Uncertainty propagation through an aeroelastic wind turbine model using polynomial surrogates

Murcia Leon, Juan Pablo; Réthoré, Pierre-Elouan Mikael; Dimitrov, Nikolay Krasimirov; Natarajan, Anand; Sørensen, John Dalsgaard; Graf, Peter; Kim, Taeseong

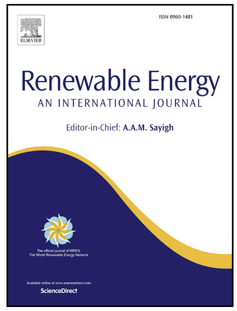
Published in:
Renewable Energy

DOI (link to publication from Publisher):
10.1016/j.renene.2017.07.070

Publication date:
2018

Document Version
Accepted author manuscript, peer reviewed version

Citation for published version (APA):
Murcia Leon, J. P., Réthoré, P-E. M., Dimitrov, N. K., Natarajan, A., Sørensen, J. D., Graf, P., & Kim, T. (2018). Uncertainty propagation through an aeroelastic wind turbine model using polynomial surrogates. *Renewable Energy*, *119*, 910-922. https://doi.org/10.1016/j.renene.2017.07.070



Uncertainty propagation through an aeroelastic wind turbine model using polynomial surrogates

Juan Pablo Murcia, Pierre-Elouan Réthoré, Nikolay Dimitrov, Anand Natarajan, John Dalsgaard Sørensen, Peter Graf, Taeseong Kim

PII: S0960-1481(17)30698-5

DOI: 10.1016/j.renene.2017.07.070

Reference: RENE 9045

To appear in: *Renewable Energy*

Received Date: 27 September 2016

Revised Date: 17 April 2017

Accepted Date: 14 July 2017

Uncertainty propagation through an aeroelastic wind turbine model using polynomial surrogates

Juan Pablo Murcia[a], Pierre-Elouan Réthoré[b], Nikolay Dimitrov[b], Anand Natarajan[b], John Dalsgaard Sørensen[b,c], Peter Graf[d], Taeseong Kim[b]

[a]*jumu@dtu.dk, Department of Wind Energy, Technical University of Denmark. Risø Campus, Frederiksborgvej 399 Building 125. 4000 Roskilde. Denmark*
[b]*Department of Wind Energy, Technical University of Denmark*
[c]*Department of Civil Engineering, Aalborg University*
[d]*National Renewable Energy Laboratory, Colorado USA*

Abstract

Polynomial surrogates are used to characterize the energy production and lifetime equivalent fatigue loads for different components of the DTU 10 MW reference wind turbine under realistic atmospheric conditions. The variability caused by different turbulent inflow fields are captured by creating independent surrogates for the mean and standard deviation of each output with respect to the inflow realizations. A global sensitivity analysis shows that the turbulent inflow realization has a bigger impact on the total distribution of equivalent fatigue loads than the shear coefficient or yaw miss-alignment. The methodology presented extends the deterministic power and thrust coefficient curves to uncertainty models and adds new variables like damage equivalent fatigue loads in different components of the turbine. These surrogate models can then be implemented inside other work-flows such as: estimation of the uncertainty in annual energy production due to wind resource variability and/or robust wind power plant layout optimization. It can be concluded that it is possible to capture the global behavior of a modern wind turbine and its uncertainty under realistic inflow conditions using polynomial response surfaces. The surrogates are a way to obtain power and load estimation under site specific characteristics without sharing the proprietary aeroelastic design.

Keywords: Wind energy, uncertainty quantification, aeroelastic wind turbine model, annual energy production, lifetime equivalent fatigue loads

1. Introduction

The wind turbine design standard IEC 61400-1 [1] provides wind climate specifications which are used as a reference for the structural design of the wind turbines. For achieving type certification of a new turbine model, the designer has to demonstrate that the structural capacity of the turbine is sufficient for withstanding the reference wind conditions over the entire lifetime of the turbine. Such a demonstration is normally given by dynamic load simulations which characterize the behavior of the turbine under the reference wind conditions. Once certification is achieved, the given turbine model can safely be installed on sites where the wind conditions are identical or more benign than the reference standard conditions. However, in many occasions one or more of the parameters describing the site environmental conditions will be outside the ranges which are sufficiently covered by the IEC reference conditions. In such cases, it is necessary to estimate the actual loads which the turbine will experience over its entire lifetime, by considering the full joint distribution of the variables that describe the turbulent inflow. This is similar to a propagation of uncertainty problem in which the distribution of the atmospheric conditions on the site needs to be propagated through the aeroelastic model of the turbine.

If a full design load case setup similar to the IEC 61400-1 design cases is used for that purpose, the problem quickly becomes time-consuming as new dynamic simulations would be required for each site. As an example, the number of simulations required to predict within 1% error the lifetime equivalent fatigue loads on a floating wind turbine where the inflow conditions (sea/wind) are characterized by five stochastic variables can reach up to $3,200,000 = 20^5$ using regular grid-based estimates or in the order of 50,000 using Monte-Carlo (MC) simulation [2]. An approach that alleviates these issues is mapping the turbine response to different environmental inputs by means of a fast and accurate surrogate model. Several techniques can be used to predict the behavior of the turbine from a limited set of model evaluations such as: interpolation techniques, response surface techniques [3], Gaussian process (Kriging) [4] and machine learning techniques [5, 6].

Polynomial chaos expansion is a methodology used to efficiently propagate input uncertainties through a non-linear model. This methodology consists in building a polynomial response surface to capture the global dependency of the output as a function of the uncertain inputs. PCE is widely used in the uncertainty quantification field because of its simplicity and fast convergence in comparison to a full MC simulation based on the original model [7, 8, 9, 10, 11]. Furthermore, adaptive PCE training algorithms can be used to obtain a sparse surrogate that minimizes the number of terms that have multiple variable dependency, making the surrogates extremely efficient response surfaces in multiple dimensions [12, 13, 14]. In the case of smooth continuous models with multiple input variables, sparse polynomial chaos expansion methodology is the most efficient technique to build the surrogates in terms of the number of model evaluations required, the number of input dimensions they can handle and the rate of convergence [12].

One of the main difficulties in building a surrogate of an aeroelastic wind turbine model is the fact that the turbulent inflow realization (TIR, i.e. turbulent structures in the flow field) causes variations in the different wind turbine model outputs: such as power, thrust, fatigue and extreme loads in the different components of the turbine. This can be restated as: an aeroelastic wind turbine model has stochastic/non-deterministic outputs. Many studies have analyzed the difficulties of studying fatigue and extreme loads under different turbulent inflow realizations [15, 16, 17, 4, 3]. Different TIR activate different dynamics of the structure and have different control system responses; therefore are an important source of uncertainty in the prediction of the outputs of the model [15]. The high variability in the model response to certain turbulent inflow structures has also been shown to be problematic when MC simulation was used to predict lifetime averages of fatigue loads on a floating wind turbine [2].

1.1. Response to the problem

The aim of the present study is to demonstrate a method for building a quick and accurate surrogate of a wind turbine model that predicts the turbine response as a function of multiple stochastic input variables that describe the turbulent inflow on a site ($\mathbf{x}$). The surrogate for the turbine model is a set of two independent sparse polynomial response surfaces that allow to predict the variability caused by different input variable distributions and by different turbulent inflow field realizations (TIR). One response surface characterizes the expected output with respect to TIR: $\hat{y}_{\mathbb{E}}(\mathbf{x}) \approx \mathbb{E}_{\text{TIR}}(y|\mathbf{x})$. The other one describes the standard deviation of the output with respect TIR: $\hat{y}_{\mathbb{S}}(\mathbf{x}) \approx \sqrt{\mathbb{V}_{\text{TIR}}(y|\mathbf{x})}$; which is a model that predicts the uncertainty in the turbine response due to different turbulent structures hitting the turbine. Finally, a sample can be obtained from the normal distribution constructed using the mean and the standard deviation surrogates in order to make a prediction of the variability in the output at a given input point:

$$\hat{y}(\mathbf{x}) \sim \text{Normal}(\hat{y}_{\mathbb{E}}(\mathbf{x}),\, \hat{y}_{\mathbb{S}}(\mathbf{x})) \tag{1}$$

The final surrogate $\hat{y}(\mathbf{x})$ can then be used to obtain distributions of the wind turbine power and fatigue loads in a given year whose input parameters (wind, wind/sea, or wind/geological conditions) follow the distribution used to train the surrogate PDF($\mathbf{x}$). Since the surrogate is a response surface it can also be used to predict the distribution of the outputs when the input distributions is close but not exactly the distribution used for training the surrogate. This setup is considered a multi-leveled uncertainty propagation and it is the scenario that occurs when there is uncertainty in the parameters that characterize the WS distribution for example. This approach is necessary to estimate the uncertainty in annual energy production and lifetime averaged equivalent fatigue load.

1.2. Article overview

A general overview of the PCE methodology in multiple dimensions is presented in section 2. This section describes the Rosenblatt transformation, the design of experiments used to define the training simulation points, the approach used to train sparse polynomial response surfaces and the logistic transformation used to limit the output. In section 3, the methodology is then applied to the response of the DTU 10 MW reference wind turbine HAWC2 model [18] to turbulent inflow fields characterized by four input parameters. The four input parameters are the 10-min averaged hub height wind speed (WS), the turbulent standard deviation of the instantaneous wind speed in the streamwise component (σ_1), the shear exponent (α) and the yaw misalignment angle (γ). A study of how many independent realizations of the turbulent inflow field are required to achieve a certain error tolerance in the surrogate is presented in the section 3.7. Finally in section 3.8, the surrogates are used in an example of prediction of the uncertainty in the annual energy production and the uncertainty in lifetime averaged equivalent fatigue loads.

2. Methods

This article proposes the use of two different variable transformations to simplify the polynomial response surface fitting problem, see figure 1. The first transformation is the Rosenblatt transformation [19], which is used to de-correlate the set of D input variables $\mathbf{x} = (x_0, x_1, \ldots x_{D-1})$ into a set of independent uniform variables, $\mathbf{w} = (w_0, w_1, \ldots w_{D-1})$. The second transformation is a logistic transformation, and it is used to enforce constraints on the polynomial surrogates [20]. This transformation enables the use of polynomial surrogates in problems where the output has a minimum and/or maximum value. Without the logistic transformation the polynomial surrogates will present oscillations in the regions where the model has a constant output. The power production of a turbine is an example of a variable with a strict upper constraint corresponding to the rated power.

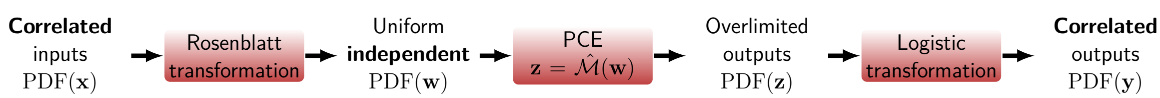


Figure 1: Transformation of variables to build efficient polynomial response surface.

2.1. 1D PCE theory

Consider a model with a single uncertain input (x) and a single output (y). PCE consists of defining a polynomial family that is orthogonal with respect to the input distribution, PDF(x). Orthogonal polynomial families with respect to the most important distributions are well known, see table 1. For details on how to define new polynomial basis to an arbitrary input distributions refer to Gautschi et al [21].

Distribution	Polynomial Family
Uniform	Legendre
Normal	Hermite
Exponential	Laguerre

Table 1: Classical orthogonal polynomial families.

The orthogonal polynomials are used to build a polynomial approximation of the output, i.e. a polynomial response surface, see equation 2. Where, $\phi_l(x)$ is the l

order orthogonal polynomial, c_l is its correspondent coefficient and M represents the truncation order of the PCE.

$$y(x) \approx \hat{y}(x) = \sum_{l=0}^{M} c_l \, \phi_l(x) \tag{2}$$

There are two different approaches to determine the c_l coefficients:

Semi-Spectral projection consists in using quadrature rules to approximate the inner product definition of the coefficient, see equation 3. Many quadrature rules exist to approximate the integrals; but all quadrature rules give N_n nodes for model evaluation (x_i) and their corresponding weights (ω_i). Gaussian quadrature rules are widely used because they are accurate for smooth function integration with respect a weight function, in this case the PDF(x), see equation 3.

$$c_l = \langle y, \phi_l \rangle \equiv \int y(x) \, \phi_l(x) \, \mathrm{PDF}(x) \, dx \approx \sum_{i=0}^{N_n} \omega_i \, y(x_i) \, \phi_l(x_i) \tag{3}$$

In general, semi-spectral projection is an efficient method for low number of input dimensions, but the number of model evaluations required grows exponentially with the number of dimensions. Additionally, quadrature rules can be unstable for heavy tailed PDFs such as the Weibull distribution [21].

Point collocation consists in fitting the polynomial basis to a small sample of model evaluations. Traditionally, this fit can be done using least squares algorithm, but some other optimization algorithms can be used to obtain PCE approximations that minimize the number of terms in the surrogate [12, 13, 14]. This techniques are explained in the section 2.5. In general, point collocation is robust and the advanced optimization algorithms are designed to handle large number of dimensions, to avoid over-fitting and to achieve sparsity in the final surrogate. The present study focuses only in the point collocation techniques since the number of model evaluations required to fit a multiple dimensional PCE is smaller [12] than in other methods.

2.2. Rosenblatt transformation

To build the PCE of a model with multiple correlated inputs ($\mathbf{x}$), it is required to initially transform the correlated input space into an uncorrelated space ($\mathbf{w} = R^{-1}(\mathbf{x})$). In this article, the Rosenblatt transformation is used because the input distribution of the turbulent inflow field parameters are usually defined in a sequence of conditional relationships [19]. Refer to Dimitrov et al [22] and Graf et al [2] for examples of such distributions used for offshore and floating wind turbine fatigue and extreme load analysis.

Since all the variables are transformed into uncorrelated unitary uniform variables then the PCE only requires the use of the Legendre polynomials: $y(\mathbf{x}) = y(R(\mathbf{w})) \approx \hat{y}(\mathbf{w})$.

2.3. Multi-dimensional PCE

A D-dimensional polynomial is constructed as the sum of the product between individual one dimensional polynomials for each of the D uniform input variables, $\mathbf{w} = [w_0, \ldots, w_{D-1}]$. The D-dimensional surrogate is written using a set of multiple indexes $\mathcal{I} \subset \mathbb{N}^D$. An element $J \in \mathcal{I}$ contains the order of the polynomial in each dimension: $J = [l_0, \ldots, l_{D-1}]$. Additionally, the multiple indexes are enumerated, $J \leftrightarrow j \in \mathbb{N}$. A surrogate that contains N_c terms can be written as:

$$y(\mathbf{x}) = y(R(\mathbf{w})) \approx \sum_{j=0}^{N_c-1} c_j \, \boldsymbol{\phi}_j(\mathbf{w}) \tag{4}$$

where an element in the multidimensional polynomial basis is given as:

$$\boldsymbol{\phi}_j(\mathbf{w}) = \phi_{l_0}(w_0) \times \cdots \times \phi_{l_{D-1}}(w_{D-1}) \tag{5}$$

2.4. Training point selection

The Rosenblatt transformation enables the use of multiple variance reduction MC sampling techniques to define the training points of a surrogate [23]. Latin hypercube sampling [24], Sobol sequence [25] and Hammersley sequence [26] are some examples of such techniques. These techniques are designed to sample from the unitary hypercube of D dimensions, i.e. the uniform distributed variables: $\mathbf{w}_i \sim \text{PDF}(\mathbf{w})$. Finally, the Rosenblatt transformation is used to transform each realization in the uniform sample into the correlated input space, $\mathbf{x}_i = R(\mathbf{w}_i) \sim \text{PDF}(\mathbf{x})$.

The number of unknown coefficients c_j in a D-dimensional PCE depends of the total polynomial order of the PCE. The total order is defined as the maximum sum of the one dimensional orders. If the PCE is truncated to a total order M then the number of unknown coefficients is given by the following combination:

$$N_c = \binom{M+D}{M} = \frac{(M+D)!}{M!\,D!} \tag{6}$$

The number of model evaluations should be between 2 or 3 times the number of unknowns in order to have extra data to test the accuracy of the surrogate and to implement strategies to avoid over-fitting [12]. Note that the maximum order is only used to estimate the number of model evaluations. Advanced regression techniques allow to explore higher order terms [14, 12]. The maximum order M can be increased in order to achieve higher accuracy surrogates but at the cost of having more model evaluations and the requirement of assuring that there is no over-fitting.

2.5. Point collocation and the LASSO problem

The least absolute shrinkage and selection operator (LASSO) problem is a modified least squares optimization problem that adds a term that penalizes the amount of active terms in the surrogate (terms with non zero coefficients). LASSO is used to achieve sparsity and to avoid over fitting in the polynomial surrogate. Additionally, the number of model evaluations required for solving the LASSO problem is smaller in comparison to a least squares regression that has the same maximum total polynomial order [12].

A LASSO problem can be described as finding the set of coefficients c_j that minimizes the sum of squared errors plus the sum of the absolute values of all coefficients (ℓ_1 norm regularization term) [14]:

$$\min_{c_j} \quad \sum_{i=0}^{N-1} \left[\sum_{j=0}^{N_c-1} c_j \boldsymbol{\phi}_j(\mathbf{w}_i) - y(\mathbf{x}_i) \right]^2 + \alpha \sum_{j=0}^{N_c-1} |c_j| \tag{7}$$

where the number of model/surrogate evaluation points N is fixed. Note that the input and surrogate evaluation points are related by the Rosenblatt transformation $\mathbf{x}_i = R(\mathbf{w}_i)$. The maximum number of possible terms of the surrogate N_c is fixed by selecting a maximum total multi-dimensional polynomial order.

The regularization coefficient α controls the amount of active terms in the final solution. Smaller values allow to have more active terms while larger values will prefer final surrogates with few active terms. A sparse surrogate has the advantage of making the evaluation of the multi-dimensional surrogate faster in comparison to the full least squares solution; this advantage becomes critical in high number of input dimensions.

There are two algorithms widely used to solve the LASSO problem: coordinate descent [14] and least angle regression (LAR) [12]. Coordinate decent is used in the present work because it tends to be more stable for high dimensional problems [13]. The reason for this is that coordinate descent operates on a given regularization coefficient instead of exploring the full space of α's as in LAR algorithm.

Cross-validation is used to select the regularization coefficient α that minimizes over fitting of the data. A k-fold cross-validation consists in splitting the dataset into k groups of data. All the points in k-1 groups are used for training and the remaining group is used for cross-validation. This means that the surrogate fitted using k-1 groups is used to predict the output in each of the elements of the remaining group. The mean squared error of the prediction of the surrogate is then computed. This process is repeated leaving out each individual fold and for multiple regularization parameters. The regularization parameter that gives the lowest mean cross-validation mean squared errors is then selected to train the whole dataset. This translates as selecting the sparse model that performs the best by predicting missing data, i.e. that has less over-fitting.

2.6. Logistic transformation

A logistic transformation is applied to an output of the model in order to avoid oscillations in the regions where the model is constant. In practice this transformation is used to impose strict restrictions on the polynomial surrogates. The transformation consists in applying the *logit* function, $L(p) = \ln\left(\frac{p}{1-p}\right)$, to the model output at the training points $y_i = y(\mathbf{x}_i)$ into the over-shooting variable space: $z_i = L(a_1\, y_i + a_0)$ [20]. Finally, each time the surrogate is evaluated, the prediction of the surrogate is transformed back to the original output space $\hat{y} = (L^{-1}(\hat{z}) - a_0)/a_1$. The constants of the transformation are calibrated in order to impose the constraints of the output and to avoid numerical instabilities that are inherent to the logit function.

2.7. Global sensitivity analysis

Global sensitivity analysis (SA) is a methodology to determine how important each input is to explain the variance of the output. SA can be obtained with a Sobol variance decomposition [27]. In this technique, the variance of the output is explained into the different terms of variance of each of the inputs, in a process similar to the analysis of the variance of experiments (ANOVA) [28]. Total effect Sobol indices are widely used as measures of how much of the variance of a given output is explained by the variance of an input, including possible interactions with other variables. This method is the most recognized method for global sensitivity analysis because it accounts for non-linear dependencies and for interactions between variables [29].

Variance decomposition can be expressed as the sum of the variance of the marginal expected value of a subset of input variables, see equation 8. Note that this decomposition is not an infinite series expansion, it is truncated to the maximum number of variable interactions.

$$\begin{aligned}
\mathbb{V}(y) &= \sum_{k=0}^{D-1} \mathbb{V}_k + \sum_{k=0}^{D-1}\sum_{l>k}^{D-1} \mathbb{V}_{kl} + \sum_{k=0}^{D-1}\sum_{l>k}^{D-1}\sum_{m>l}^{D-1} \mathbb{V}_{klm} + \cdots + \mathbb{V}_{0\ldots D-1} \\
\mathbb{V}_k &= \mathbb{V}\left(\mathbb{E}_{\forall n\neq k}\left(\mathcal{M}(\mathbf{x}|x_k)\right)\right) \\
\mathbb{V}_{kl} &= \mathbb{V}\left(\mathbb{E}_{\forall n\neq k,l}\left(\mathcal{M}(\mathbf{x}|x_k, x_l)\right)\right) \\
\mathbb{V}_{klm} &= \mathbb{V}\left(\mathbb{E}_{\forall n\neq k,l,m}\left(\mathcal{M}(\mathbf{x}|x_k, x_l, x_m)\right)\right)
\end{aligned} \tag{8}$$

The global sensitivity measure is defined by normalizing eq. 8 with the total variance of the output $\mathbb{V}(y)$. From this normalization one can define the Sobol index of a given degree of interaction between input variables as:

$$S_k = \frac{\mathbb{V}_k}{\mathbb{V}(y)} \qquad S_{kl} = \frac{\mathbb{V}_{kl}}{\mathbb{V}(y)} \qquad S_{klm} = \frac{\mathbb{V}_{klm}}{\mathbb{V}(y)} \quad \ldots \tag{9}$$

The total effect Sobol index of an input variable x_i is then the sum of all the Sobol indices that include the variable in any interaction:

$$S_{\mathrm{total}\,x_i} = S_i + \sum_{\substack{k=0\\k\neq i}}^{D-1} S_{ik} + \ldots \tag{10}$$

The sensitivity analysis of the response of the turbine should consider the effect of having different turbulent inflow realizations which is modeled with the two independent polynomial response surfaces for the local mean and local standard deviation. The Sobol indexes are not computed directly from the PCE coefficients as for classical problems, see Sudret et al [30], because the Logistic transformation removes the stochastic properties of the PCE and because the coefficients of the local mean surrogate would not include the effect of the turbulence inflow realization. To solve this limitation, the approximate method proposed in Saltelli et. al [29] is used to compute the total effect Sobol indexes. This approach estimates the total effect Sobol indexes from a large MC simulation.

3. Results

3.1. Implementation

Several open source implementations of PCE methods are available such as: Chaospy [23], Dakota [31], UQLab [32] and OpenTurns [33]. In the present work we use Chaospy because of its implementation of the Rosenblatt transformation. Additionally, the present work uses the LASSO problem solvers [14] and the cross-validation capabilities available in the open source library Scikit-learn [13]. These capabilities are used inside of Chaospy for general users and are used externally in the present study to gain control over the different stages of the cross-validation.

3.2. Case description

The model consists of the DTU 10 MW reference wind turbine HAWC2 model [34, 18] with Mann turbulent inflow generation [35]. The turbulent inflow conditions are defined using the four variables described in table 2.

Input	Variable	Distribution	Parameters	
10-min mean hub height wind speed	$x_0 = \mathrm{WS}$	Rayleigh	$\mathbb{E}(\mathrm{WS}) = 10$ m/s	
Std. of the inst. wind speed in the streamwise direction during the 10-min simulation	$x_1 = \sigma_1$	Lognormal	$\mu_{\sigma_1}(\mathrm{WS})$	$\sigma_{\sigma_1}(\mathrm{WS})$
10-min mean shear exponent	$x_2 = \alpha$	Normal	$\mu_\alpha(\mathrm{WS})$	$\sigma_\alpha(\mathrm{WS})$
10-min mean yaw miss-align.	$x_3 = \gamma$	Normal	$\mu_\gamma = 0$	$\sigma_\gamma = 5$ deg.

Table 2: Wind turbine model inputs.

The dependency between WS and σ_1 is defined in the Normal Turbulence Model described in the IEC 61400-1 [1]. The present case uses a reference ambient turbulence intensity of a site Class 1A: $\mathrm{TI}_{\mathrm{ref}} = 0.16$. This dependency is given by the local statistical moments of σ_1 as: $\mathbb{E}(\sigma_1|\mathrm{WS}) = \mathrm{TI}_{\mathrm{ref}}\,(0.75\mathrm{WS} + 3.8)$ and $\mathbb{V}(\sigma_1|\mathrm{WS}) = (1.4\,\mathrm{TI}_{\mathrm{ref}})^2$.

The parameters of the σ_1 distribution are given in equation 11 as functions of WS.

$$
\begin{aligned}
\sigma_{\sigma_1} &= \left(\ln\left(\frac{\mathbb{V}(\sigma_1|\mathrm{WS})}{\mathbb{E}^2(\sigma_1|\mathrm{WS})} + 1\right)\right)^{1/2} = \left(\ln\left(\frac{1.4^2}{(0.75\mathrm{WS} + 3.8)^2} + 1\right)\right)^{1/2} \\
\mu_{\sigma_1} &= \ln\left(\mathbb{E}(\sigma_1|\mathrm{WS})\right) - \frac{\sigma_{\sigma_1}^2}{2} = \ln\left(\mathrm{TI}_{\mathrm{ref}}\,(0.75\mathrm{WS} + 3.8)\right) - \frac{\sigma_{\sigma_1}^2}{2}
\end{aligned}
\tag{11}
$$

The correlation between α and WS is based on the simplified joint distribution defined by Dimitrov et al [22]:

$$\begin{aligned} \mu_\alpha &= 0.088(\ln(\mathrm{WS}) - 1) \\ \sigma_\alpha &= 1/\mathrm{WS} \end{aligned} \tag{12}$$

Seven different model outputs are considered ($\mathbf{y}$), see table 3. The damage equivalent fatigue loads (EFL) are computed using a rainflow counting algorithm to determine the number of load cycles n_i with their corresponding load range S_i in the 10-min time series of turbine response. The EFL is then obtained using different materials' Wöhler exponent m, see equation 13 [36]. For obtaining 1Hz-equivalent fatigue loads based on 10 minute reference periods, the reference number of load cycles used is $N_{\mathrm{ref}} = 600$.

$$S_{\mathrm{eq}} = \left(\frac{\sum n_i S_i^m}{N_{\mathrm{ref}}} \right)^{\frac{1}{m}} \tag{13}$$

Output	m	Variable
10 minute mean power production	-	P
10 minute mean thrust coefficient	-	CT
EFL blade root flapwise bending moment	12	BRF
EFL tower bottom fore-aft bending moment	4	TBF
EFL tower bottom sidewise bending moment	4	TBS
EFL tower top tilt bending moment	4	TTT
EFL tower top yaw bending moment	4	TTY

Table 3: Wind turbine model outputs.

3.3. Training points

In this study, the number of model evaluations are set to be $N = 2\,N_c$, the maximum order of the polynomial is expected to be $M = 4$ and the number of input variables is $D = 4$. This leads to 140 total number of model evaluations, i.e. 140 input variables locations for which HAWC2 model is executed, see equation 6. A Hammersley sequence [26] is preferred over other variance reduction methods to generate the training sample in the uniform space as it is a sequence that can be extended to contain larger sample size without changing the previous points [23, 37]. The uniform sample is then transformed into the physical variables using the Rossenblat transformation. A similar approach is used to generate the input sample for a MC simulation on either the real model or the surrogate models; the size of the MC sample is taken to be 80000. The training input sample is shown in figure 2 as well as a the inputs sample for the MC simulation. figure 2 is a representation of the multidimensional PDF($\mathbf{x}$): the histograms represent the marginal distributions for each variable, while the plots in the lower diagonal represent the training points and bi-dimensional histograms of the MC sample. The figures in the lower diagonal show the correlations between each pairs of variables as well as the iso-pdf quantiles that enclose 68%, 95% and 99.7% of the data.

It can be observed that the training points are more densely distributed in the regions of higher probability of the inputs. This means that the surrogate is better trained in the most likely region of the input space. For applications where the quantity of interest is the tail of the output distribution, such as ultimate load estimation, the training dataset could be distributed uniformly over the region encircled by a given iso-pdf quantile of the inputs, see iso-PDF contours in figure 2. 100 different turbulent inflow realizations are generated using the Mann model for each input point, for which the mean and standard deviation of the outputs are obtained. This number is selected to test the accuracy of the prediction of the surrogates when they are trained using a reduced number of TIR as it is defined in the design load cases defined in the standard [1]. The full training sample consists of 140 × 100 HAWC2 10 minutes simulations.

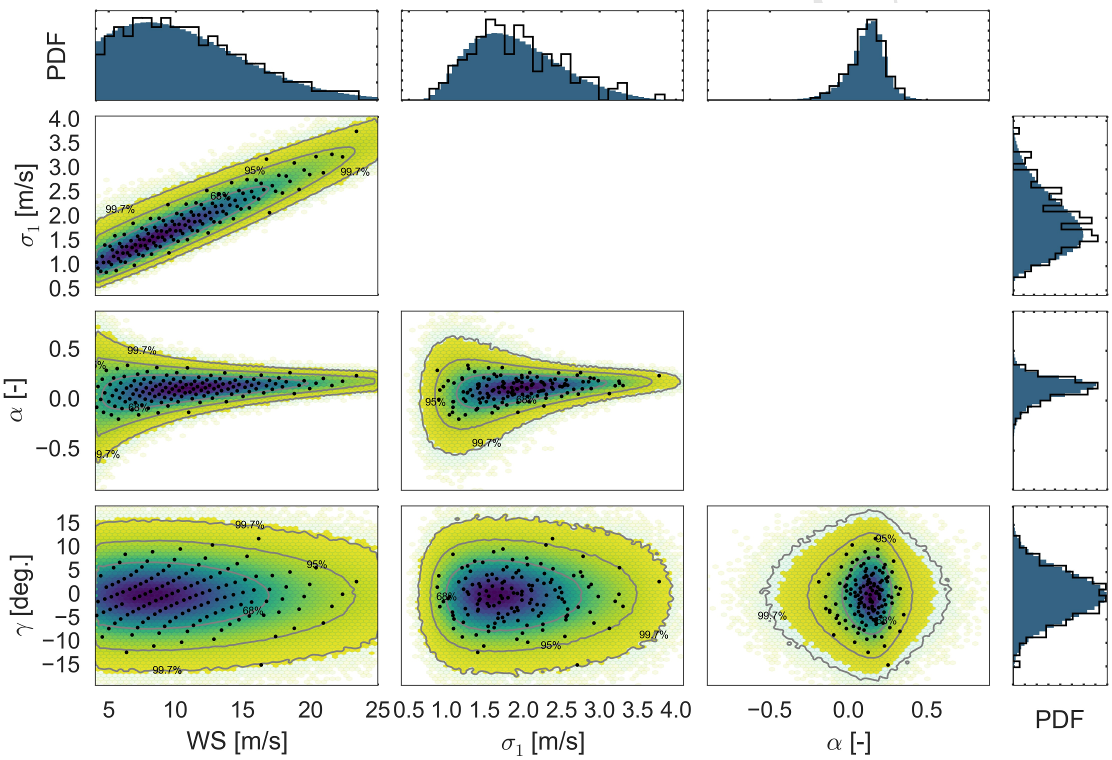


Figure 2: (Black points) Training dataset in the inputs: 140 Hammersley sequence sample of input joint distribution. (Histogram colored hex-bins) 80000 Hammersley sequence MC sample. (Contour lines) Iso-PDF lines that encircle 68%, 95% and 99.7% of the MC sample.

3.4. Example of PCE surrogates for individual statistical moments

Some examples of the distribution of $\mathbf{y}_{\mathbb{E}}$ and $\mathbf{y}_{\mathbb{S}}$ [1] are shown in figure 3. In this figure the black points represent the observed statistic of the output for the training

[1] $P_{\mathbb{S}}$ represents the standard deviation of 100 different realizations of the 10-min averaged power; this variable should not be confused with the standard deviation of the instantaneous power during the 10 minutes of simulation.

points; while the bi-dimensional histogram represents the obtained distribution of the surrogate for a 80000 MC sample. The observed histogram in the training dataset and the PDF predicted by the surrogate for $\mathbf{y}_{\mathbb{E}}$ and $\mathbf{y}_{\mathbb{S}}$ are shown in the last column in figure 3. It can be observed that the surrogates accurately capture the global PDF of the model and its dependency with respect to the 4 input variables. The surrogates of the local standard deviations, $\hat{y}_{\mathbb{S}}$, are not able to capture the behavior of some extreme cases, see the extreme points at low wind speeds in the plots for $\mathrm{CT}_{\mathbb{S}}$ and $\mathrm{BRF}_{\mathbb{S}}$. These errors are small in comparison to the overall magnitude of the output; the distribution of the errors of the surrogates and its impact in the final prediction are quantified in section 3.7. These errors can be reduced up to a tolerance level selected by the user by adding more training points (input points with their turbulent inflow realizations). The surrogates are robust enough to predict the frequency of occurrence of extreme values such as the outputs resulting from the input point with largest σ_1, see first and third row in figure 3. This point seems to be outside the main trend in WS in figure 3 because it has a large σ_1 and α given its WS, see figure 2.

3.5. Final surrogate predictions

The surrogates of $\mathbf{y}_{\mathbb{E}}$ and $\mathbf{y}_{\mathbb{S}}$ are combined to estimate the distribution of each individual output of the DTU 10 MW RWT. The prediction is done by sampling the normal distribution constructed using the surrogates of $\mathbf{y}_{\mathbb{E}}$ and $\mathbf{y}_{\mathbb{S}}$, see equation 1. These results are presented in figure 4 along with the full dataset of HAWC2 simulations. In this figure each cross represents an individual 10-min simulation, therefore the scatter of nearby simulations illustrates the stochasticity in the output of the aeroelastic simulation. The amount of local output variability due to the turbulent inflow realization varies between outputs and depends on the region of the input space. The effect of the turbulent inflow realization is more important for the fatigue loads than for power and thrust coefficient. figure 4 also presents the bi-dimensional histogram obtained with a 80000 MC simulation of the surrogate. The distribution predicted by the surrogate captures the dependency and variability of each output with respect to the four input variables; the iso-PDF quantiles that encircle the 68%, 95% and 99.7% of the MC sample are also shown in figure 4 and they give a visual estimation of how likely are the observations of the output. It can be observed that the surrogate estimates the regions that contribute more on the lifetime fatigue and even gives an estimation of the input region on which the largest damage is to be expected. Additionally the MC simulation on the surrogate gives an estimation of the PDF for each variable, see fifth column in figure 4.

The obtained distribution of power shows a similar behaviour to the operational data of wind turbines; this shows that one of the main drivers for variability in the prediction of power below rated is the TIR. Similarly, the thrust coefficient shows large variability for wind speeds below rated; this large variability can become important for wake models that use the thrust coefficient to predict the strength of the wake of a turbine and its impact on other turbines in a wind farm. The fatigue load blade root and tower top bending moments (BRF, TTT and TTY) show similar dependency on

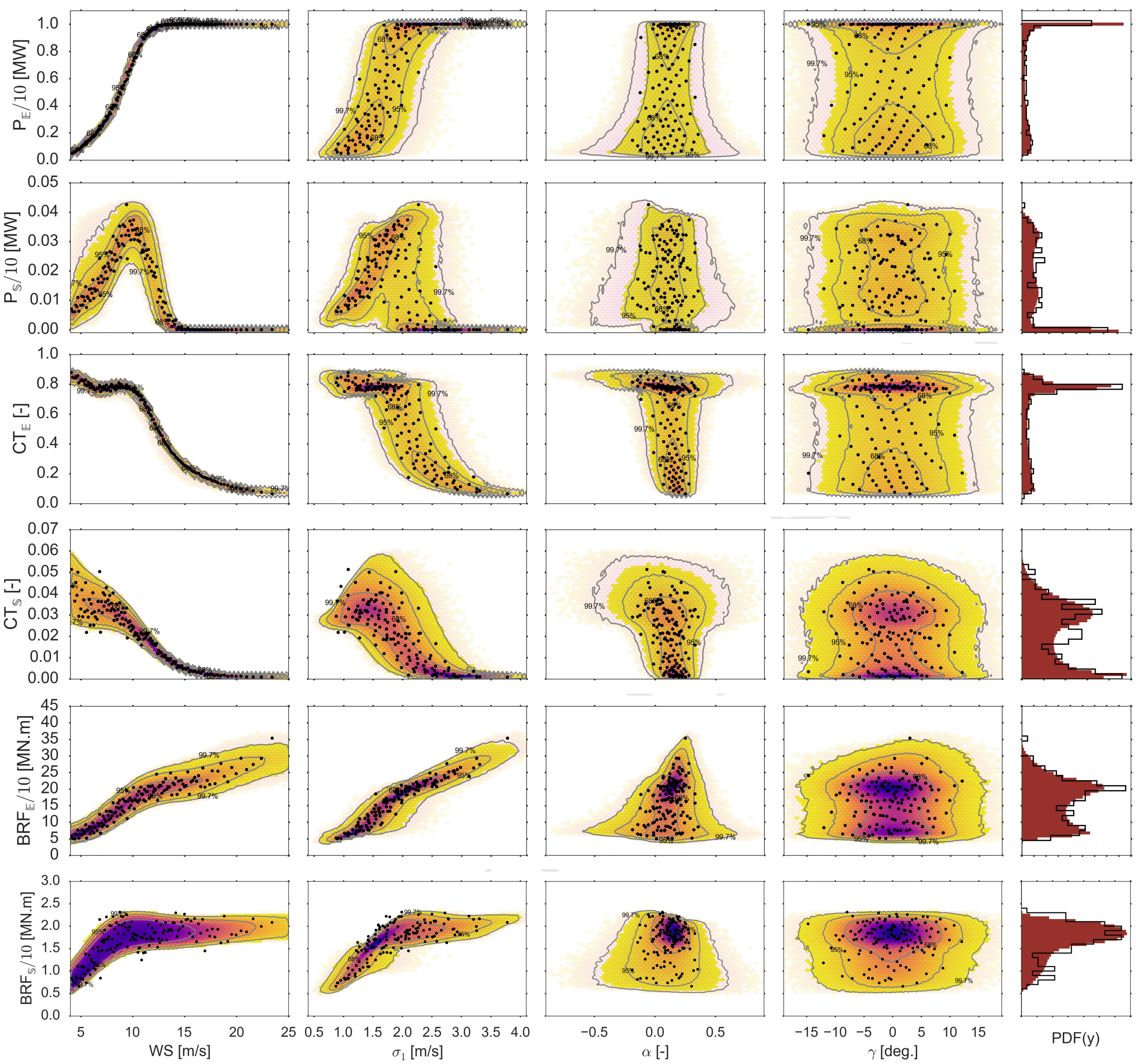


Figure 3: Example of surrogates for mean and std of the output with respect TIR. (Black points) 140 training points. (Histogram colored hex-bins) 80000 MC simulation on the surrogate. (Contour lines) Iso-PDF lines that encircle 68%, 95% and 99.7% of the MC simulation on the surrogate.

the four input variables and a similar amount of variability due to TIR; this is because they are all driven by the streamwise flow field. The fatigue loads tower bottom bending moments (TBF and TBS) show a different dependency on the input variables, mainly because they are driven by the thrust and sidewise forces; these two outputs have larger variability at lower WS which generates both the largest and lowest observations.

3.6. Sensitivity analysis

The global sensitivity analysis (SA) for the outputs are presented in table 4. The total effect Sobol indexes are computed using the approximation presented by Saltelli et al [29]. The total effect Sobol index represents the non-linear influence of the input

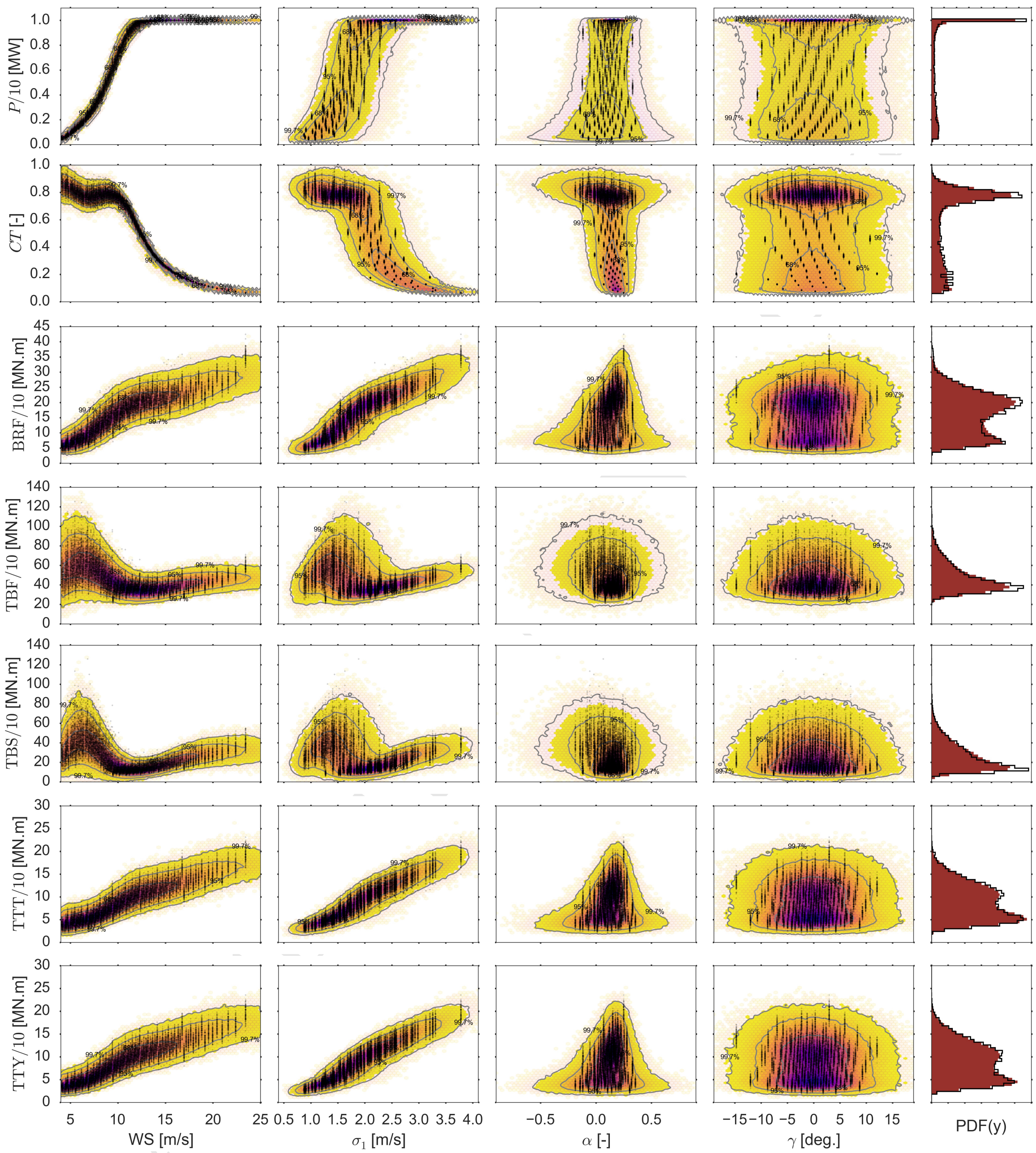


Figure 4: (Black crosses) 10-min HAWC2 simulation for the 140 input sample x 100 turbulent inflow realizations. (Histogram colored hex-bins) 80000 MC simulation of the surrogate. (Contour lines) Iso-PDF lines that encircle 68%, 95% and 99.7% of the MC simulation on the surrogate.

variable in the total variance of the output. Most of the outputs have a large total Sobol index for the wind speed. WS is clearly the main variable to explain the power and loads in a wind turbine. The SA shows that the power and thrust coefficient can be explained almost fully by the WS, since all the terms in the surrogate have WS dependency.

The variance introduced by the turbulent inflow realization is an important component for all the outputs, it has a higher influence than σ_1 for most outputs. This counter intuitive result is due to the large amount of correlation between WS and σ_1; thus a large fraction of the variance of the output generated by σ_1 is already explained by WS. The shear and yaw have reduced effects over most output variables. The yaw misalignment has reduced total effect because its assumed distribution is centered around zero. The shear exponent becomes important only for capturing the fatigue at the tower top tilt and yaw bending moments (TTT, TTY); while the yaw misalignment becomes important for modeling the fatigue at the tower bottom fore-aft moment (TBF).

	WS	σ_1	α	γ	TIR
P	1.0 1st	2.4×10^{-4} 4th	3.1×10^{-4} 3rd	8.1×10^{-5} 5th	3.1×10^{-3} 2nd
CT	9.9×10^{-1} 1st	1.2×10^{-3} 3rd	1.3×10^{-3} 4th	6.5×10^{-4} 5th	9.8×10^{-3} 2nd
BRF	8.8×10^{-1} 1st	5.6×10^{-2} 3rd	1.5×10^{-2} 4th	3.4×10^{-3} 5th	6.7×10^{-2} 2nd
TBF	5.9×10^{-1} 1st	2.1×10^{-1} 3rd	3.6×10^{-4} 5th	1.0×10^{-3} 4th	3.0×10^{-1} 2nd
TBS	7.1×10^{-1} 1st	7.6×10^{-2} 3rd	2.1×10^{-3} 5th	2.3×10^{-4} 4th	3.0×10^{-1} 2nd
TTT	8.7×10^{-1} 1st	7.1×10^{-2} 3rd	3.3×10^{-4} 5th	5.7×10^{-4} 4th	7.7×10^{-2} 2nd
TTY	8.7×10^{-1} 1st	6.8×10^{-2} 3rd	2.2×10^{-4} 5th	9.6×10^{-4} 4th	7.2×10^{-2} 2nd

Table 4: Total influence Sobol index.

The sensitivity analysis conditioned on WS for the outputs are presented in table 5. It can be observed that for power and thrust coefficient the influence of TIR goes from being the main source of variability at WS below rated to become the least important for WS above rated; this result summarizes the influence of the pitch controller enforcing the power and limiting the thrust. The effect of TIR in the fatigue loads is more uniform through all the ranges of operation. Similarly to the global SA , the main variables required to explain the equivalent fatigue loads are TIR and σ_1. This is also true for the power and thrust coefficient for WS bellow rated.

3.7. Convergence

A leave-one-out cross-validation (LOO) is done to estimate the distribution of the prediction error of each surrogate as a function of the number of independent turbulent seeds per input points used in the surrogate training. A LOO is a cross validation in which the surrogate is trained leaving one point out. Then, the local statistical

	WS=8 ms^{-1}				WS=12 ms^{-1}				WS=16 ms^{-1}			
	σ_1	α	γ	TIR	σ_1	α	γ	TIR	σ_1	α	γ	TIR
P	1.1×10^{-1} 3rd	1.4×10^{-1} 2nd	2.8×10^{-2} 4th	7.9×10^{-1} 1st	7.8×10^{-2} 2nd	3.7×10^{-2} 3rd	2.5×10^{-2} 4th	9.8×10^{-1} 1st	3.0 2nd	1.6 3rd	3.7 1st	9.7×10^{-1} 4th
CT	5.1×10^{-2} 3rd	1.1×10^{-1} 2nd	3.7×10^{-2} 4th	8.6×10^{-1} 1st	2.4×10^{-1} 2nd	2.1×10^{-1} 3rd	1.5×10^{-1} 4th	6.4×10^{-1} 1st	6.1×10^{-1} 1st	4.3×10^{-1} 2nd	3.3×10^{-1} 3th	2.0×10^{-1} 4th
BRF	4.8×10^{-1} 2nd	3.3×10^{-2} 3rd	1.1×10^{-2} 4th	5.0×10^{-1} 1st	3.9×10^{-1} 2nd	1.0×10^{-1} 3rd	9.2×10^{-3} 4th	5.1×10^{-1} 1st	3.5×10^{-1} 2nd	1.8×10^{-1} 3rd	2.7×10^{-2} 4th	4.6×10^{-1} 1st
TBF	3.7×10^{-1} 2nd	4.6×10^{-4} 4th	1.9×10^{-3} 3rd	6.5×10^{-1} 1st	5.6×10^{-1} 1st	2.1×10^{-3} 3rd	1.9×10^{-3} 4th	4.5×10^{-1} 2nd	5.2×10^{-1} 1st	3.6×10^{-3} 4th	4.0×10^{-3} 3rd	4.8×10^{-1} 2nd
TBS	1.9×10^{-1} 2nd	3.2×10^{-3} 3rd	6.8×10^{-4} 4th	8.3×10^{-1} 1st	2.4×10^{-1} 2nd	8.7×10^{-4} 4th	1.7×10^{-3} 3rd	7.8×10^{-1} 1st	2.2×10^{-1} 2nd	1.4×10^{-3} 4th	1.5×10^{-3} 3rd	7.9×10^{-1} 1st
TTT	5.6×10^{-1} 1st	2.2×10^{-3} 3rd	4.0×10^{-3} 4th	4.5×10^{-1} 2nd	4.6×10^{-1} 2nd	1.3×10^{-3} 4th	3.6×10^{-3} 3rd	5.5×10^{-1} 1st	4.6×10^{-1} 2nd	2.5×10^{-3} 4th	3.5×10^{-3} 3rd	5.4×10^{-1} 1st
TTY	5.3×10^{-1} 1st	1.9×10^{-3} 3rd	1.9×10^{-3} 4th	4.8×10^{-1} 2nd	4.6×10^{-1} 2nd	5.6×10^{-4} 4th	4.5×10^{-3} 3rd	5.5×10^{-1} 1st	4.7×10^{-1} 2nd	1.7×10^{-3} 4th	1.2×10^{-2} 3rd	5.3×10^{-1} 1st

Table 5: Total influence Sobol index at different WS.

moments of the output predicted by the surrogates at the missing point are compared against the statistics computed using the surrogate. In this article, the prediction errors are normalized with respect to the maximum scale of the output variable, which means that the errors represent the fraction of the total scale that should be considered as an extra uncertainty due to the inadequacy of the surrogate. The prediction error for the local surrogates are defined as:

$$\begin{aligned} \epsilon_{y\,\mathbb{E}} &= \frac{y_{\mathbb{E}}(\mathbf{x}_{LO}) - \hat{y}_{\mathbb{E}}(\mathbf{x}_{LO})}{\max(y)} \\ \epsilon_{y\,\mathbb{S}} &= \frac{y_{\mathbb{S}}(\mathbf{x}_{LO}) - \hat{y}_{\mathbb{S}}(\mathbf{x}_{LO})}{\max(y)} \end{aligned} \tag{14}$$

The convergence of the prediction error of the statistical moments is shown in figure 5. It can be seen that all the prediction errors tend to be distributed around zero and their standard deviations converge as the number of turbulent inflow realizations per input are increased. The errors converge to the distribution of the errors to the current surrogate. New input points need to be added to the training data set in order to further narrow the converged distribution of surrogate errors. In this figure the outliers are the extreme cases of selecting seeds with similar outputs, therefore, they are those cases that have large errors in the statistical moments. Finally, the converged distribution can be used to estimate the uncertainty in the final prediction of the output as:

$$\hat{y}(\mathbf{x}) \sim \text{Normal}(\hat{y}_{\mathbb{E}}(\mathbf{x}) + \epsilon_{y\,\mathbb{E}}\, max(y), \quad \hat{y}_{\mathbb{S}}(\mathbf{x}) + \epsilon_{y\,\mathbb{S}}\, max(y)) \tag{15}$$

where the errors of the surrogates can be sampled from the distribution predicted using LOO cross validation, see figure 5:

$$\epsilon_{y\,\mathbb{E}} \sim \text{Normal}(\mathbb{E}(\epsilon_{y\,\mathbb{E}}), \mathbb{S}(\epsilon_{y\,\mathbb{E}})) \quad \epsilon_{y\,\mathbb{S}} \sim \text{Normal}(\mathbb{E}(\epsilon_{y\,\mathbb{S}}), \mathbb{S}(\epsilon_{y\,\mathbb{S}})) \tag{16}$$

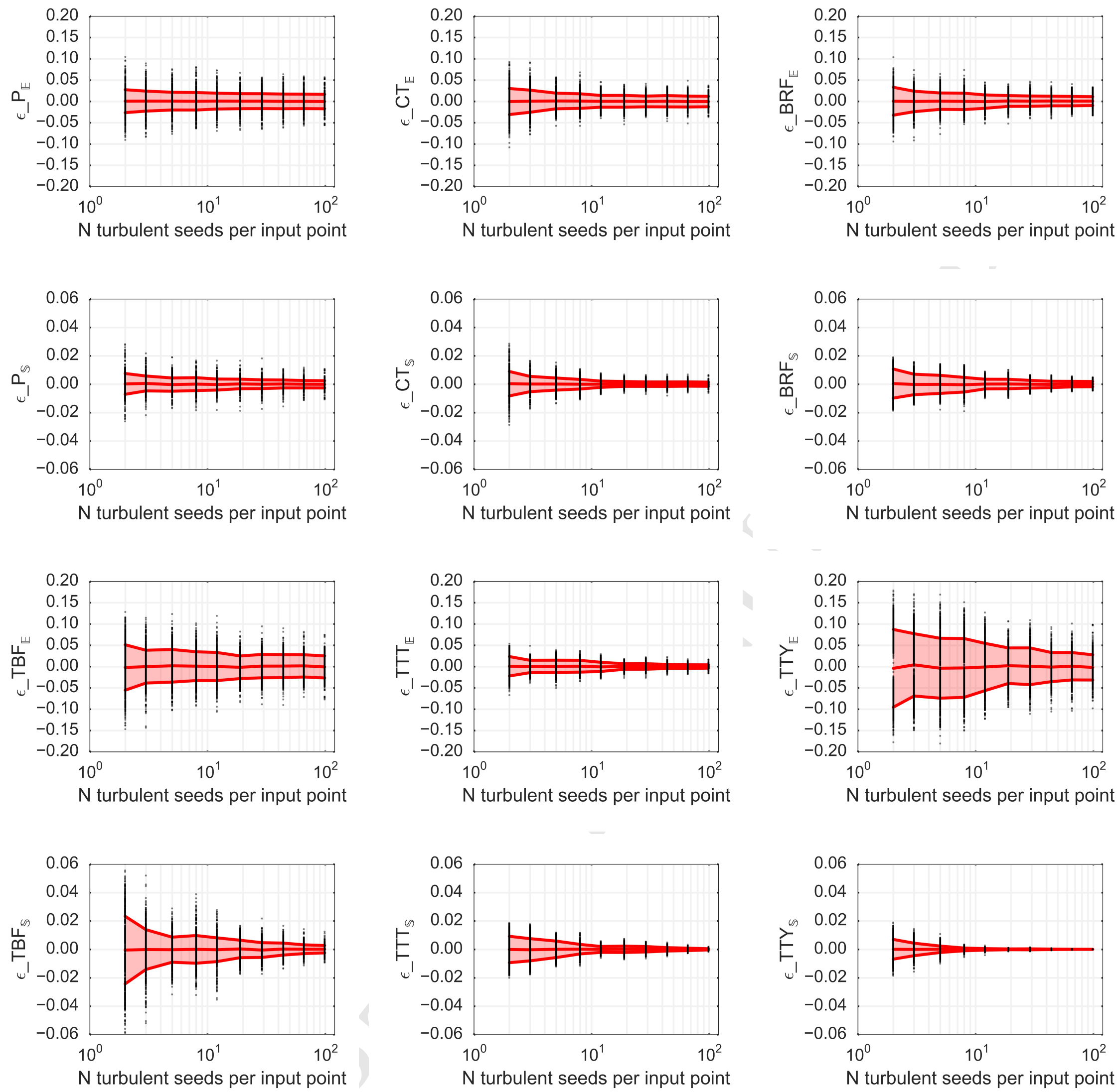


Figure 5: Convergence of the LOO cross-validation prediction error as a function of the number of turbulent seeds per input point used in PCE training. (Pink area) One standard deviation confidence interval around the mean $\mathbb{E}(\epsilon) \pm \mathbb{S}(\epsilon)$.

3.8. Example of using the surrogates for the estimation of the uncertainty in annual energy production and lifetime equivalent fatigue loads

This section presents an example to illustrate the use of the surrogates of the DTU 10 MW RWT to estimate the uncertainty in the distribution of expected energy production and of equivalent fatigue loads $\mathbb{E}_{\mathbf{x}}(\mathbf{y})$ in a given period; here the averaging period is either 1 year or 20 years. In this example a single turbine is planned to operate in a location from which the uncertainty in the wind resources has been estimated before hand. This uncertainty can represent the year-to-year variability, the effect of the long-term correction, uncertainty in the wind resources assessment tool, among other sources of uncertainty. The propagation of uncertainty is done in two steps as described in figure 6. The inner level predicts the distribution of the turbine outputs PDF($\mathbf{y}$) given a joint distribution of the turbulent inflow parameters PDF($\mathbf{x}$); the inner level returns the expected value of the output to the outer level. In the outer level the uncertainty in the resources is propagated through the inner level to estimate the uncertainty of the expected value of each output.

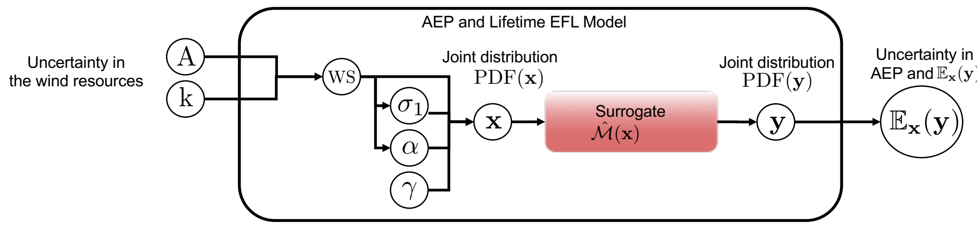


Figure 6: 2 levels of propagation of uncertainty.

The distribution of the variability of the wind resources is presented in table 6. The main difference with the distribution used for training the surrogates is the fact that the WS follows a Weibull distribution with uncertain shape and scale parameters. This distribution of the Weibull parameters is used to characterize the variability/uncertainty in the wind resources. Nevertheless, the conditional distributions of σ_1, α and γ with respect WS follow the same dependency described in table 2.

Variable	Distribution	Parameters	
A	Normal	$\mu_A = 9$	$\sigma_A = 0.5$ m/s
k	Normal	$\mu_k = 2$	$\sigma_k = 0.1$
$x_0 = \mathrm{WS}$	Weibull	scale= A	shape= k
$x_1 = \sigma_1$	Lognormal	$\mu_{\sigma_1}(\mathrm{WS})$	$\sigma_{\sigma_1}(\mathrm{WS})$
$x_2 = \alpha$	Normal	$\mu_\alpha(\mathrm{WS})$	$\sigma_\alpha(\mathrm{WS})$
$x_3 = \gamma$	Normal	$\mu_\gamma = 0$	$\sigma_\gamma = 5$ deg.

Table 6: Uncertainty in wind resources.

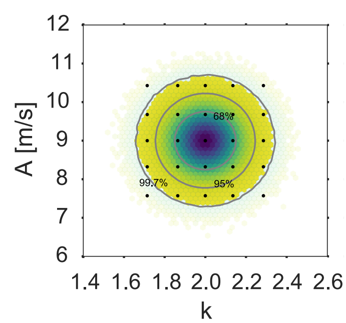


Figure 7: Joint distribution of the Weibull parameters and semi-spectral projection nodes for outer level propagation of uncertainty.

The propagation of uncertainty in the outer level is done using both a 1000 MC sample and a PCE with semi-spectral projection, for which a total of 25 Weibull parameters nodes are evaluated with their corresponding Gaussian quadrature weights, see figure 7 and equation 3. Each node or element of the outer level MC sample represents a realization of the wind resources in a given year. For each of these nodes, a large inner level sample of the inputs of the surrogate, $\mathbf{x} = [\mathrm{WS}, \sigma_1, \alpha, \gamma]$, is generated. The size of the inner level MC sample is the number of 10-min cases in a year, $365 \times 24 \times 6 = 52,560$ cases. The power and EFL are evaluated using the surrogate and the mean power and mean EFL for a given year are calculated $\mathbb{E}_{\mathbf{x}}(\mathbf{y})$. Note that the definition of the lifetime damage equivalent fatigue load (see eq. 13) requires to take the average of the individual 10-min EFL to the Wöhler exponent, which translates in taking a higher order statistical moment: $\mathbb{E}_{\mathbf{x}}(\mathbf{y}^m)$. Each individual surrogate evaluation has its own realization of the local distribution of the outputs due to the turbulence inflow realization, see equation 1. Additionally, the effect of the errors of the surrogate are considered, by sampling the distribution of the errors for each evaluation of the outputs, see equation 15. There are no differences in the distributions of $\mathbb{E}_{\mathbf{x}}(\mathbf{y})$ obtained using the surrogate or the ones obtained including the uncertainty of the surrogate due to the large sample size of the inner level $(52, 560)$; this means that the errors of the surrogate cancel out when computing their mean on a given year.

A 1000 MC sample of the distribution of one year $\mathbb{E}_{\mathbf{x}}(\mathbf{y})$ is generated using the PCE of the outer level in order to have an equivalent database of 1000 years as the one obtained in the outer MC simulation. A bootstrap of the outer level sample is used to estimate the variation in the expected value during 20 years of operation. This means that the average of 20 randomly selected years is computed for several realizations of 20 years. The central limit theorem is also used to estimate the distribution of the average of 20 randomly selected (independent) years. The distributions of the 1 year and 20 years capacity factor and of lifetime equivalent fatigue loads are presented in figure 8. It can be observed how the 20-year-averaged distribution has a narrower distribution, $\sigma_{20yr} = \sigma_{1yr}/\sqrt{20}$. Note that the yearly distribution of average output is required in order to estimate the uncertainty in the 20-year-averaged output. In this example coefficient of variations (CoV $= \sigma/\mu$) of 5.6% for the scale parameter (A) and 5.0% for the shape parameter (k) of the WS Weibull distribution give a coefficient of variation of 2.4% in AEP and a 9.5% in year-to-year expected power production. The coefficient of variation in the 20-year damage equivalent BRF is 8.0% while the CoV of the year-to-year damage equivalent BRF is 35.0%. The CoV for the TBF are 1.0% for the 20-year damage equivalent load and 4.0% for the year-to-year variation. Note that this coefficients of variations will be modified if the correlation between the WS and the other turbulent inflow parameters changes from year to year. It is important to realize that the distribution for the year-year equivalent damage BRF is skewed due to the large Wöhler exponent of the composite blades used in this study (12). Nevertheless, the lifetime equivalent damage BRF converges to a Normal distribution which can be estimated from the mean and variance of the PCE of the yearly distribution.

4. Discussion

The present article presents a methodology to implement sparse polynomial surrogates for aeroelastic wind turbine models. PCE are widely used in the uncertainty quantification field due to their efficiency to compute the statistical properties of the output and because the sensitivity analysis is obtained without any additional effort. The main two limitations in the use of PCE for wind energy are: (1) The input atmospheric parameters are usually jointly distributed with several layers of dependency (2) Some of the output have discontinuities and/or are restricted to certain values (e.g.

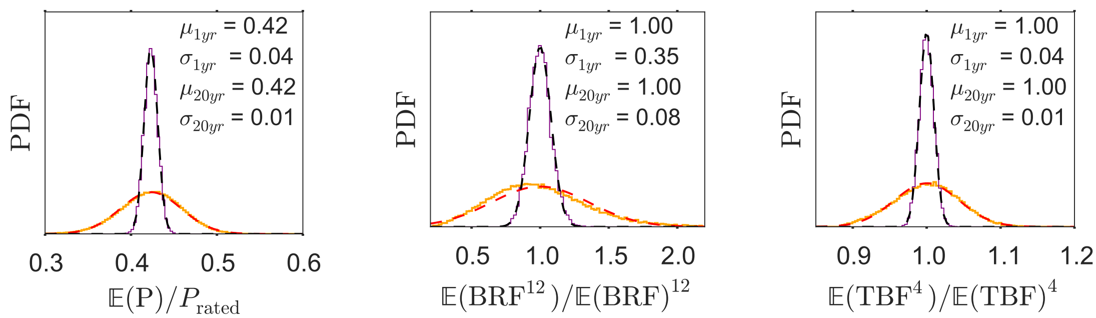


Figure 8: Distribution of the capacity factor and of the expected BRF and TBF equivalent loads. (Red) Normal distribution with the mean and variance predicted with the PCE distribution of the 1 year expected output. (Orange) 1000 MC sample of the 1 year expected output. (Black) Central limit distribution of 20-year-averaged output. (Purple) 1000 Bootstraps of the 20-year-averaged output.

only positive). The present article has shown how to solve these two problems: the implementation of an iso-probabilistic transformation to de-correlate the inputs, and the use of a logistic transformation to implement restrictions on the outputs. The benefits of using the logistic transformation can be seen in figure 3, note that the polynomial surrogates do not present oscillations in the constant regions.

The final surrogate can be used to generate an output sample that covers the full output space, and that will predict the general details of the distributions of the outputs. One of the main limitations of the present surrogates is that the local distribution of the output is assumed to be normal, this is not the case for the operating region close to rated wind speed. Since this assumption only affects the turbulent inflow realization, it is considered to be an acceptable approximation. The local distributions of most outputs are not normal in reality, because the wind turbine controller has different strategies in each operating region, which creates skewness in the local distributions.

The results presented in this article show that there are multiple dependencies between the input variables as well as between inputs and outputs. Such complicated inter-dependencies are difficult to capture when applying other methods such as interpolation or Gaussian processes. For example, advanced interpolation methods such as radial basis functions will not account for the likelihood of an extreme training point and will generate trends that always pass through all the model observations. This behavior penalizes the capacity of the surrogate to generalize and to predict the output in new conditions. The sparse PCE are ideal for this class of problems because the k-fold cross validation is a step inside the training. Additionally, the correlations between the outputs are fully captured when using the presented surrogates; this occurs because each of the outputs has a dependency on the inputs. The full pair plot of the training dataset and the resulting surrogate for all inputs and outputs is presented in the extra material accompanying this article.

The final results presented in figures 4 and 8 show a promising new approach to communicate the performance characteristics of a wind turbine between the turbine manufacturers and project developers. The wind turbine producers normally do not share the detailed structural and aerodynamic model information of their products due to intellectual property concerns. As a result, often the wind project planners and operators do not have the full information about the expected performance of a turbine at the site they are developing. Furthermore, typically there is no model for the uncertainty of the turbine performance. A possible application of the multiple polynomial surrogates of a wind turbine could involve fitting the model by the manufacturer, and consequent distribution of the surrogate to users and clients. With this approach, project developers could get a useful tool for assessing site feasibility including uncertainty estimation, while not requiring access to detailed engineering models. Consequently, the use of more refined site assessment can potentially lead to improved overall estimation of levelized cost of energy and its uncertainty.

Obtaining the PDF(P) and PDF(EFL) is useful as they can be used for uncertainty estimation of the levelized cost of energy on a yearly basis. The surrogates can be evaluated on a long time series of the local wind resources (in multiple variables) such as the ones predicted by Weather Research and Forecasting (WRF) models without considerable extra computational effort. The power surrogate can then predict the annual variation of energy production while the EFL can be used to estimate the operation and maintenance costs. Such a probabilistic output can be the input to a decision support tool.

A surrogate of the DTU 10 MW RWT within a 4-dimensional turbulent inflow parameter space can be built using only 140 input cases (with multiple turbulent inflow realizations per case) and can be used to predict the distribution of the power, thrust coefficient and equivalent fatigue loads on the turbine. In contrast, traditional approaches require in the order of 20^4 gridsearch/interpolation (full factorial design with 20 points per dimension) or $10^5 - 10^6$ MC sample of the inputs with variance reduction [22]. Furthermore, the present approach enables to build an uncertainty model around the 10 minutes performance of the turbine that captures the effect of the turbulent inflow realization.

The combined PCE surrogate approach can also be used to improve traditional designs in which a conservative scenario for shear and turbulence intensity is considered. The fast evaluation of the joint probability distributions for loads based on the surrogate model opens possibilities for performing structural reliability analysis and probability based design.

5. Conclusions

In the present study, a polynomial surrogate model of wind turbine fatigue loads and energy output was defined and demonstrated for the DTU 10 MW reference wind turbine. Using only 140 input cases was found to be sufficient for building a surrogate of the DTU 10MW model within a 4-dimensional turbulent inflow parameter space. The presented approach was demonstrated as an efficient alternative of the traditional techniques for characterizing the global behavior of an aeroelastic wind turbine model under multiple uncertain turbulent inflow parameters.

The surrogate has enabled us to perform a global sensitivity analysis on the DTU 10 MW turbine. This study showed that the hub height wind speed is the most important variable to predict the power of the turbine, followed by the turbulent inflow realization (TIR); this is a consequence of the correlation between turbulence intensity, shear and hub height wind speed. The turbulence intensity is of similar importance as the TIR in the prediction of blade root flapwise (BRF), and tower top tilt (TTT) and yaw (TTY) equivalent fatigue loads.

The surrogate can be used in a two-level propagation of uncertainty example. In the example presented in this article the year-to-year variability in the shape and scale parameters of the hub height wind speed Weibull distribution are propagated into a variation of AEP and of lifetime equivalent fatigue loads. Coefficient of variations of 5.6% for the scale and of 5% for the shape parameters give a coefficient of variation of 2.4% in AEP, of 1.8% in lifetime $\mathbb{E}(BRF)$ and of 0.5% in lifetime $\mathbb{E}(TBF)$.

Finally, the methodology presented in this article can be used in other applications in which there are fields which might take multiple realizations such as marine structures (wave and current fields), offshore structures (wave and wind fields) or soil-foundation structures (soil properties fields) among others.

6. Acknowledgments

This work was supported by the International Collaborative Energy Technology R&D Program of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), granted financial resource from the Ministry of Trade, Industry & Energy, Republic of Korea. (No. 20138520021140). The authors will like to thank Michael McWilliam for the suggestion of the use of logistic transformations to enforce strict restrictions to the polynomial surrogates.

References

[1] I. E. Commission, et al., Iec 61400-1: Wind turbines part 1: Design requirements, International Electrotechnical Commission.

[2] P. A. Graf, G. Stewart, M. Lackner, K. Dykes, P. Veers, High-throughput computation and the applicability of monte carlo integration in fatigue load estimation of floating offshore wind turbines, Wind Energy.

[3] H. S. Toft, L. Svenningsen, W. Moser, J. D. Sørensen, M. L. Thøgersen, Assessment of wind turbine structural integrity using response surface methodology, Engineering Structures 106 (2016) 471–483.

[4] I. Abdallah, A. Natarajan, J. D. Sørensen, Influence of the control system on wind turbine loads during power production in extreme turbulence: Structural reliability, Renewable Energy 87 (2016) 464–477.

[5] A. Clifton, L. Kilcher, J. Lundquist, P. Fleming, Using machine learning to predict wind turbine power output, Environmental research letters 8 (2) (2013) 024009.

[6] A. Clifton, M. Daniels, M. Lehning, Effect of winds in a mountain pass on turbine performance, Wind Energy 17 (10) (2014) 1543–1562.

[7] C. Soize, R. Ghanem, Physical systems with random uncertainties: chaos representations with arbitrary probability measure, SIAM Journal on Scientific Computing 26 (2) (2004) 395–410.

[8] O. Le Maıtre, O. Knio, H. Najm, R. Ghanem, Uncertainty propagation using wiener–haar expansions, Journal of computational Physics 197 (1) (2004) 28–57.

[9] S.-K. Choi, R. V. Grandhi, R. A. Canfield, C. L. Pettit, Polynomial chaos expansion with latin hypercube sampling for estimating response variability, AIAA journal 42 (6) (2004) 1191–1198.

[10] M. Berveiller, B. Sudret, M. Lemaire, Stochastic finite element: a non intrusive approach by regression, European Journal of Computational Mechanics/Revue Européenne de Mécanique Numérique 15 (1-3) (2006) 81–92.

[11] D. Xiu, J. S. Hesthaven, High-order collocation methods for differential equations with random inputs, SIAM Journal on Scientific Computing 27 (3) (2005) 1118–1139.

[12] G. Blatman, B. Sudret, Adaptive sparse polynomial chaos expansion based on least angle regression, Journal of Computational Physics 230 (6) (2011) 2345–2367.

[13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., Scikit-learn: Machine learning in python, The Journal of Machine Learning Research 12 (2011) 2825–2830.

[14] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society. Series B (Methodological) (1996) 267–288.

[15] P. Moriarty, Database for validation of design load extrapolation techniques, Wind Energy 11 (6) (2008) 559.

[16] A. Natarajan, D. R. Verelst, Outlier robustness for wind turbine extrapolated extreme loads, Wind Energy 15 (5) (2012) 679–697.

[17] C. Tibaldi, L. C. Henriksen, C. Bak, Investigation of the dependency of wind turbine loads on the simulation time, in: European Wind Energy Conference & Exhibition 2014, 2014.

[18] C. Bak, R. Bitsche, A. Yde, T. Kim, M. H. Hansen, F. Zahle, M. Gaunaa, J. P. A. A. Blasques, M. Døssing, J.-J. Wedel Heinen, et al., Light rotor: The 10-mw reference wind turbine, in: EWEA 2012-European Wind Energy Conference & Exhibition, 2012.

[19] M. Rosenblatt, Remarks on a multivariate transformation, The annals of mathematical statistics 23 (3) (1952) 470–472.

[20] P. Y. Simard, Y. A. LeCun, J. S. Denker, B. Victorri, Transformation invariance in pattern recognitiontangent distance and tangent propagation, in: Neural networks: tricks of the trade, Springer, 1998, pp. 239–274.

[21] W. Gautschi, Algorithm 726: Orthpol–a package of routines for generating orthogonal polynomials and gauss-type quadrature rules, ACM Transactions on Mathematical Software (TOMS) 20 (1) (1994) 21–62.

[22] N. Dimitrov, A. Natarajan, M. Kelly, Model of wind shear conditional on turbulence and its impact on wind turbine loads, Wind Energy 18 (11) (2015) 1917–1931.

[23] J. Feinberg, H. P. Langtangen, Chaospy: An open source tool for designing methods of uncertainty quantification, Journal of Computational Science 11 (2015) 46–57.

[24] M. D. McKay, R. J. Beckman, W. J. Conover, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics 42 (1) (2000) 55–61.

[25] I. M. Sobol', On the distribution of points in a cube and the approximate evaluation of integrals, Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki 7 (4) (1967) 784–802.

[26] J. M. Hammersley, Monte carlo methods for solving multivariable problems, Annals of the New York Academy of Sciences 86 (3) (1960) 844–874.

[27] I. M. Sobol, Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates, Mathematics and computers in simulation 55 (1) (2001) 271–280.

[28] S. E. Maxwell, H. D. Delaney, Designing experiments and analyzing data: A model comparison perspective, Vol. 1, Psychology Press, 2004.

[29] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index, Computer Physics Communications 181 (2) (2010) 259–270.

[30] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, Reliability Engineering & System Safety 93 (7) (2008) 964–979.

[31] M. S. Eldred, A. A. Giunta, B. G. van Bloemen Waanders, S. F. Wojtkiewicz, W. E. Hart, M. P. Alleva, DAKOTA, a multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis: Version 4.1 reference manual, Citeseer, 2007.

[32] S. Marelli, B. Sudret, UQLab: a framework for uncertainty quantification in MATLAB, ETH-Zürich, 2014.

[33] G. Andrianov, S. Burriel, S. Cambier, A. Dutfoy, I. Dutka-Malen, E. De Rocquigny, B. Sudret, P. Benjamin, R. Lebrun, F. Mangeant, et al., Open turns, an open source initiative to treat uncertainties, risks n statistics in a structured industrial approach, in: Proceedings of ESREL, 2007.

[34] T. J. Larsen, A. M. Hansen, How 2 HAWC2, the user's manual, Risø National Laboratory, 2007.

[35] J. Mann, Wind field simulation, Probabilistic Engineering Mechanics 13 (4) (1998) 269 – 282.

[36] M. A. Miner, et al., Cumulative damage in fatigue, Journal of applied mechanics 12 (3) (1945) 159–164.

[37] S. Hosder, R. W. Walters, M. Balch, Efficient sampling for non-intrusive polynomial chaos applications with multiple uncertain input variables, in: Proceedings of the 48th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, AIAA paper, Vol. 1939, 2007.

Highlights

- Sparse polynomials are proposed as surrogates of an aeroelastic wind turbine model.
- The surrogates can be used to predict the distribution of the 10-min mean power and equivalent fatigue loads under realistic atmospheric conditions.
- The surrogates are used in a two-level uncertainty propagation scenario to estimate the uncertainty in annual energy production and in lifetime equivalent fatigue loads.