s PSD	<u>2.44e-01</u>	<u>7.64e-01</u>	6.07e-03	1.37e-04	3.59e-04	4.46e-03
Tower middle side-to-side acceleration’s PSD	<u>3.84e-01</u>	<u>3.95e-01</u>	<u>1.38e-01</u>	6.73e-04	<u>8.60e-02</u>	7.17e-03
Tower top fore-aft acceleration’s PSD	<u>1.21e-01</u>	<u>6.70e-01</u>	2.09e-03	<u>5.91e-02</u>	<u>6.76e-02</u>	<u>1.05e-01</u>
Tower top side to side acceleration’s PSD	<u>6.56e-02</u>	<u>6.24e-01</u>	1.36e-03	<u>9.67e-02</u>	<u>1.39e-01</u>	<u>9.52e-02</u>

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

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

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

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

388 For the sake of numerical time consumption of this test case, we use a number of inference iterations  
 389  $T$  equals to 20 and a number of members of the ensemble set at 100. Fig. 6 shows the results of the  
 390 identification. It can be observed that the considered input parameters are well recovered. Tab. 4  
 391 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.

	$e$ [%]	$\beta_{TR}$ [-]	$\alpha_{BF}$ [%]	$\alpha_{BE}$ [%]	$\alpha_{mass}$ [%]	$\eta_B$ [%]
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

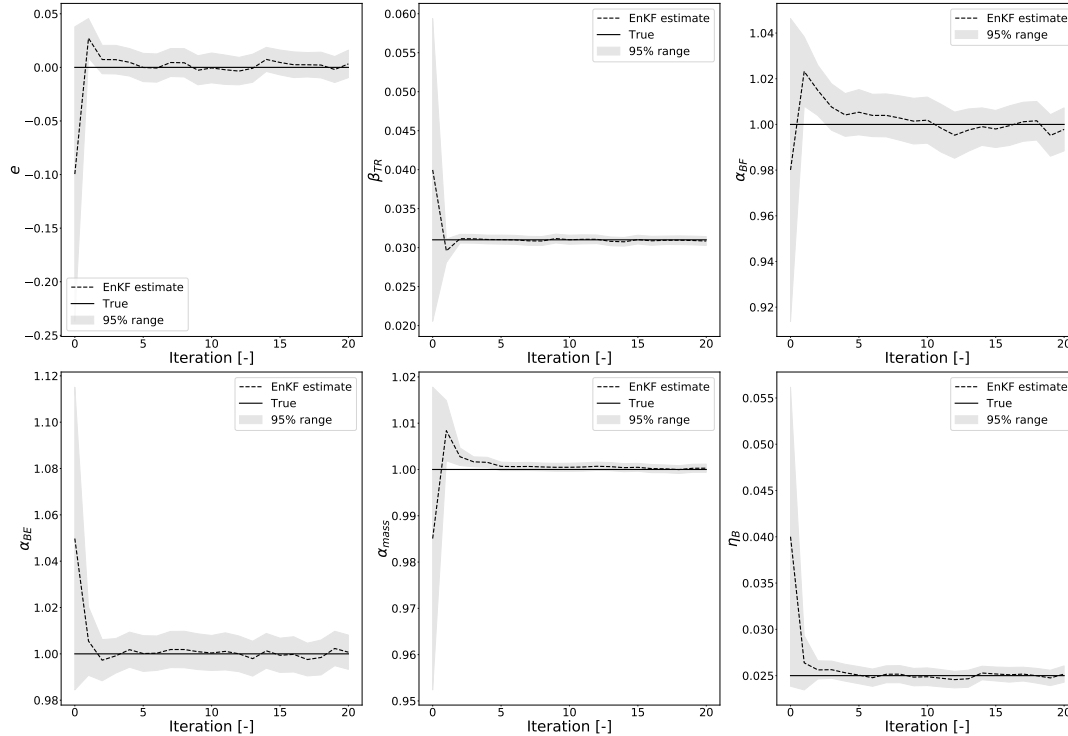


Figure 6: Iteration evolution of the a posteriori estimates of the input parameters. Results obtained by running EnKF presented in Section 2 with  $N = 100$  members of the ensemble used for the estimation and considering pseudo-experimental numerical observations.

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

## 401 Conclusion

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

404 The contributions of this paper are twofold. First, we have proposed a global sensitivity analysis  
405 based on Sobol' indices using a Gaussian process model with heteroscedastic nugget effect to quantify

406 uncertainties of a stochastic and time-consuming wind turbine numerical model. The procedure we  
407 present is efficient to balance the inherent uncertainty of the stochastic numerical model and the one  
408 from the input parameters. More precisely the GSA has been performed on the fatigue quantities  
409 of interest and showed that only a restricted number of unknown parameters happens to influence  
410 these responses. Since fatigue quantities of interest are a crucial wind turbine design and life criteria,  
411 these influent input parameters have to be properly estimated in order to give confidence in fatigue  
412 analysis results.

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

## 423 References

- 424 [1] E. De Rocquigny, N. Devictor, and S. Tarantola, *Uncertainty in industrial practice: a guide to*  
425 *quantitative uncertainty management*. John Wiley & Sons, 2008.
- 426 [2] R. C. Smith, *Uncertainty quantification: theory, implementation, and applications*, vol. 12.  
427 Siam, 2013.
- 428 [3] T. Perdrizet, J.-C. Gilloteaux, D. Teixeira, G. Ferrer, L. Piriou, D. Cadiou, J.-M. Heurtier, and  
429 C. Le Cunff, “Fully coupled floating wind turbine simulator based on nonlinear finite element  
430 method: Part ii validation results,” in *ASME 2013 32nd International Conference on Ocean,*  
431 *Offshore and Arctic Engineering*, pp. V008T09A052–V008T09A052, Citeseer, 2013.
- 432 [4] B. J. Jonkman, “Turbsim user’s guide: Version 1.50,” tech. rep., National Renewable Energy  
433 Lab (NREL), Golden, CO, USA, 2009.
- 434 [5] G. Freebury and W. Musial, “Determining equivalent damage loading for full-scale wind turbine  
435 blade fatigue tests,” in *2000 ASME Wind Energy Symposium*, p. 50, 01 2000.
- 436 [6] I. M. Sobol’, “Sensitivity estimates for nonlinear mathematical models,” *Mathematical mod-*  
437 *elling and computational experiments*, vol. 1, no. 4, pp. 407–414, 1993.
- 438 [7] L. Le Gratiet, S. Marelli, and B. Sudret, “Metamodel-Based Sensitivity Analysis: Polynomial  
439 Chaos Expansions and Gaussian Processes,” in *Handbook of Uncertainty Quantification*, pp. 1–  
440 37, Springer International Publishing, 2017.
- 441 [8] G. Evensen, *Data assimilation: the ensemble Kalman filter*. Springer Science & Business Media,  
442 2009.
- 443 [9] Hadamard, Jacques, “Sur les problèmes aux dérivées partielles et leur signification physique,”  
444 *Princeton university bulletin*, pp. 49–52, 1902.

- 445 [10] S. Dobre, T. Bastogne, C. Profeta, M. Barberi-Heyob, and A. Richard, “Limits of variance-  
446 based sensitivity analysis for non-identifiability testing in high dimensional dynamic models,”  
447 *Automatica*, vol. 48, no. 11, pp. 2740–2749, 2012.
- 448 [11] M. Lamboni, H. Monod, and D. Makowski, “Multivariate sensitivity analysis to measure global  
449 contribution of input factors in dynamic models,” *Reliability Engineering and System Safety*,  
450 vol. 96, no. 4, pp. 450–459, 2011.
- 451 [12] A. Saltelli, K. Chan, and E. M. Scott, eds., *Sensitivity analysis*. Wiley series in probability  
452 and statistics, New York and Chichester and Weinheim etc.: J. Wiley & sons, 2000.
- 453 [13] I. M. Sobol’, “On sensitivity estimation for nonlinear mathematical models,” *Matematicheskoe*  
454 *modelirovanie*, vol. 2, no. 1, pp. 112–118, 1990.
- 455 [14] W. Hoeffding, *A class of statistics with asymptotically normal distribution*. Springer, 1992.
- 456 [15] L. Le Gratiet, C. Cannamela, and B. Iooss, “A bayesian approach for global sensitivity analysis  
457 of (multifidelity) computer codes,” *SIAM/ASA Journal on Uncertainty Quantification*, vol. 2,  
458 no. 1, pp. 336–363, 2014.
- 459 [16] D. G. Krige, M. Guarascio, and F. A. Camisani-Calzolari, “Early South African geostatistical  
460 techniques in today’s perspective,” *Geostatistics*, vol. 1, pp. 1–19, 1989.
- 461 [17] C. E. Rasmussen, “Gaussian processes in machine learning,” in *Summer School on Machine*  
462 *Learning*, pp. 63–71, Springer, 2003.
- 463 [18] A. Marrel, B. Iooss, B. Laurent, and O. Roustant, “Calculations of sobol indices for the gaussian  
464 process metamodel,” *Reliability Engineering & System Safety*, vol. 94, no. 3, pp. 742 – 751,  
465 2009.
- 466 [19] O. Roustant, D. Ginsbourger, and Y. Deville, “DiceKriging, DiceOptim: Two R packages  
467 for the analysis of computer experiments by kriging-based metamodeling and optimization,”  
468 *Journal of Statistical Software*, vol. 51, no. 1, pp. 1–55, 2012.
- 469 [20] B. Iooss, A. Janon, G. Pujol, with contributions from Baptiste Broto, K. Boumhaout, S. D.  
470 Veiga, T. Delage, J. Fruth, L. Gilquin, J. Guillaume, L. Le Gratiet, P. Lemaitre, B. L. Nelson,  
471 F. Monari, R. Oomen, O. Rakovec, B. Ramos, O. Roustant, E. Song, J. Staum, R. Sueur,  
472 T. Touati, and F. Weber, *sensitivity: Global Sensitivity Analysis of Model Outputs*, 2019. R  
473 package version 1.16.0.
- 474 [21] N. B. Kovachki and A. M. Stuart, “Ensemble kalman inversion: a derivative-free technique for  
475 machine learning tasks,” *Inverse Problems*, vol. 35, no. 9, p. 095005, 2019.
- 476 [22] R. E. Kalman *et al.*, “Contributions to the theory of optimal control,” *Boletin de la Sociedad*  
477 *Matematica Mexicana*, vol. 5, no. 2, pp. 102–119, 1960.
- 478 [23] M. E. Johnson, L. M. Moore, and D. Ylvisaker, “Minimax and maximin distance designs,”  
479 *Journal of Statistical Planning and Inference*, vol. 26, no. 2, pp. 131 – 148, 1990.
- 480 [24] C. Snyder and F. Zhang, “Assimilation of simulated doppler radar observations with an en-  
481 semble kalman filter.,” *Monthly Weather Review*, vol. 131, no. 8, 2003.

- 482 [25] G. Welch and G. Bishop, “An introduction to the kalman filter,” tech. rep., USA, 1995.
- 483 [26] R. E. Kopp and R. J. Orford, “Linear regression applied to system identification for adaptive  
484 control systems,” *Aiaa Journal*, vol. 1, no. 10, pp. 2300–2306, 1963.
- 485 [27] P. L. Houtekamer and H. L. Mitchell, “A sequential ensemble kalman filter for atmospheric  
486 data assimilation,” *Monthly Weather Review*, vol. 129, no. 1, pp. 123–137, 2001.
- 487 [28] J. T. Ambadan and Y. Tang, “Sigma-point kalman filter data assimilation methods for strongly  
488 nonlinear systems,” *Journal of the Atmospheric Sciences*, vol. 66, no. 2, pp. 261–285, 2009.
- 489 [29] G. W. Lawson and J. A. Hansen, “Implications of stochastic and deterministic filters as  
490 ensemble-based data assimilation methods in varying regimes of error growth,” *Monthly weather  
491 review*, vol. 132, no. 8, pp. 1966–1981, 2004.
- 492 [30] The National Renewable Energy Laboratory, “Openfast.” [http://openfast.readthedocs.  
493 io](http://openfast.readthedocs.io), 2018.
- 494 [31] DNV GL, “Bladed: Wind turbine design software,” 2013.
- 495 [32] T. J. Larsen and A. M. Hansen, *How 2 HAWC2, the user’s manual*, 2007.
- 496 [33] Principia, “Principia 2019 deeplines wind.”
- 497 [34] C. Le Cunff, J.-M. Heurtier, L. Piriou, C. Berhault, T. Perdrizet, D. Teixeira, G. Ferrer,  
498 and J.-C. Gilloteaux, “Fully coupled floating wind turbine simulator based on nonlinear finite  
499 element method: Part i methodology,” in *ASME 2013 32nd International Conference on Ocean,  
500 Offshore and Arctic Engineering*, pp. V008T09A050–V008T09A050, Citeseer, 2013.
- 501 [35] H. J. Sutherland, “On the fatigue analysis of wind turbines,” tech. rep., Sandia National Labs.,  
502 Albuquerque, NM (US); Sandia National Labs , 1999.
- 503 [36] N. Cosack, *Fatigue load monitoring with standard wind turbine signals*. PhD thesis, 01 2011.
- 504 [37] “Loads and site conditions for wind turbines,” standard, DNV GL, November 2016.
- 505 [38] D. Witcher, “Uncertainty Quantification Techniques in Wind Turbine,” 2017.
- 506 [39] A. N. Robertson, K. Shaler, L. Sethuraman, and J. Jonkman, “Sensitivity of uncertainty in  
507 wind characteristics and wind turbine properties on wind turbine extreme and fatigue loads,”  
508 *Wind Energy Science Discussions*, pp. 1–41, 2019.
- 509 [40] C. Koukoura, *Validated Loads Prediction Models for Offshore Wind Turbines for Enhanced  
510 Component Reliability*. PhD thesis, Technical University of Denmark, 2014.
- 511 [41] J. Holierhoek, H. Korterink, R. van de Pieterman, L. Rademakers, and D. Lekou, “Recom-  
512 mended Practices for Measuring in Situ the ‘Loads’ on Drive Train, Pitch System and Yaw  
513 System.,” *Energy Research Center of the Netherlands (ECN)*, 2010.
- 514 [42] J. Quick, J. Annoni, R. King, K. Dykes, P. Fleming, and A. Ning, “Optimization under un-  
515 certainty for wake steering strategies,” in *Journal of Physics: Conference Series*, vol. 854,  
516 p. 012036, IOP Publishing, 2017.

- 517 [43] D. Simms, S. Schreck, M. Hand, and L. J. Fingersh, “Nrel unsteady aerodynamics experiment in  
518 the nasa-ames wind tunnel: a comparison of predictions to measurements,” tech. rep., National  
519 Renewable Energy Lab., Golden, CO (US), 2001.
- 520 [44] P. Welch, “The use of fast fourier transform for the estimation of power spectra: a method  
521 based on time averaging over short, modified periodograms,” *IEEE Transactions on audio and*  
522 *electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.
- 523 [45] M. Stein, “Large sample properties of simulations using latin hypercube sampling,” *Techno-*  
524 *metrics*, vol. 29, no. 2, pp. 143–151, 1987.
- 525 [46] M. J. Jansen, “Analysis of variance designs for model output,” *Computer Physics Communi-*  
526 *cations*, vol. 117, no. 1-2, pp. 35–43, 1999.
- 527 [47] S. Wold, K. Esbensen, and P. Geladi, “Principal component analysis,” *Chemometrics and*  
528 *intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.