Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

From wind to loads: wind turbine site-specific load estimation using databases with high-fidelity load simulations

Nikolay Dimitrov, Mark Kelly, Andrea Vignaroli, and Jacob Berg

DTU Wind Energy, Technical University of Denmark
Frederiksborgvej 399, 4000 Roskilde, Denmark

Correspondence to: Nikolay Dimitrov (nkdi@dtu.dk)

Abstract. We define and demonstrate a procedure for quick assessment of site-specific lifetime fatigue loads, using surrogate models calibrated by means of a database with high-fidelity load simulations. The performance of six surrogate models is assessed by comparing site-specific lifetime fatigue load predictions at ten sites. The surrogate methods include polynomial-chaos expansion, quadratic response surface, universal Kriging, importance sampling, and nearest-neighbor interpolation. Practical bounds for the database and calibration are defined via nine environmental variables, and their relative effects on the fatigue loads are evaluated by means of Sobol sensitivity indices.

Of the surrogate-model methods, polynomial-chaos expansion provided an accurate and robust performance in prediction of the different site-specific loads. Although the Kriging approach showed slightly better accuracy, it also demanded more computational resources. Taking into account other useful properties of the polynomial chaos expansion method within the performance comparisons, we consider it to generally be the most useful for quick assessment of site-specific loads.

Copyright statement. This article is subject to copyright.

1 Introduction

Before installing a wind turbine at a particular site, it needs to be ensured that the wind turbine structure is sufficiently robust to withstand the environmentally-induced loads during its entire lifetime. As the design of serially-produced wind turbines is typically based on a specific set of wind conditions, i.e. a site class defined in the IEC (2005) standard, any site where the conditions are more benign than the reference conditions is considered feasible. However, often one or more site-specific parameters will be outside this envelope—and disqualify the site as infeasible, unless it is shown that the design load limits are not going to be violated under site-specific conditions. Such a demonstration requires carrying out simulations over a full design load basis, which adds a significant burden to the site assessment process.

Various methods and procedures have been attempted for simplified load assessment such as expansions based on statistical moments, (Kashef and Winterstein, 1999; Manuel et al., 2001), multivariate regression models of first order (Mouzakis et al., 1999) and second order (Toft et al., 2016), and expansions using orthogonal polynomial basis (Murcia et al., 2016).

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

In the present work, we analyze, refine and expand the existing simplified load assessment methods, and provide a structured approach for practical implementation of a high-fidelity load database for site feasibility assessment. The study aims at fulfilling the following three specific goals:

- define a simplified load assessment procedure which can take into account all the relevant external parameters required for full characterization of the wind fields used in load simulations;
- define feasible ranges of variation of the wind-related parameters, dependent on wind turbine rotor size; and
- obtain estimates of the statistical uncertainty and parameter sensitivities.

The scope of the present study is loads generated under normal power production, which encompasses design load cases (DLC) 1.2 and 1.3 from the IEC 61400-1 standard (IEC, 2005). These load cases are the main contributors to the fatigue limit state (DLC1.2) and often the blade extreme design loads (DLC1.3). The methodology used can easily be applied to other load cases governed by wind conditions with a probabilistic description. Loads generated during fault conditions (e.g. grid drops) or under deterministic wind conditions (e.g. operational gusts without turbulence) will in general not be (wind climate) site-specific.

2 High-fidelity loads database

2.1 Definition of variable space

The turbulent wind field serving as input to aeroelastic load simulations can be fully characterized statistically by the following variables:

- mean wind field across the rotor plane as described by the
  - average wind speed at hub height (U),
  - vertical wind shear exponent (α),
  - wind veer (change of mean flow direction with height, $\Delta\varphi$);
- turbulence described via
  - variance of wind fluctuations, σ_u^2,
  - turbulence probability density function (e.g. Gaussian),
  - turbulence spectrum defined by the Mann (1994) model with parameters
    - turbulence length scale L,
    - anisotropy factor Γ,
    - turbulence dissipation parameter $\alpha\varepsilon^{2/3}$;

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

- air density ρ;
- mean wind inflow direction relative to the turbine in terms of
  - vertical inflow (tilt) angle $\bar{\varphi}_v$ and
  - horizontal inflow (yaw) angle $\bar{\varphi}_h$.

All of the parameters above, except the parameters defining mean inflow direction, are probabilistic and site-dependent in nature. The mean inflow direction parameters represent a combination of deterministic factors (i.e. terrain inclination or yaw direction bias in the turbine) and random fluctuations due to e.g. large-scale turbulence structures or variations in atmospheric stability. Mean wind speed, turbulence and wind shear are well known to affect loads and are considered in the IEC61400-1 standard. In Kelly et al. (2014) a conditional relation describing the joint probability of wind speed, turbulence and wind shear was defined. The effect of implementing this wind shear distribution in load simulations was assessed in Dimitrov et al. (2015), showing that wind shear has importance especially for blade deflection. The Mann model parameters L and Γ were also shown to have a noticeable influence on wind turbine loads (Dimitrov et al., 2017). By definition, for given L and Γ the $\alpha\varepsilon^{2/3}$ parameter from the Mann model is directly proportional to $\sigma_u^2 L^{-2/3}$ (Mann, 1994; Kelly, 2018), and can therefore be omitted from the analysis. The probability density function (pdf) typically used to synthesize time series of velocity components from Mann-model spectra is Gaussian. For a slightly smaller turbine, the NREL 5MW turbine, the assumption of Gaussian turbulence has been shown to not impact the fatigue loads (Berg et al., 2016). The final list of inflow-related parameters thus reads

$$\{U, \sigma_u, \alpha, L, \Delta\varphi, \Gamma, \bar{\varphi}_h, \bar{\varphi}_v, \rho\}.$$

The loads experienced by a wind turbine are a function of the wind-derived factors described above, and of the structural properties and control system of the wind turbine. Therefore, a load characterization database taking only wind-related factors into account is going to be turbine-specific.

The variables describing the wind inflow often have a significant correlation between them, and any site-specific load or power assessment has to take this into account using an appropriate description of the joint distribution of input variables. At the same time, most probabilistic models require inputs in terms of a set of independent and identically distributed (i.i.d) variables. The mapping from the space of i.i.d variables to joint distribution of physical variables requires applying an isoprobabilistic transformation as e.g. the Nataf transform (Liu and Der Kiureghian, 1986), and the Rosenblatt transformation (Rosenblatt, 1952). In the present case, it is most convenient to apply the Rosenblatt transformation, which maps a vector of n dependent

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

variables $\mathbf{X}$ into a vector of independent components $\mathbf{Y}$ based on conditional relations:

$$\mathbf{X} \rightarrow \mathbf{Y} = \begin{pmatrix} F_1(X_1) \\ \vdots \\ F_{k|1,\ldots,k-1}(X_k|X_1,\ldots,X_{k-1}) \\ \vdots \\ F_{n|1,\ldots,n-1}(X_n|X_1,\ldots,X_{n-1}) \end{pmatrix}. \quad (1)$$

Further mapping of $\mathbf{Y}$ to a standard Normal space vector $\mathbf{U}$ is sometimes applied, i.e.

$$\mathbf{Y} \rightarrow \mathbf{U} = \begin{pmatrix} \Phi^{-1}(Y_1) \\ \ldots \\ \Phi^{-1}(Y_n) \end{pmatrix}. \quad (2)$$

For the currently considered set of variables, the Rosenblatt transformation can be applied in the order defined in Table 1 - i.e., the wind speed is considered independent of other variables, the turbulence is dependent on the wind speed, the wind shear is conditional on both wind speed and turbulence, etc. For any variable in the sequence, it is not necessary that it is dependent on all higher-order variables (it may only be conditional on a few of them or even none), but it is required that it is independent from lower-order variables.

2.2 Defining the ranges of input variables

The choice for ranges of variation of the input variables needs to ensure a balance between two objectives: a) covering as wide a range of potential sites as possible, while b) ensuring that the load simulations produce valid results. To ensure validity of load simulations, the major assumptions behind the generation of the wind field and computation of aerodynamic forces should not be violated, and the instantaneous wind field should have physically meaningful values.

The turbulence intensity, $I_u = \sigma_u/U$, upper limit can be written as the IEC-prescribed form (ed. 3, sub-class A) with $I_{\mathrm{ref,A}}$=18%, plus a constant (representing the larger expected range of TI, to span different sites) and a term that encompasses low-windspeed sites and regimes which have higher turbulent intensities. This form is basically equivalent to $\sigma_{u,\mathrm{IEC}} + I_{\mathrm{ref,A}} U_{\text{cut-in}}[1 + (U_{\text{cut-out}}/U)]$ with $\{U_{\text{cut-in}}, U_{\text{cut-out}}\}$={3,25}m/s. The bounds for turbulence intensity as function of mean wind speed are shown on Figure 2. The limits on shear exponent were chosen following the derivations and findings of Kelly et al. (2014) for $P(\alpha|U)$, expanding on the established $\sigma_\alpha(U)$ form to allow for a reasonably wide and inclusive range of expected cases, and also accounting for rotor size per height above ground. This includes an upper bound which allows for enhanced shear due e.g. to lower-level jets and terrain-induced shear; the lower bound also includes the R/z dependence, but does not expand the space to the point that it includes jet-induced negative shear (these are generally found only in the top portion of the rotor). The condition $L > \max\{7.5\mathrm{m}, (15\mathrm{m})|\alpha|^{-2/3}\}$ arises from consideration of the relationship between L, α, σ_u, and ε; small shear tends to correlate with larger motions (as in convective well-mixed conditions). The minimum scale (7.5 m) and proportionality constant (15 m) are set to allow a wide range of conditions (though most sites will likely have a scaling factor

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Table 1. Bounds of variation of the variables considered. All values are defined as statistics over 10-minute reference period.

Variable	Lower bounds	Upper bounds
U	$U \geq 4\text{m/s}$	$U \leq 25\text{m/s}$
σ_u	$\sigma_u \geq 0.025 \cdot U(\text{m/s})$	$\sigma_u \leq 0.18\left(6.8+0.75U+3\left(\frac{10}{U}\right)^2\right)(\text{m/s})$
α	$\alpha \geq \alpha_{ref,LB} - 0.23\left(\frac{U_{max}}{U}\right)\left(1-\left(0.4\log\frac{R}{z}\right)^2\right)$	$\alpha \leq \alpha_{ref,UB} + 0.4\left(\frac{R}{z}\right)\left(\frac{U_{max}}{U}\right)$
L	$L \geq \max\{7.5\text{m} \quad (15\text{m})\cdot\lvert\alpha\rvert^{-2/3}\}$	$L \leq 275\text{m}$
Γ	$\Gamma \geq 1$	$\Gamma \leq 5$
$\Delta\varphi_h$	$\Delta\varphi_h \geq -0.1D\left(\frac{5}{U}\right)$	$\Delta\varphi_h \leq \min\{60^\circ \sin\lvert\phi\rvert, \quad 1.0D\left(\frac{5}{U}\right)^2\}$
$\bar{\varphi}_h$	$\bar{\varphi}_h \geq -10^\circ$	$\bar{\varphi}_h \leq 10^\circ$
$\bar{\varphi}_v$	$\bar{\varphi}_v \geq -10^\circ$	$\bar{\varphi}_v \leq 10^\circ$
ρ	$\rho \geq 1.1\text{kg/m}^3$	$\rho \leq 1.35\text{kg/m}^3$
Where - R is the rotor radius, D the rotor diameter; - $\alpha_{ref,LB} = 0.15, \alpha_{ref,UB} = 0.22$ are reference wind shear exponents at 15m/s wind speed; - $U_{max} = 25\text{m/s}$ is the upper bound of the wind speed; - ϕ is the reference latitude (here chosen as 50°).		

larger than 15 m). The maximum Mann-model length scale is chosen based on the limits of where the model can be fitted to measured spectra; this is dictated also by the limits of stationarity in the atmospheric boundary layer. The range of Γ is also dictated by the minimum expected over non-complex terrain within reasonable use of the turbulence model (smaller Γ might occur for spectra fitted at low heights over hills, but such spectra should be modelled in a different way, as in e.g. Mann (2000)). The range of veer is limited in a way analogous to shear exponent, i.e. it has a basic $1/U$ dependence; this range also depends upon the rotor size, just as $(dU/dz)|_{\text{rotor}} = \alpha D/U$ (Kelly and van der Laan, 2018). The limits for $\Delta\varphi_h$ above peak follow from the limits on α, while for unstable conditions ($\Delta\varphi_h < \Delta\varphi_{h,\text{peak}}$, e.g. all $\Delta\varphi_h < 0$) then the veer limit follows a semi-empirical form based on observed extremes of $\partial\varphi_h/\partial z$.

2.3 Reference high-fidelity load database

Building a large database with high-fidelity load simulations covering the entire variable space is a central task in the present study as such a database can serve several purposes:

1) be directly used as a site assessment tool by probability-weighting the relative contribution of each point to the design loads;

2) serve as an input for calibrating simplified models such as orthogonal-polynomial based expansions;

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

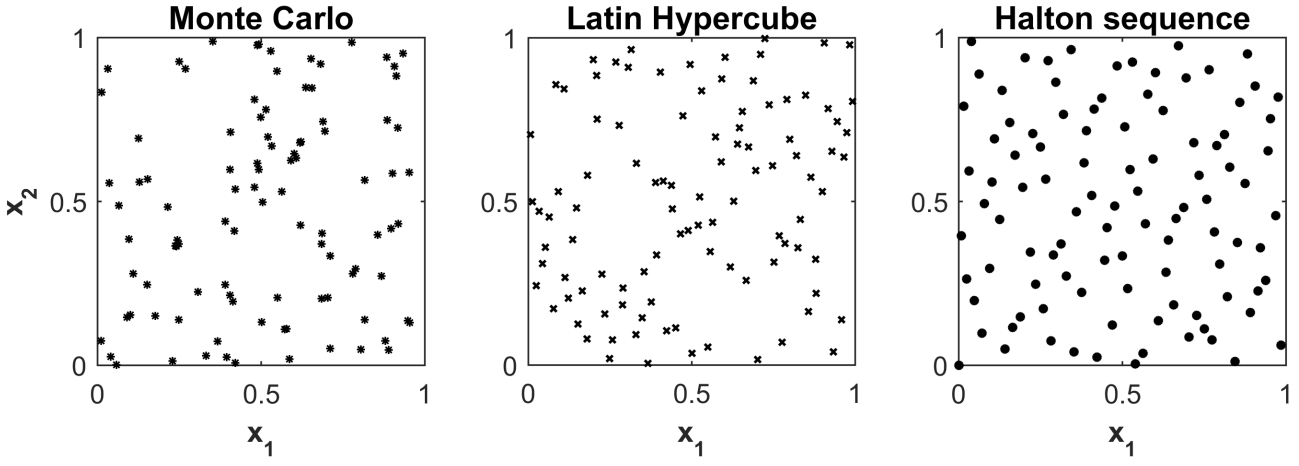

Figure 1. Comparison of several simulation-based experimental design approaches. Examples show random (Monte Carlo and Latin Hypercube) or pseudorandom (Halton sequence) samples of size 100 drawn from a uniform distribution within a unit hypercube.

3) be used as reference for the performance of load models calibrated by other means.

Characterizing the load behaviour of a wind turbine over a range of input conditions requires an experimental design covering the range of variation of all variables with sufficient resolution. In the case of having more than 3-4 dimensions, a full factorial design with multiple levels quickly becomes impractical due to the exponential increase in the number of design points as function of number of dimensions. Therefore, in the present study we resort to a Monte Carlo (MC) simulation as the main approach for covering the joint distribution of wind conditions. For assuring better and faster convergence, we use the low-discrepancy Halton sequence in a quasi-Monte Carlo approach (Caflisch, 1998). Figure 1 shows an experimental design based on Halton sequence, compared to crude Monte Carlo and Latin Hypercube designs (Mckay et al., 2000). While a crude Monte Carlo integration has a convergence rate proportional to the square root of the number of samples N, i.e., the mean error $\bar{\varepsilon} \propto N^{-0.5}$, the convergence rate for a pseudo-Monte Carlo with a low-discrepancy sequence results in $\bar{\varepsilon} \propto N^{-\lambda}$, $0.5 \leq \lambda \leq 1$. For low number of dimensions and smooth functions, the pseudo-Monte Carlo sequences show a significantly improved performance over the Monte Carlo, e.g. $\lambda \to 1$, however for multiple dimensions and discontinuous functions the advantage over crude Monte Carlo is reduced (Morokoff and Caflisch, 1995). Nevertheless, even for the full 9-dimensional problem discussed here, it is expected that $\lambda \approx 0.6$ (Morokoff and Caflisch, 1995), which still means about an order of magnitude advantage, e.g., 10^4 pseudo-Monte Carlo samples should result in about the same error as 10^5 crude Monte Carlo samples.

2.4 Database specification

A large-scale generic load database is generated using Halton low-discrepancy sample points within the 9-dimensional variable space defined in section 2.3 (Figure 2 shows the bounds for the first 6 variables). The database setup is the following:

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

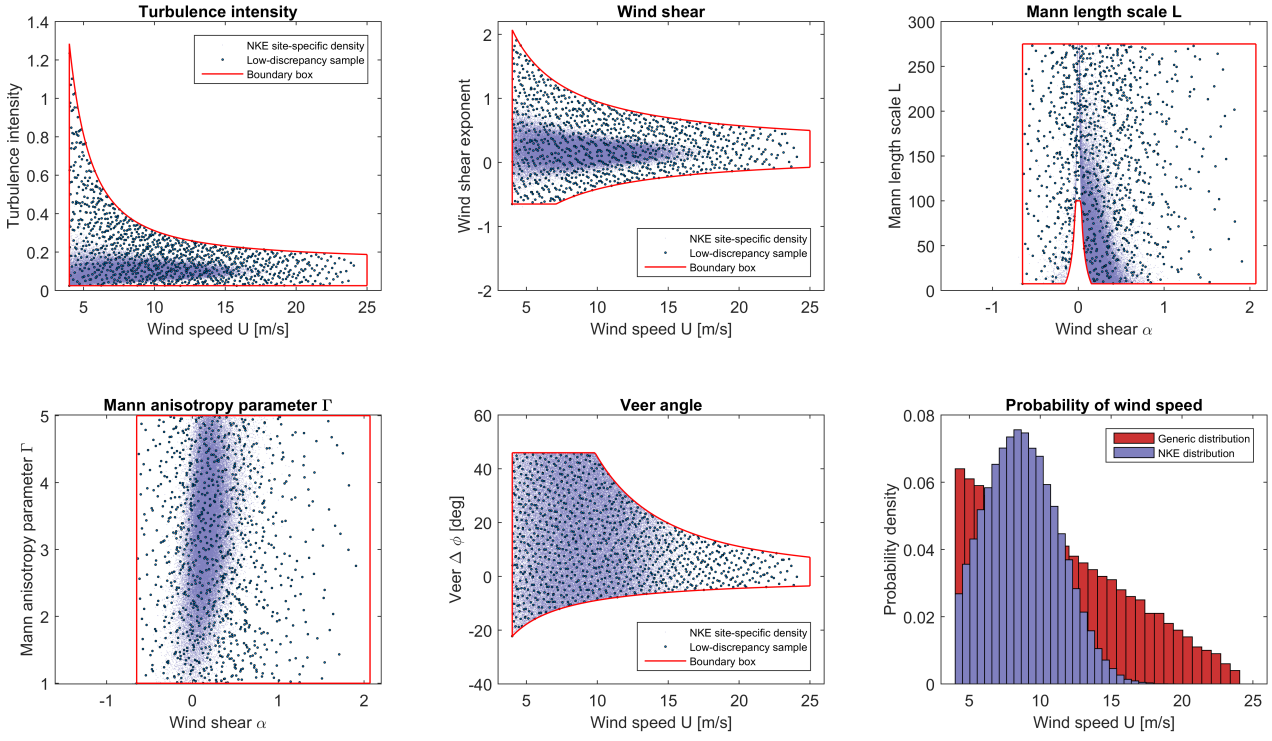

Figure 2. Sample distributions obtained using 1024 low-discrepancy points within a 6-dimensional variable space $\{U, I_u, \alpha, \Delta\phi_h, L, \Gamma\}$. Here U is Beta-distributed, while the other variables are uniformly distributed within their ranges. Blue shading shows the site-specific variable distribution for the NKE reference site (site 0, c.f. Table 2/Section 3).

- Up to 10^4 pseudo-random MC sample points are generated, following a low-discrepancy sequence for obtaining evenly distributed points within the parametric space.
- For each sample point, eight simulations, with 1h duration each, are carried out.
- The simulation parameters are tuned to match the required 10-minute statistics (1 h statistics are slightly different due to longer sampling time).
- Each 1h time-series can be split into six 10-minute series, which on average will have the required statistics.
- Simulation conditions are kept stationary over each 1 h simulation period.
- The DTU 10MW reference wind turbine model (Bak et al., 2013), with the basic DTU Wind Energy controller (Hansen and Henriksen, 2013), is used in the HAWC2 aeroelastic software (Larsen and Hansen, 2012).

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

2.5 Postprocessing and analysis

2.5.1 Postprocessing

The main quantities of interest from the load simulation output are the short-term fatigue damage-equivalent loads (DEL), and the 10-minute extremes (minimum or maximum, depending on the load type). For each load simulation, four statistics (mean, standard deviation, minimum and maximum values) are calculated for each load channel. For several selected load channels, the 1 Hz DEL for a reference period T_{ref} are estimated using the expression

$$S_{\text{eq}} = \left[\sum \frac{n_i S_i^m}{N_{\text{ref}}} \right]^{1/m} \tag{3}$$

where $N_{\text{ref}} = f \cdot T_{\text{ref}}$ is a reference number of cycles ($N_{\text{ref}} = 3600$ for obtaining 1 Hz-equivalent DEL over a 1 h period), S_i are load range cycles estimated using a rainflow counting algorithm (Rychlik, 1987), and n_i are the number of cycles observed in a given range. For a specific material with fatigue properties characterized by an S-N curve of the form $K = N \cdot S^m$ (where K is the material-specific Wöhler constant), the fatigue damage D accumulated over one reference period equals

$$D(T_{\text{ref}}) = \frac{N_{\text{ref}}}{K} S_{\text{eq}}^m. \tag{4}$$

2.5.2 Definition of lifetime damage-equivalent loads

Obtaining site-specific lifetime fatigue loads from a discrete set of simulations requires integrating the short-term damage contributions over the long-term joint distribution of input conditions. The lifetime damage-equivalent fatigue load is defined as

$$S_{\text{eq,lifetime}} = \left[\int_{\mathbf{X} \in R^9} [S_{\text{eq}}(\mathbf{X})]^m f(\mathbf{X}) d\mathbf{X} \right]^{1/m} \tag{5}$$

where $f(\mathbf{X})$ is the joint distribution of the multidimensional vector of input variables $\mathbf{X}$. With the above definition, $S_{\text{eq,lifetime}}$ is a function of the expected value of the short-term equivalent loads conditional on the distribution of environmental variables. The integration in (5) is typically performed numerically over a finite number of realizations drawn from the joint distribution of the input variables, e.g. by setting up a look-up table or carrying out a Monte Carlo simulation. Thus the continuous problem is transformed into a discrete one:

$$S_{\text{eq,lifetime}} = \left[\sum_{i=1}^{N} [S_{\text{eq}}(\mathbf{X}_i)]^m p(\mathbf{X}_i) \right]^{1/m}, \tag{6}$$

where $\mathbf{X}_i, i = 1 \ldots N$, is the i^{th} realization of $\mathbf{X}$ out of N total realizations, and $p(\mathbf{X}_i)$ is the relative, discretized probability of $\mathbf{X}_i$, which is derived by weighting the joint pdf values of $\mathbf{X}$ so that they satisfy the condition $\sum_{i=1}^{N} p(\mathbf{X}_i) = 1$. For a standard Monte Carlo simulation, each realization is considered to be equally likely, and $p(\mathbf{X}_i) = 1/N$.

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

2.5.3 Obtaining site-specific results using Importance Sampling

Figure 2 shows the distributions of input variables from our high-fidelity database (§2.4), along with the site-specific distributions for reference site 0 (c.f. Table 2 for site list).

One of the simplest and most straightforward (but not necessarily most precise) ways of carrying out the integrations needed to obtain predicted statistics is to use Importance Sampling ('IS'), where probability weights are applied on each of the database sample points (see e.g. Ditlevsen and Madsen, 1996). The integration (importance sampling) function for determining the expected value of a function $g(\mathbf{x})$ is given by

$$\mathbb{E}[g(\mathbf{x})] = \frac{1}{N}\sum_{i=1}^{N} g(\mathbf{x})\frac{f(\mathbf{x}_i)}{h(\mathbf{x}_i)}, \tag{7}$$

where

- $i = 1 \ldots N$ is the sample point number;
- $\mathbf{x}_i = [x_{1,i}, x_{2,i}, \ldots, x_{9,i}]$ is a 9-component vector array specifying the values of the 9 environmental variables considered at sample point i;
- $f(\mathbf{x}_i) = f(x_{1,i}) \cdot f(x_{2,i}|x_{1,i}) \cdot \ldots \cdot f(x_{9,i}|x_{8,i}, \ldots, x_{1,i})$ is the joint pdf of sample point i according to the **site-specific** probability distribution; and
- $h(\mathbf{x}_i) = h(x_{1,i}) \cdot h(x_{2,i}|h_{1,i}) \cdot \ldots \cdot h(x_{9,i}|x_{8,i}, \ldots, x_{1,i})$ is the joint pdf of sample point i according to the **generic** probability distribution used to generate the database for the 9 variables.

Based on the above, it is clear that only points in the database which also have a high probability of occurrence in the site-specific distribution will have a significant contribution to the lifetime load estimate. Therefore, the IS procedure has relatively slow convergence compared to e.g. a pseudo-MC simulation. Figure 3 shows an example of the convergence of an IS integration for reference site 0, based on using a high-fidelity database with 10^4 points.

2.5.4 Obtaining site-specific results using multi-dimensional interpolation

Estimating an expected function value with a true multi-dimensional interpolation from the high-fidelity database would require finding a set of neighboring points which form a convex polygon. For problem dimensions higher than 3, this is quite challenging due to the non-structured sample distribution. However, it is much easier to find a more crude approximation by simply finding the database point closest to the function evaluation point in a nearest-neighbor approach. This is similar to the table look-up technique often used with structured grids; the denser the distribution of the sample points is, the closer will the results be to an actual Monte Carlo simulation. Finding the nearest neighbor to a function evaluation point requires determining the distances between this point and the rest of the points in the sample space. This is done most consistently in a normalized space, i.e. where the input variables have equal scaling. The cdf (cumulative distribution function) of the variables is an example of such a space, as all cdf's have the same range of $(0, 1)$. Thus, the normalized distance between a new evaluation point

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

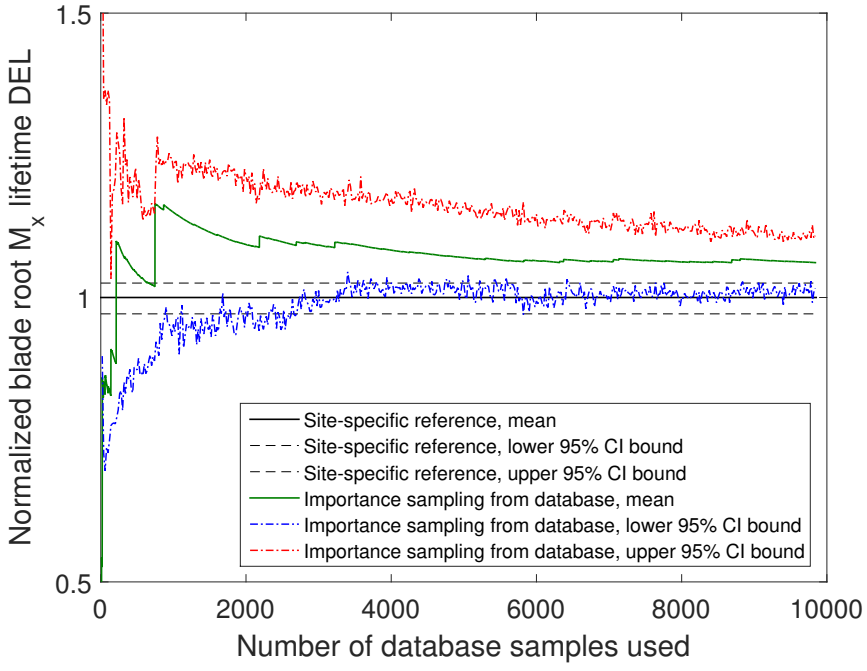

Figure 3. Convergence of an importance sampling (IS) calculation of the blade root moment from the hi-fi database, towards site-specific lifetime fatigue loads for reference site (site 0).

and an existing sample is computed as the vector norm of the (e.g. 9-dimensional vector) differences between the marginal cdf for the two points:

$$|x| = \sqrt{\mathbf{D}^T\mathbf{D}} \tag{8}$$

where $\mathbf{D} = \mathbf{Y} - \hat{\mathbf{Y}}$ is the difference between the current evaluation point $\mathbf{Y}$ and the existing sample points in the reference database, $\hat{\mathbf{Y}}$. The vector $\mathbf{Y}^T = [F_1(X_1), F_2(X_2|X_1), \ldots, F_n(X_n|X_1, \ldots, X_{n-1})]$ consists of the marginal cdf functions of the input variables $\mathbf{X}$ as obtained using a Rosenblatt transformation.

Since some of the input variables may have significantly bigger influence on the result than other variables, it may be useful to weight the cdf of different variables according to their importance (e.g. by making the weights proportional to the variable sensitivity indices; see Section 4.1.2).

2.5.5 Uncertainty estimation and confidence intervals

With the present problem of evaluating the uncertainty in aeroelastic simulations, for any specific combination of environmental conditions, $\mathbf{X}_i$, there will be uncertainty in the resulting damage-equivalent loads, $S_{\text{eq}}(\mathbf{X}_i)$. Part of this uncertainty is statistical by nature and is caused by realization-to-realization variations in the turbulent wind fields used as input to the load simulations.

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

This uncertainty is normally taken into account by carrying out load simulations with multiple realizations (seeds) of turbulence inputs. Provided that enough load realizations have been generated, the seed-to-seed uncertainty can be characterized by the sample statistics (mean and standard deviation) and with an assumption about the statistical distribution. For the damage-equivalent loads which are non-negative by definition, a log-normal distribution is a suitable choice; then the confidence intervals for $(S_{eq,lifetime})^m$ for confidence level α can be found by:

$$\begin{aligned} CI^{+}_{(S_{eq,lifetime})^m}(\alpha) = &\quad \exp\left(\mu_{LN} + \Phi(1-\alpha/2)\cdot\sigma_{LN}\right) \\ CI^{-}_{(S_{eq,lifetime})^m}(\alpha) = &\quad \exp\left(\mu_{LN} - \Phi(\alpha/2)\cdot\sigma_{LN}\right) \end{aligned} \tag{9}$$

where

$$\mu_{LN} = \log\left(\frac{\mathrm{E}[(S_{eq,lifetime})^m]}{\sqrt{1+\frac{\sigma^2_{(S_{eq,lifetime})^m}}{\mathrm{E}[(S_{eq,lifetime})^m]^2}}}\right) \quad ; \quad \sigma_{LN} = \sqrt{\log\left(1+\frac{\sigma^2_{(S_{eq,lifetime})^m}}{\mathrm{E}[(S_{eq,lifetime})^m]^2}\right)}$$

are the parameters of the log-normal distribution, and Φ denotes the standard Normal cumulative distribution function. It should be noted that the confidence intervals defined above are given in terms of the lifetime DEL raised to the power of the Wøhler slope m, meaning that they actually reflect the range of variation of the lifetime fatigue damage, as visible from eq. (4). Using the sample information, confidence intervals can also be determined in a straightforward way using the bootstrapping technique (Efron, 1979). Its main advantage is robustness and no necessity for assuming a statistical distribution of the uncertain variable. With this approach, each function realization is given an integer index, e.g., from 1 to N for N function realizations. Then, a "bootstrap" sample is created by generating random integers from 1 to N, and, for each random integer, assigning the original sample point with the corresponding index, as part of the new bootstrap sample. Since the generation of random integers allows number repetitions, the bootstrap sample will in most cases differ from the original sample. To obtain a measure of the uncertainty in the original sample, a large number of bootstrap samples are drawn, and the resultant quantity of interest (e.g. the lifetime fatigue load) is computed for each of them. Then, the empirical distribution of the set of outcomes is used to define the confidence intervals. If M bootstrap samples have been drawn, and R is the set of outcomes ranked by value in ascending order, then the bounds for confidence level α equal

$$\begin{aligned} CI^{+}_{S_{eq,lifetime}}(\alpha) = &\quad R_{[(1-\alpha/2)M]} \\ CI^{-}_{S_{eq,lifetime}}(\alpha) = &\quad R_{[\alpha M/2]} \end{aligned} \tag{10}$$

where the square brackets $[x]$ indicate the integer part of x, and $R_{[x]}$ means the value in R with rank equal to $[x]$.

Note that the confidence intervals estimated with the above procedure describe only the statistical uncertainty due to the finite number of random samples, and due to seed-to-seed variation. Narrowing them down by e.g. creating a large number of model realizations does not eliminate other model uncertainties as well as uncertainties in the input variables.

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

3 Reference sites

The site-specific load calculation methods presented in this study are validated against a set of reference site-specific load calculations on a number of different sites which cover a wide fraction of the variable domain included within the high-fidelity database. In order to show a realistic example of situations where a site-specific load estimation is necessary, the majority of the sites chosen are characterized with conditions which slightly exceed the standard conditions specified by a certain type-certification class. Exceptions are site 0 which has the most measured variables available and is therefore chosen as a main reference site, and the "sites" representing standard IEC class conditions. The IEC classes are included as test sites as they are described by only one independent variable (mean wind speed). They are useful test conditions as it may be challenging to correctly predict loads as function of only one variable using a model based on up to 9 random variables. The list of test sites is given in Table 2.

Table 2. Reference sites used for validation of the site-specific load estimation methods.

Site No.	Location	Terrain	Specific condition	Variables included
0	Denmark	Flat agricultural	-	$U, \sigma_u, \alpha, L, \Gamma, \Delta\varphi$
1	Denmark	Flat agricultural	IIIC exceedance	U, σ_u, α
2	N. Denmark	Forested	IIIB exceedance	U, σ_u, α
3	N. Denmark	Forested	IA exceedance	U, σ_u, α
4	N. Denmark	Forested	IIA exceedance	U, σ_u, α
5	USA, Colorado	Mountain foothills	Low-wind	U, σ_u, α
6	USA, Colorado	Mountain foothills	Low-wind	U, σ_u, α
IEC IA, NTM	-	-	Standard reference class	U
IEC IIB, NTM	-	-	Standard reference class	U
IEC IIIC, NTM	-	-	Standard reference class	U
IEC IIB, ETM	-	-	Standard reference class	U

For each site, the joint distributions of all variables are defined in terms of conditional dependencies, and generating simulations of site-specific conditions is carried out using the Rosenblatt transformation, (1). The conditional dependencies are described in terms of functional relationships between the governing variable and the distribution parameters of the dependent variable, e.g. the mean and standard deviation of the turbulence are modelled as linearly dependent on the wind speed as recommended by the IEC 61400-1 standard, while the mean wind shear is dependent on the mean wind speed and on the turbulence, as defined by (Kelly et al., 2014). With this procedure, Pseudo-Monte Carlo samples of the environmental conditions at each site are generated from the respective joint distribution, and fed as input to load simulations. The resulting reference lifetime equivalent loads are then defined as the sample means from the Monte Carlo simulations, while the uncertainty in the lifetime loads is estimated using bootstrapping.

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

3.1 Site locations

Site 0 (also also referred to herein as the reference site) is located at the Nørrekær Enge wind farm in Northern Denmark (Borraccino et al., 2017), over flat, open agricultural terrain (Fig. 4). Site 1 is a flat-terrain, near-coastal site at the the National

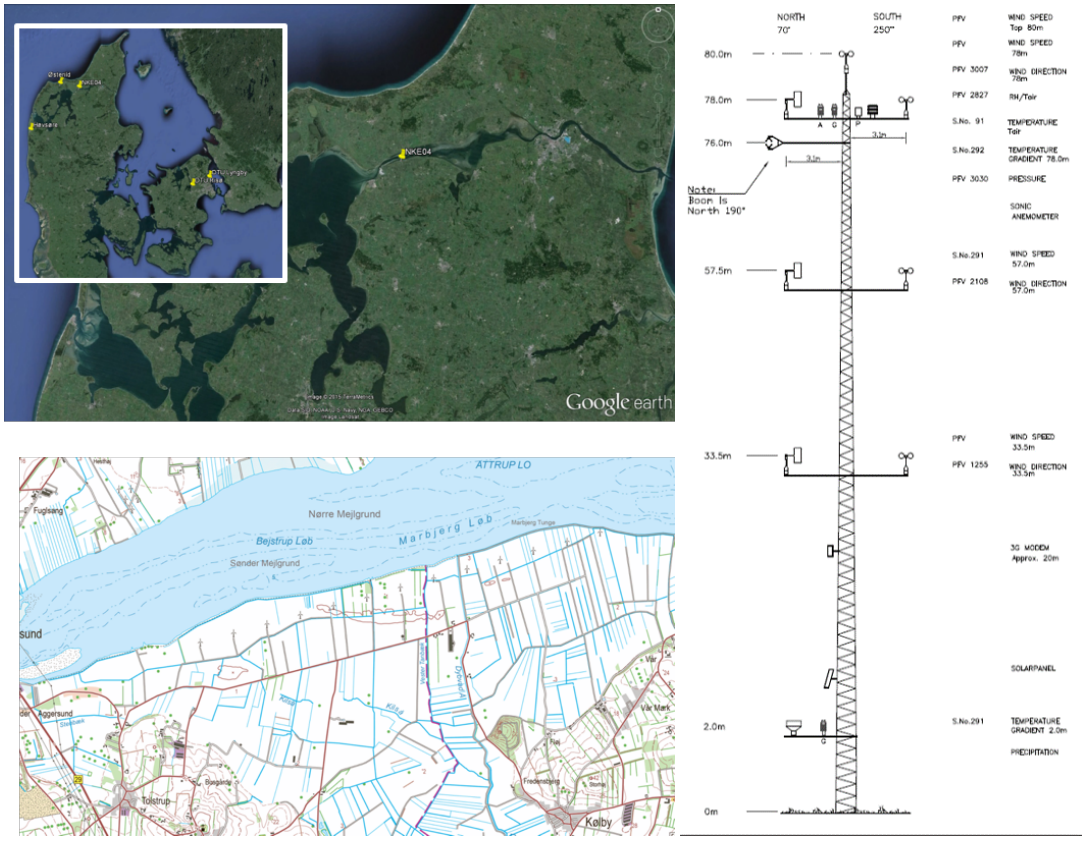

Figure 4. Nørrekær Enge site and mast. *Permissions by DTU Wind Energy.*

Centre for Wind Turbines at Høvsøre, Denmark (Peña et al., 2016, and shown here in Fig. 5). Sites 2 to 4 are based on the wind conditions measured at the Østerild Wind Turbine Test Field which is located in a large forest plantation in North-western Denmark (Hansen et al., 2014) (Fig. 6). Due to the forested surroundings of the site, the flow conditions are more complex than those in Nørrekær Enge and Høvsøre. By applying different filtering according to wind direction, three imaginary site climates are generated and considered as sites 2–4.

Sites 5 and 6 are located at NREL's National Wind Technology Center (NWTC), near the base of the Rocky Mountain foothills just south of Boulder, Colorado (see Fig. 8; c.f. Clifton et al., 2013). Similar to Østerild, directional filtering is applied to the NTWC data to split it into two virtual sites—accounting for the different conditions and wind climates from the two ranges of directions considered.

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

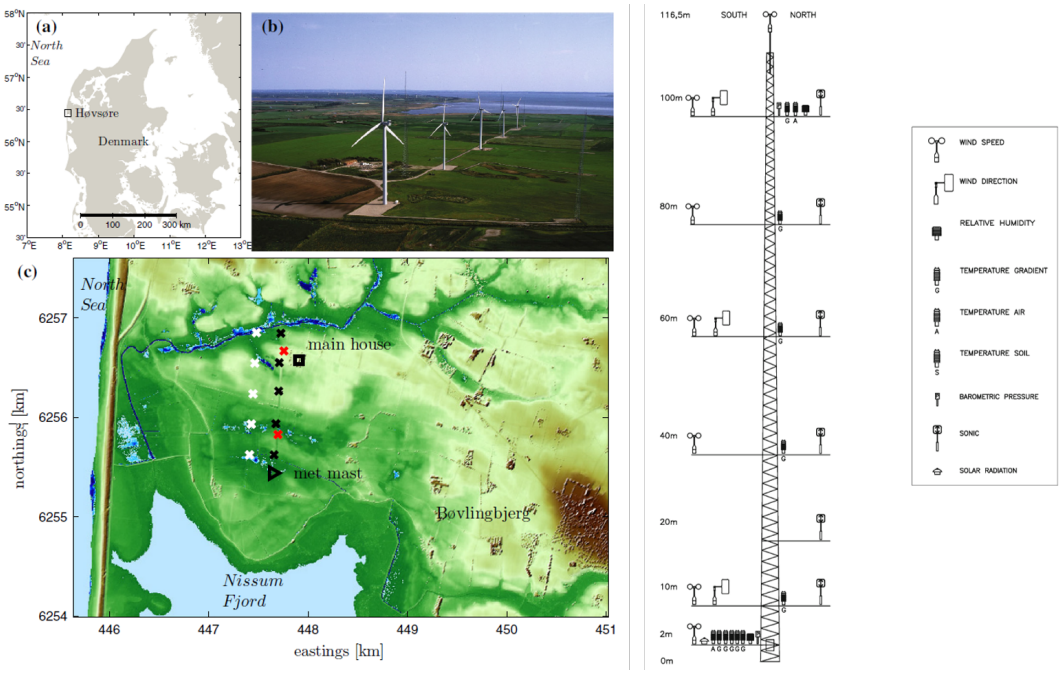

Figure 5. Høvsøre site and mast. *Permissions by DTU Wind Energy.*

4 Reduced order models

In this section we present three different reduced order models: 1) Polynomial chaos expansion, 2) Universal Kriging, and 3) Quadratic response surface. All three methodologies are to be calibrated to the database presented in Section 2.4 in order to predict site-specific loads.

4.1 Polynomial chaos expansion

Polynomial Chaos Expansion (PCE) is a popular method for approximating a stochastic function of multiple random variables using an orthogonal polynomial basis. In the classical definition of PCE (Ghanem and Spanos, 1991) the input random variables $\mathbf{X}$ are defined on $(-\infty, \infty)$, with Hermite polynomials typically used as the polynomial basis.[1] Choosing a polynomial basis which is orthogonal to a non-Gaussian probability measure turns the PCE problem into the so-called Wiener-Askey or Generalized chaos, (Xiu and Karniadakis, 2002). For the present problem, a Generalized PCE using Legendre polynomials is considered most suitable as the Legendre polynomials $P_n(\xi)$ are orthogonal with respect to a uniform probability measure in the interval $\xi = [-1, 1]$, which means that the PCE can conveniently be applied on the cumulative distribution functions of the

[1] In the classical definition of the PC decomposition used in e.g. spectral stochastic finite element methods (Ghanem and Spanos, 1991), the input random variables are Normally distributed (Gaussian), which means that the Hermite polynomials are a suitable Hilbertian basis—since the Hermite polynomials are orthogonal with respect to the Gaussian probability measure. In this case, the properties of the Hermite polynomials dictate that the random variables $\mathbf{X}$ are defined on $(-\infty, \infty)$.

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

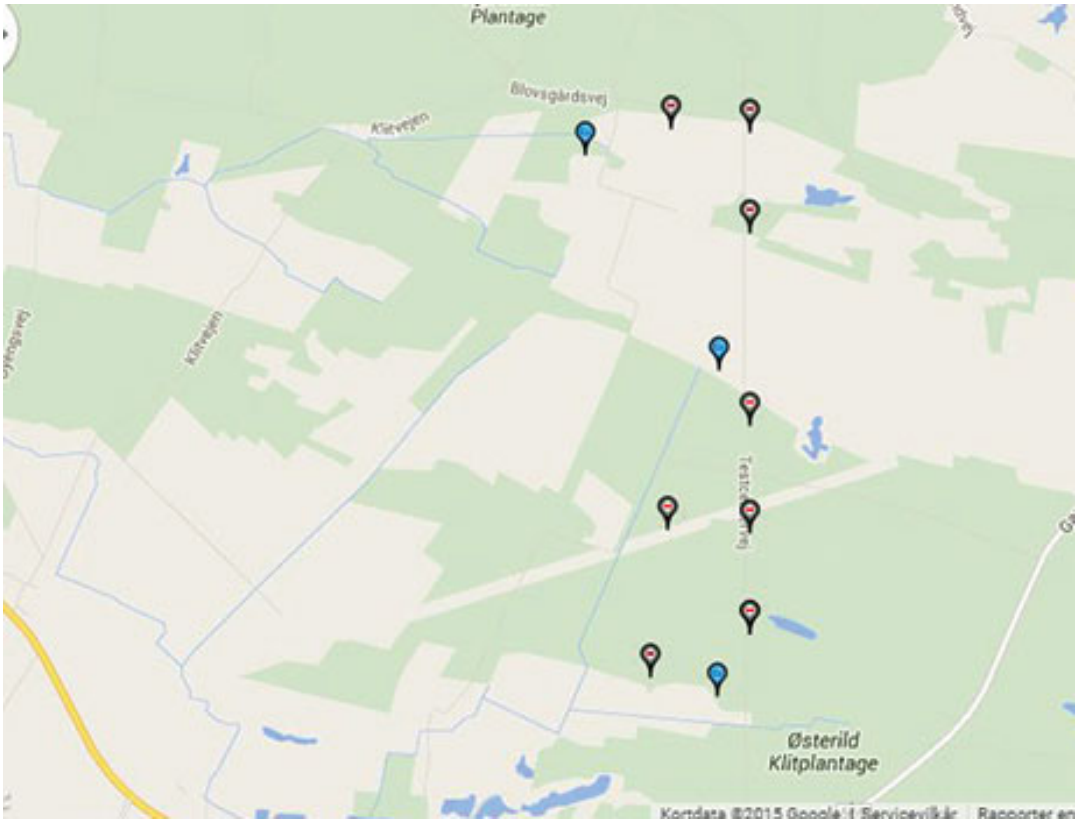

Figure 6. The fully instrumented mast at Østerild is the blue dot furthest South/Down on the map. *Permissions by DTU Wind Energy.*

Figure 7. Left: Cup anemometer (left) and wind vane (right) at 244 m height, Østerild test site. *Permissions by DTU Wind Energy.*

variables $\mathbf{X}$ which are defined in the interval $[0,1]$ so that

$$\xi_i = 2F(X_i) - 1, \tag{11}$$

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

Figure 8. Aerial view of NWTC. *Permissions by NREL.*

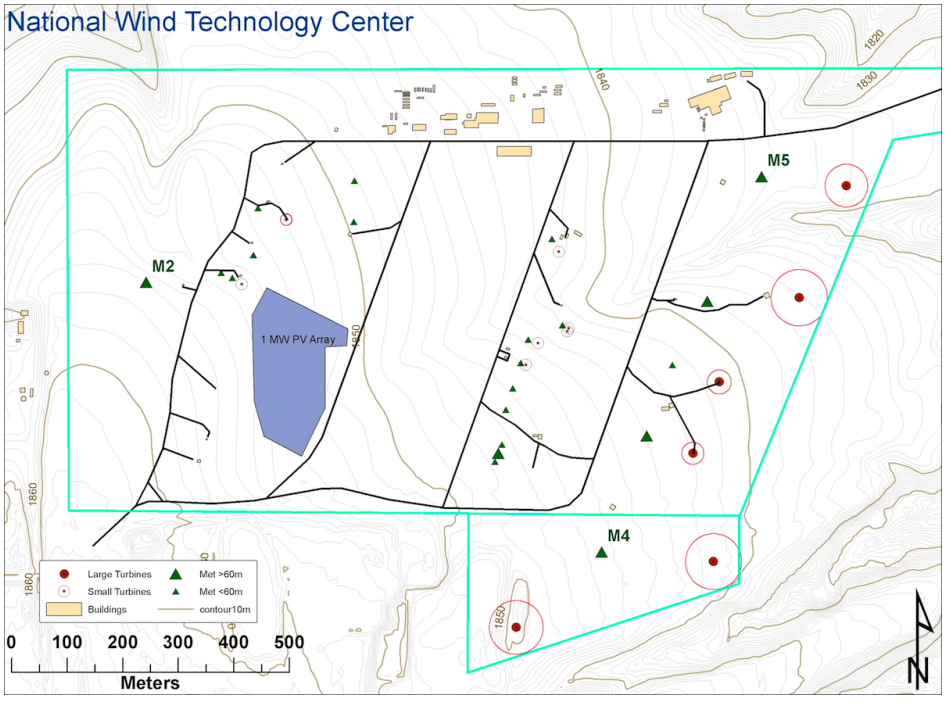

Figure 9. Map of NWTC. *Permissions by NREL.*

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

where $F(X_i)$ is the cumulative distribution function of a variable $X_i \in \mathbf{X}$, $i = 1, \ldots, M$. With this definition, the PCE represents a model applied to a set of transformed variables which due to the applied transformation are independent and identically distributed ('i.i.d.'). Note that eq. (11) and the evaluation of the cumulative distribution in general does not ensure independence - so an appropriate transformation should also account for the dependence between variables. In the present case, it is convenient to apply the Rosenblatt transformation as defined in eq. (1). For the current implementation of PCE, only eq. (1) is required since the expansion is based on the Legendre polynomials, however the transformation to standard Normal space in eq. (2) is used for other procedures, e.g. the response surface model discussed later.

Using the notation defined by Sudret (2008), we consider the family of univariate Legendre polynomials $P_n(\xi)$. A multivariate, generalized PCE with M dimensions and maximum polynomial degree p is defined as the product of univariate Legendre polynomials where the maximum degree is less than or equal to p. The univariate polynomial family for dimension i can be defined as

$$P_{\alpha_i}(\xi), \qquad \text{where} \quad i = 1, \ldots, M, \quad \alpha_i \in \mathbb{N}, \quad \sum_{i=1}^{M} \alpha_i \leq p(\alpha_i \geq 0). \tag{12}$$

The multivariate polynomial of dimension M is then defined as

$$\Psi_{\boldsymbol{\alpha}} = \prod_{i=1}^{M} P_{\alpha_i}(\xi_i) \tag{13}$$

The total number of polynomials of this type is

$$N_p = \frac{(M+p)!}{M!\,p!} \tag{14}$$

The aim of using PCE is to represent a scalar quantity $S = g(\mathbf{X})$ in terms of a truncated sequence $\tilde{S}(\mathbf{X}) + \varepsilon$ where ε is a zero-mean residual term. With this definition, the multivariate generalized PCE of dimension M and maximum degree p is given by

$$\tilde{S}(\boldsymbol{\xi}) = \sum_{j=0}^{N_p-1} S_j \Psi_{\boldsymbol{\alpha},j}(\boldsymbol{\xi}) \tag{15}$$

where $S_j \in \mathbf{S} = [S_1, \ldots, S_{Np}]$ are unknown coefficients which need to be determined, and $\boldsymbol{\xi} = [\xi_1, \ldots \xi_M]$ are functions of $\mathbf{X}$ as defined in (11). The most straightforward way of determining $\mathbf{S}$ is minimizing the variance of the residual ε using a least-squares regression approach:

$$\mathbf{S} = \min \left\{ \frac{1}{N_e} \left| \sum_{i=1}^{N_e} \left[g(\boldsymbol{\xi}^{(i)}) - \sum_{j=0}^{N_p-1} S_j \Psi_{\boldsymbol{\alpha},j}(\boldsymbol{\xi}^{(i)}) \right]^2 \right| \right\} \tag{16}$$

where N_p is the number of polynomial coefficients in the PCE and N_e is the number of sampling points in the experimental design. For this purpose, a design experiment has to be set up and the so-called design matrix $\boldsymbol{\Psi}$ needs to be constructed:

$$\boldsymbol{\Psi}_{ij} = \Psi_{\boldsymbol{\alpha},j}(\xi^{(j)}); \qquad i = 1, \ldots, N_e, \qquad j = 1, \ldots, N_p. \tag{17}$$

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Under the condition that the residuals are (approximately) Normally-distributed, the solution to equation (16) is given by

$$\mathbf{S} = (\mathbf{\Psi}^{\mathrm{T}}\mathbf{\Psi})^{-1} \cdot \mathbf{\Psi}^{\mathrm{T}} \cdot \mathbf{y}, \tag{18}$$

with $\mathbf{y} = g(\mathbf{x}^{(i)})$ being a vector with the outcomes of the functional realizations obtained from the design experiment, where $i = 1 \ldots N_e$.

The solution of eq. (18) requires that the so-called information matrix $(\mathbf{\Psi}^{\mathrm{T}}\mathbf{\Psi})$ is well-conditioned, which normally requires that the number of collocation points N_e is significantly larger than the number of expansion coefficients N_p. Subsequently, the problem grows steeply in size when M and p increase. In such situations, it may be desirable to limit the number of active coefficients by carrying out a Least Absolute Shrinkage and Selection Operator (LASSO) regression (Tibshirani, 1996), which regularizes the regression by penalizing the sum of the absolute value of regression coefficients:

$$\mathbf{S} = \min \left\{ \left| \frac{1}{2N_e} \sum_{i=1}^{N_e} \left[g(\boldsymbol{\xi}^{(i)}) - \sum_{j=0}^{N_p-1} S_j \Psi_{\boldsymbol{\alpha},j}(\boldsymbol{\xi}^{(i)}) \right]^2 + \lambda \sum_{j=0}^{N_p-1} |S_j| \right| \right\} \tag{19}$$

where λ is a positive regularization parameter; larger values of λ increase the penalty and reduce the absolute sum of the regression coefficients, while $\lambda = 0$ is equivalent to ordinary least-squares regression.

4.1.1 Convergence of PCE

We assess the convergence of PCE by calculating the normalized root-mean-square (NRMS) error between a set of observed quantities (i.e. damage-equivalent loads from simulations) $\mathbf{y} = g(\mathbf{X}^{(i)}), i = 1 \ldots N$, and the PCE predictions, $\tilde{\mathbf{y}} = \tilde{S}(\mathbf{X}^{(i)}), i = 1 \ldots N$, over the same set of N sample points $\mathbf{X}^{(i)}$:

$$\varepsilon_{N,\mathrm{RMS}} = \frac{1}{\mathrm{E}[\mathbf{y}]} \sqrt{\frac{\sum_{i=1}^{N} (\tilde{y}_i - y_i)^2}{N}} \tag{20}$$

where $\mathrm{E}[\mathbf{y}]$ is the expected value of the observed variable. Figure 10 shows the NRMS error for a PCE of order 6 and with 6 dimensions, as function of the number of samples used to train the PCE, and the hours of load simulations used for each sample point. The NRMS error shown on Figure 10 is calculated based on a set of 500 pseudo-MC points sampled from the joint pdf of reference site 0, and represents the difference in blade root flapwise DEL observed in each of the 500 points vs. the DEL predicted by a PC expansion trained on a selection of points from the high-fidelity database described in Section 2.

Figure 10 illustrates a general tendency that using a few thousand training samples leads to convergence of the values of the PC coefficients, and the remaining uncertainty is due to seed-to-seed variations and due to the order of the PCE being lower than what is required for providing an exact solution at each sample point. Using longer simulations per sample point does not lead to further reduction in the statistical uncertainty due to seed-to-seed variations - with 4000 training samples, the RMSE error for 1h simulation per sample is almost identical to the error with 8h simulation per sample. The explanation for this observation is that the seed-to-seed variation introduces an uncertainty not only between different simulations within the same sampling point, but also between different sampling points. This uncertainty materializes as an additional variance which is

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

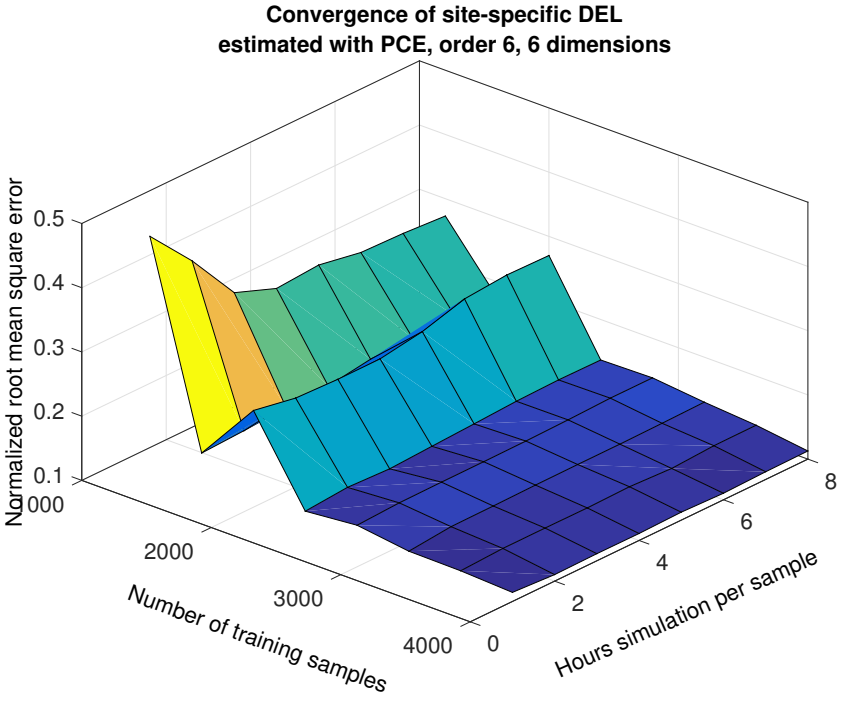

Figure 10. Convergence of a PCE of dimension 6 and order 6, as function of number of collocation points and hours of simulation per collocation point. The z-axis represents the NRMS error obtained from the difference between 500 site-specific pseudo-MC samples of blade root flapwise DEL for reference site 0, and the corresponding predictions from PCE.

not explained by the smooth PCE surface. Therefore, the RMSE error achieved under these conditions can be considered near to the smallest possible for the given PCE order and number of dimensions - and further increase in the number of training points or simulation length will not introduce noticeable improvement. However, it should be noted that the number of training points required for such convergence will differ according to the order and dimension of the PCE, and higher order and more dimensions will require more than the ~ 3000 points which seem sufficient for a PCE of order 6 and dimension 6, as shown on Figure 10.

4.1.2 Sensitivity indices and model reduction

One useful corollary of the orthogonality in the PCE polynomial basis is that the total variance of the expansion can be expressed as the sum of the contributions from individual terms (Sudret, 2008):

$$\mathrm{Var}\left[\tilde{S}(\boldsymbol{\xi})\right] = \mathrm{Var}\left[\sum_{j=0}^{N_p-1} S_j \Psi_{\boldsymbol{\alpha},j}(\boldsymbol{\xi})\right] = \sum_{j=1}^{N_p-1} S_j^2 \mathrm{E}\left[\Psi_{\boldsymbol{\alpha},j}^2(\boldsymbol{\xi})\right] \tag{21}$$

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Each of the terms in the sum in eq. (21) represents the contribution of the variables contained in the respective multivariate polynomials $\Psi_{\boldsymbol{\alpha},j}$ where $j = 0 \dots N_p - 1$. This property can be used for eliminating polynomials which do not contribute significantly to the variance of the output, thus achieving a sparse, more computationally efficient reduced model. By combining the variance truncation and the LASSO regression technique in eq. (19), a model reduction of an order of magnitude or more can be achieved. For example, for a 9-dimensional PCE of order 6, using LASSO regularization parameter $\lambda = 1$ and retaining the polynomials which have a total variance contribution of 99.5%, resulted in a reduction of the number of polynomials from 5005 to about 200.

Denoting by $\mathscr{F}_{i_1,\dots,i_s}$ the set of all polynomials dependent on a specific combination of input variables $(i_1, \dots, i_s)$ (and only on them), the sum of variance contributions over $\mathscr{F}_{i_1,\dots,i_s}$ normalized with the total variance represents the PCE-based Sobol' index with respect to variable set $\mathscr{F}_{i_1,\dots,i_s}$ (Sudret, 2008):

$$SU_{i_1,\dots,i_s} = \left(\sum_{\boldsymbol{\alpha} \in \mathscr{F}_{i_1,\dots,i_s}} S_{\boldsymbol{\alpha}}^2 \mathrm{E}\left[\Psi_{\boldsymbol{\alpha}}^2(\boldsymbol{\xi})\right] \right) \cdot \left(\mathrm{Var}\left[\tilde{S}(\boldsymbol{\xi})\right] \right)^{-1}. \tag{22}$$

Based on eq. (22) it is also straightforward to obtain the total Sobol indices for a given variable j by summing all $SU_{i_1,\dots,i_s}$ where $j \in (i_1, \dots, i_s)$. Note that since each variable appears in multiple cross-terms in the expansion, the contributions of some polynomial coefficients are included multiple times in the total Sobol' indices, meaning that the sum of the total indices will typically exceed 1.

The Sobol indices estimated using the above procedure represent the relative contribution to the model variance from variables following the joint input distribution used to calibrate the PCE. In the present case, this distribution would span the uniform variable space of the high-fidelity database defined in Section 2, and the indices will correspond to the load variation within the entire variable ranges as defined in Table 1. It may be more relevant to compute site-specific Sobol indices. This can be carried out efficiently by a Monte Carlo simulation on the PCE. For number of dimensions equal to M and for N (pseudo) Monte-Carlo samples the required experimental design represents an $N \times 2M$ matrix. This is divided into two $N \times M$ matrices, $\mathbf{A}$ and $\mathbf{B}$. Then, for each dimension i, $i = 1 \dots M$, a third matrix $\mathbf{AB}_i$ is created by taking the i^{th} column of $\mathbf{AB}_i$ equal to the i^{th} column from $\mathbf{B}$, and all other columns taken from $\mathbf{A}$. The load response function, i.e. the PCE, is then evaluated for all three matrices, resulting in three model estimates: $f(\mathbf{A})$, $f(\mathbf{B})$, and $f(\mathbf{AB}_i)$. By repeating this for $i = 1 \dots M$, simulation-based Sobol' sensitivity indices can be estimated as

$$SU_i = \frac{1}{N} \sum_{j=1}^{N} f(\mathbf{B})_j \left(f(\mathbf{AB}_i)_j - f(\mathbf{B})_j \right) \tag{23}$$

where $j = 1 \dots N$ is the row index in the design matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{AB}_i$ (Saltelli et al., 2008).

4.2 Universal Kriging with polynomial chaos basis functions

Kriging (Sacks et al., 1989; Santher et al., 2003) is a stochastic interpolation technique which assumes the interpolated variable follows a Gaussian process. A Kriging metamodel is described (Sacks et al., 1989) by

$$Y(\mathbf{x}) = \boldsymbol{\beta}^T \mathbf{f}(\mathbf{x}) + Z(\mathbf{x}), \tag{24}$$

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

where $\mathbf{x}$ represents the input variables, and $Y(\mathbf{x})$ is the output. The term $\boldsymbol{\beta}^T\mathbf{f}(\mathbf{x})$ is the mean value of the Gaussian process (a.k.a. the "trend") represented as a set of basis functions $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}),\ldots,f_P(\mathbf{x})]$ and regression coefficients $\boldsymbol{\beta} = [\beta_1,\ldots,\beta_P]$; $Z(\mathbf{x},\mathbf{w})$ is a stationary, zero-mean Gaussian process. The probability distribution of the Gaussian process is characterized by its covariance

$$V(\mathbf{w},\mathbf{x}) = \sigma^2 R(\mathbf{w},\mathbf{x},\boldsymbol{\theta}) \tag{25}$$

where σ^2 is the process variance which is assumed to be constant, and $R(\mathbf{w},\mathbf{x},\boldsymbol{\theta})$ is the correlation between $Z(\mathbf{x})$ and $Z(\mathbf{w})$. The hyperparameters $\boldsymbol{\theta}$ define the correlation behavior, in terms of e.g. a correlation length. Given a set of points $\mathbf{x} = [x_1, x_2, \ldots x_N]$ where the true function values $\mathbf{y} = Y(\mathbf{x})$ are known, the aim is to obtain a model prediction at a new point, x'. Based on Gaussian theory, the known values $Y(\mathbf{x})$ and the Kriging predictor $\hat{Y}(x')$ will be jointly Gaussian distributed, and $\hat{Y}(x')$ will have the following mean and variance (Santher et al., 2003):

$$\begin{aligned} \mu_{\hat{Y}}(x') &= \mathbf{f}(x')^T\boldsymbol{\beta} + r(x')^T\mathbf{R}^{-1}(\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\beta}) \\ \sigma^2_{\hat{Y}}(x') &= \sigma^2(1 - r(x')^T\mathbf{R}^{-1}r(x') + u(x')^T[\boldsymbol{\Psi}^T\mathbf{R}^{-1}\boldsymbol{\Psi})^{-1}u(x')]. \end{aligned} \tag{26}$$

Here

$\boldsymbol{\Psi}$ is the design matrix collecting the terms constituting the basis functions $\mathbf{f}(\mathbf{x})$,

$$\Psi_{ij} = f_j(x_i) \quad \text{for} \quad i = 1\ldots N \quad \text{and} \quad j = 1\ldots P;$$

$r(x')$ is the vector of cross-correlations between the prediction point x' and the known points $\mathbf{x}$;

$\mathbf{R}$ is the correlation matrix of the known points,

$$R_{ij} = R(x_i, x_j, \theta) \quad \text{for} \quad i,j = 1,\ldots,N; \quad \text{and}$$

$$u(x') = \boldsymbol{\Psi}^T\mathbf{R}^{-1}r(x') - f(x')\ .$$

Using the predictor functions above requires determining the regression coefficients ($\boldsymbol{\beta}$), the field variance (σ^2), and the correlation hyperparameters ($\boldsymbol{\theta}$). A suitable approach is to find the values of $\boldsymbol{\beta}$, σ^2 and $\boldsymbol{\theta}$ which maximize the likelihood of $\mathbf{y}$, i.e. minimize the model error in a least-squares sense (Lataniotis et al., 2015):

$$\mathcal{L}(\mathbf{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{\theta}) = \frac{\det(\mathbf{R})^{-1/2}}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\beta})^T\mathbf{R}^{-1}(\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\beta})\right]. \tag{27}$$

Here the hyperparameters, $\boldsymbol{\theta}$, appear within the correlation matrix $\mathbf{R}$. Having set up the design matrix $\boldsymbol{\Psi}$, the expansion coefficients $\boldsymbol{\beta}$ can be determined with the least-squares approach,

$$\boldsymbol{\beta} = \beta(\boldsymbol{\theta}) = (\boldsymbol{\Psi}^T\mathbf{R}^{-1}\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^T\mathbf{R}^{-1}\mathbf{y}. \tag{28}$$

The field variance is obtained as

$$\sigma^2 = \sigma^2(\boldsymbol{\theta}) = \frac{1}{N}(\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\beta})^T\mathbf{R}^{-1}(\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\beta}). \tag{29}$$

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

From (28) and (29) it follows that $\boldsymbol{\beta}$ and σ^2 can be expressed as functions of $\boldsymbol{\theta}$. Therefore, calibrating the Kriging model amounts to finding the values of $\boldsymbol{\theta}$ which maximize the likelihood. By combining eqns. 27–29 this leads to the optimization problem

$$\boldsymbol{\theta} = \arg\min_{D_\theta} \left(\frac{1}{2} \log\left(\det(\mathbf{R})\right) + \frac{N}{2} \log\left(2\pi\sigma^2\right) + \frac{N}{2} \right). \tag{30}$$

For a problem with M dimensions, we assume that the correlation between sample points can be modelled using an anisotropic separable correlation function ((Sacks et al., 1989; Lataniotis et al., 2015), which assumes a different correlation parameter for each dimension. The total correlation is expressed as the product of the individual one-dimensional correlation functions,

$$R(\mathbf{x}, \mathbf{x}', \boldsymbol{\theta}) = \prod_{i=1}^{M} R(x_i, x_i', \theta_i). \tag{31}$$

The one-dimensional correlation functions are assumed to follow an exponential relation to the distance $h = (x_i - x_i')$ between points,

$$R(h, \theta) = \exp\left(-\frac{|h|}{\theta} \right). \tag{32}$$

The functional form of the mean field $\mathbf{f}(\mathbf{x})^T \boldsymbol{\beta}$ is identical to the generalized PCE defined in eq. (18), meaning that the PCE is a possible candidate model for the mean in a Kriging interpolation. We adopt this approach and define the Kriging mean as a PCE with properties as described in section 4.1.

The main practical difference between regression- or expansion-type models such as regular PCE and the Kriging approach is in the way the training sample is used in the model: in the pure regression-based approaches the training sample is used to only calibrate the regression coefficients, while in Kriging as in other interpolation techniques the training sample is retained and used in every new model evaluation. As a result the Kriging model may have an advantage in accuracy since the model error tends to zero in the vicinity of the training points; however this comes at the expense of an increase in the computational demands for new model evaluations. The extra computational burden is mainly the time necessary to assemble $\mathbf{r}(\mathbf{x}')$, the matrix of cross-correlations between the prediction points and the training sample, and the time to multiply $\mathbf{r}(\mathbf{x}')$ with other equation terms. Thus, while for a PCE the model evaluation time $t(N)$ for a sample of size N would follow $t(N) = \mathcal{O}(N)$, for a Kriging model $t(N) = \mathcal{O}(N^2)$. For a Kriging model, a gain in accuracy over the model represented by the trend function will only materialize in problems where there is a noticeable correlation between the residual values at different training points. In a situation where the model error is independent from point to point (as e.g. in the case when the error is only due to seed-to-seed variations in turbulence) the inferred correlation length will tend to zero and the Kriging estimator will be represented by the trend function alone.

4.3 Quadratic response surface

A quadratic-polynomial response surface (RS) method based on Central Composite Design (CCD) is a reduced-order model which, among other applications, is also used for wind turbine load prediction (Toft et al., 2016). The procedure involves

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

fitting a quadratic polynomial regression (a response surface) to a normalized space of i.i.d variables which are derived from the physical variables using an isoprobabilistic transformation as the Rosenblatt transformation given in eqs. (1) and (2). The design points used for calibrating the response surface form a combination of a central point, axial points with unit shifts in one dimension at a time, and a 2^k factorial design where k equals the number of dimensions and there are two levels per variable dimension (Figure 11). Due to the structured design grid required, it is not possible to use this approach with the sample points from the high-fidelity database described in section 2. Therefore, we implement the procedure using an additional set of simulations. Due to the low-order of the response surface, it is also not possible to fully characterize the turbine response as function of mean wind speed using a single response surface. Therefore, multiple response surfaces are calibrated for wind speeds from 4 to 25 m/s in 1 m/s steps. However, due to the exponential increase of the number of design points with increasing problem dimension, it is not practical to fit response surface covering all 9 variables considered. Instead, we choose to replace three of the variables (yaw, tilt, and air density) with their mean values. The result is a 6-dimensional problem which requires 946 design points in total. Analogically to the high-fidelity database, 8h of simulations are carried out for characterizing each design point. The polynomial coefficients of the response surface are then defined using least-squares regression with the same closed-form solution defined by eq. (18). For any sample point p in the central composite design, the corresponding row in the design matrix is defined as

$$\mathbf{\Psi}_p = \left[\{1\}, \quad \{U_1, \ldots U_n\}, \quad \{U_1^2, \ldots U_n^2\}, \quad \{U_i \cdot U_j,\, i{=}1 \ldots n,\, j{=}1 \ldots (i{-}1)\}\right] \tag{33}$$

where $\mathbf{U}$ are standard Normal variables derived from the physical variables $\mathbf{X}$ by an isoprobabilistic transformation.

5 Results from site-specific calculations

5.1 Lifetime fatigue loads

The lifetime damage-equivalent (DEL) loads are computed for all reference sites in Table 2, using the six methodologies defined above: 1) full pseudo-Monte Carlo simulation; 2) quadratic response surface; 3) polynomial chaos expansion, 4) importance sampling, 5) nearest-neighbor interpolation; and 6) Kriging with the mean defined by polynomial chaos basis functions. Methods 3–6 are based on data from the high-fidelity load database described in Section 2. The load predictions with the Monte Carlo approach are obtained by carrying out HAWC2 aeroelastic simulations with the DTU 10MW reference wind turbine for each validation site. A total of approx.1000 pseudo-MC samples are drawn from the joint distribution of environmental input variables characterizing each site, and 8h of aeroelastic simulations are carried out for each pseudo-MC sample. An exception are the IEC-based sites, where the standard IEC procedure is followed and loads are evaluated for 22 wind speeds from cut-in to cut-out in 1m/s steps. For each site, the full Monte Carlo simulation is then used as a reference to test the performance of the other five methods. The load predictions from the PCE, Kriging, the quadratic RS and the nearest-neighbor interpolation procedures all use a pseudo-MC simulation of the respective model with inputs drawn from the site-specific distributions of environmental variables. The load predictions with Importance Sampling are based on the probability-weighted contributions from all samples in a high-fidelity load database. Based on computations from PCE with different number of dimensions and

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

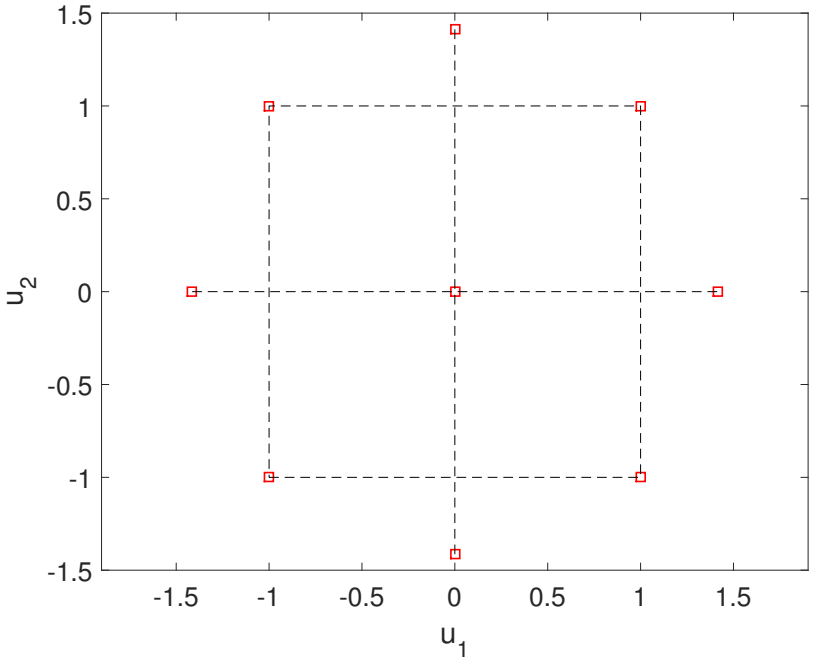

Figure 11. Example of a rotatable Central Composite Design (CCD) in a 2-dimensional standard Normal space $[u_1, u_2]$. The CCD consists of a central point, a 2^k factorial design with 2 levels and $k = 2$ dimensions, and axial points at distance $u = \sqrt{2}$, meaning that all the outer points lie on a circle.

different maximum order, it was observed that expansions of order 4 or 5 may not be sufficiently accurate for some response channels. This is illustrated in Figure 12 where the prediction of main shaft torsion loads using order 4 and order 6 PC expansion are compared against other methods, and the order 4 calculation shows a significant bias. Therefore, the PC expansion used for reporting the results in this section is based on the same 6-dimensional variable input used with the quadratic response surface, and has a maximum order of 6. For evaluating confidence intervals from the reduced-order models (Kriging, PCE and quadratic response surface), two reduced-order models of each type are calibrated - one for the mean values, and the other for the standard deviations. This allows computing confidence intervals directly by eq. (9), or generating a number of realizations for each sampled combination of input variables, and subsequently computing confidence intervals by bootstrapping (eq. (10)). In the present, we use the latter approach, because it allows bootstrapping simultaneously a random sample of the input variables and the random seed-to-seed variations within each sample. As a result, the obtained confidence interval reflects the combination of seed-to-seed uncertainty and the uncertainty due to finite number of samples from the distribution of the input variables. This approach also allows consistency with the Importance Sampling and Nearest-Neighbor interpolation techniques, where bootstrapping is also used. Since the lifetime fatigue load is in essence an integrated quantity subject to the law of large numbers, the uncertainty in computations based on a random sample of size N will be proportional to $1/\sqrt{N}$. Comparing

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

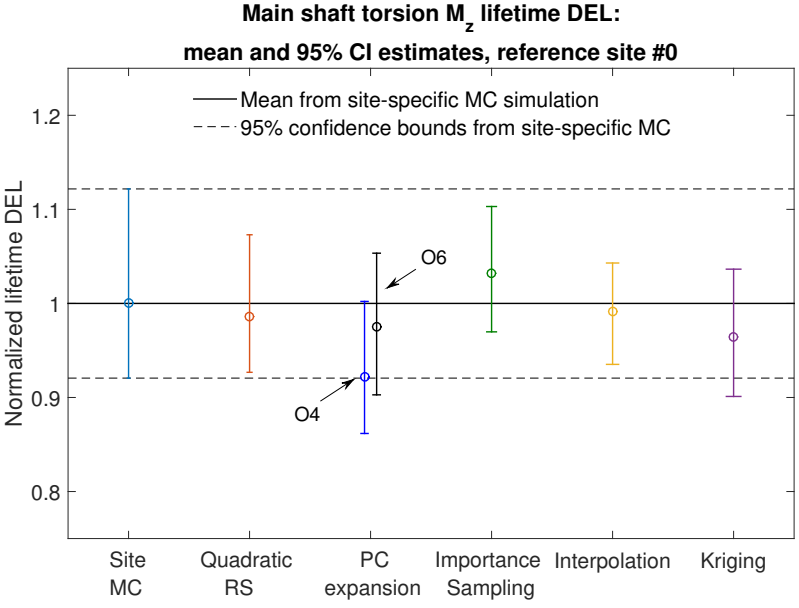

Figure 12. Comparison of predictions of the lifetime damage-equivalent loads for six different estimation approaches. All values are normalized with respect to the mean estimate from a site-specific Monte Carlo simulation, and the error bars represent the bounds of the 95% confidence intervals. Results from two PCEs are shown: the blue bar corresponds to the output of a 4^{th} order PCE, while the black bar corresponds to a 6^{th} order PCE.

uncertainties and confidence intervals as defined in Section 2.5.5 will therefore only be meaningful when approximately the same number of samples is used for all calculation methods. This approach is used for generating Figures 13 and 14, where the performance of all site-specific load estimation methods is compared for reference site 0, for 8 load channels in total. Figure 13 shows results for tower base and tower top fore-aft and side-side bending moments, and Figure 14 displays the tower top yaw moment, the main shaft torsion, and blade root flapwise and edgewise bending moments.

The results for Site 0 show that for all methods the prediction of blade root and tower top loads is more accurate than the prediction of tower base loads. Also, overall the predictions from the reduced models - the quadratic RS and the PCE, as well as from the Kriging model, are more robust than the IS and nearest-neighbor (NN) interpolation techniques. Similar performance is observed for most other validation sites. The full site-specific results for all load estimation methods are shown in Tables 3, 4, 5, 6, and 7, for the PCE, Kriging, quadratic RS, Importance Sampling, and NN-interpolation techniques respectively. In all tables, the values represent the mean estimates and are normalized to the results obtained with the direct site-specific Monte Carlo simulations. The largest observed errors amount to $\approx 9\%$ with Kriging, $\approx 10\%$ for the PCE, $\approx 10\%$ for the quadratic RS, $\approx 24\%$ for IS, and $\sim$ 15–17% for NN-interpolation. Noticeably, the low wind speed, high turbulence site 5 seems as the most difficult for prediction—for most load prediction methods this is the site where the largest error is found. All models except the Kriging also show relatively large errors for the IEC class-based sites. That can be attributed to significantly smaller number

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

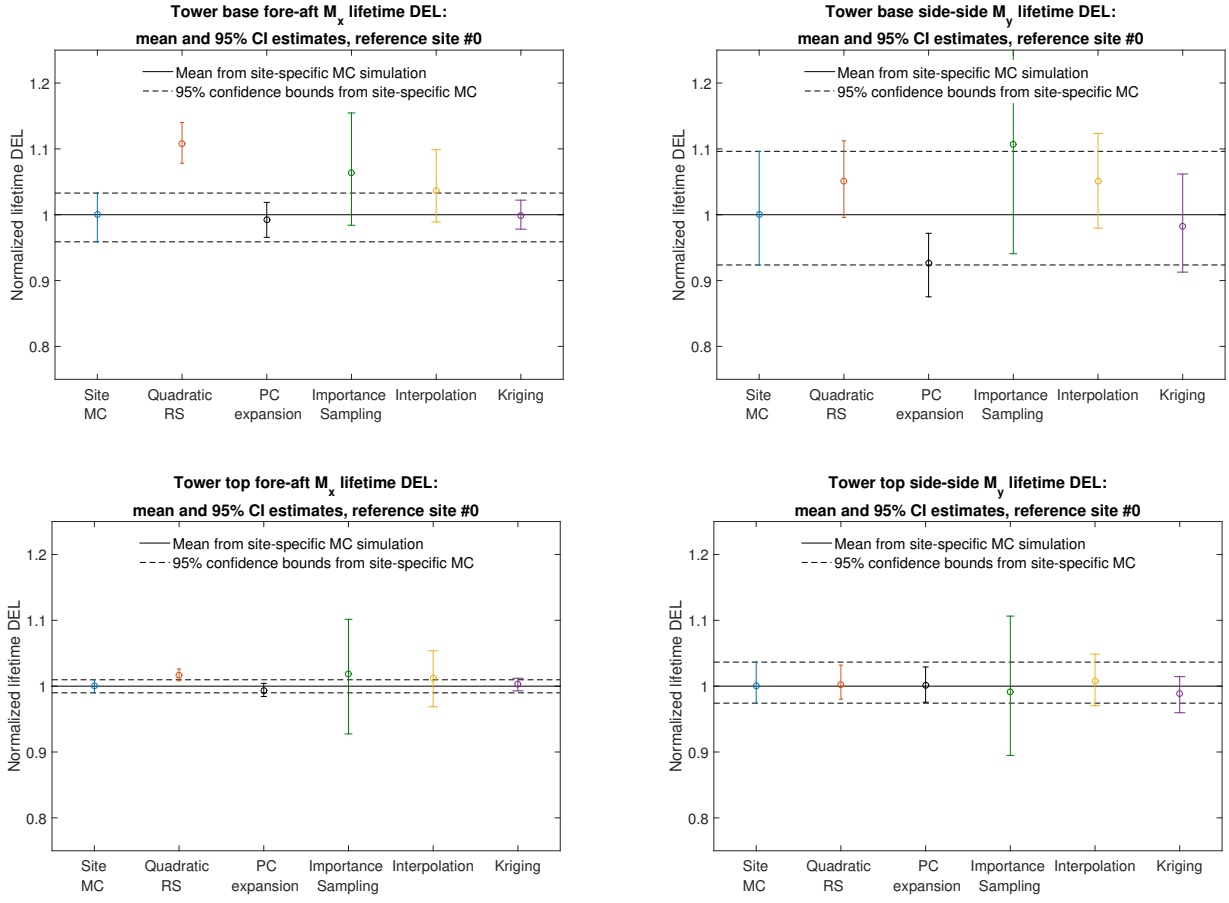

Figure 13. Comparison of predictions of the lifetime damage-equivalent loads for six different estimation approaches. All values are normalized with respect to the mean estimate from a site-specific Monte Carlo simulation.

of samples used for the ICE-based sites (22 samples where only the wind speed is varied in 1m/s steps). As mentioned above, the statistical uncertainty in the estimation of the lifetime DEL will diminish with increasing number of samples. In addition to this effect, as discussed in section 2.5.3, the uncertainty of the IS model can increase when the site-specific distribution has fewer dimensions than the model. It can be expected that this effect is strongest for the IEC class-based sites, as for them only a single variable - the wind speed - is considered stochastic.

5.2 One-to-one comparison and mean squared error

Since the prediction of lifetime fatigue loads is the main purpose of the present study, the performance of the load prediction methods with respect to estimating the lifetime DEL is the main criterion for evaluation. However, the lifetime DEL as an integrated quantity will efficiently identify model bias but may not reveal the magnitude of some uncertainties which result in zero-mean error. As an additional means of comparison we calculate the normalized root-mean square (NRMS) error, defined

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

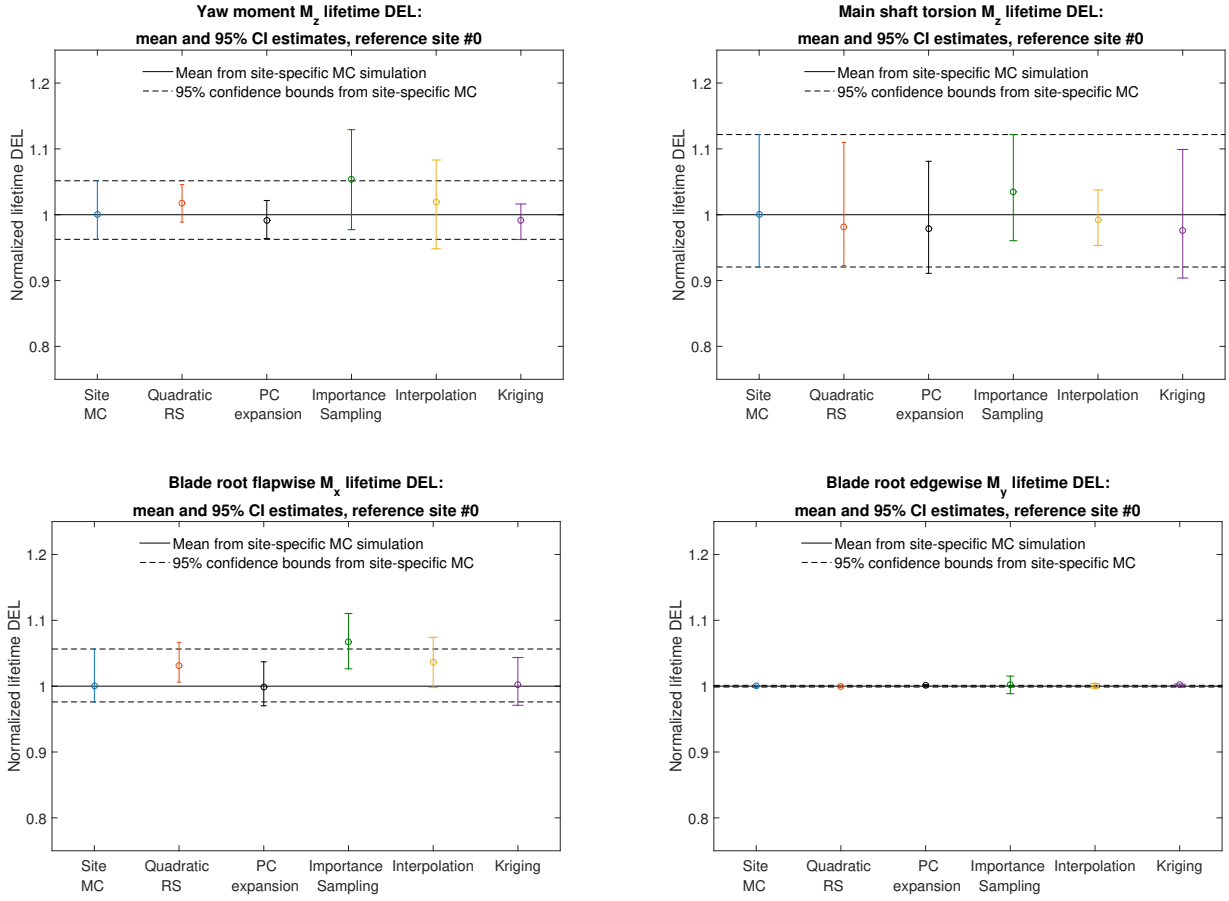

Figure 14. Comparison of predictions of the lifetime damage-equivalent loads for six different estimation approaches. All values are normalized with respect to the mean estimate from a site-specific Monte Carlo simulation.

in eq. (20) resulting from a point-by-point comparison of load predictions from a reduced-order model against actual reference values. The reference values are the results from the site-specific aeroelastic load simulations for reference site 0. At each sample point, the reference value y_i represents the mean DEL from all turbulence seeds simulated with these conditions. The values of the NRMS error for site 0 for Kriging, RS, and PCE-based load predictions are listed in Table 8. Figure 15 shows a one-to-one comparison for a short sequence of sample points.

The RMS error analysis reveals a slightly different picture. In contrast to the lifetime DEL where the Kriging, PCE and RS methods showed very similar results, the RMS error of the quadratic RS is for some channels about twice the RMS error of the other two approaches.

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Table 3. Lifetime damage-equivalent fatigue load predictions obtained using 6-dimensional Polynomial Chaos Expansion of order 6, relative to predictions from site-specific Monte Carlo simulations. Load channel abbreviations are the following: TB: tower base; TT: tower top; MS: main shaft; BR: blade root. Loading directions consist of M_x: fore-aft (flapwise) bending, M_y: side-side (edgewise) bending, and M_z: torsion.

POLYNOMIAL CHAOS EXPANSION								
	Load channels							
Site number	TB M_x	TB M_y	TT M_x	TT M_y	TT M_z	MS M_z	BR M_x	BR M_y
Site 0	1.009	0.955	1.001	1.008	1.000	1.002	1.019	1.002
Site 1	0.950	0.917	0.979	0.998	1.009	1.007	1.028	1.000
Site 2	0.977	0.942	0.997	1.005	1.010	1.016	1.003	1.001
Site 3	0.954	0.917	0.984	1.004	0.997	1.013	1.003	1.000
Site 4	1.002	0.957	0.999	0.994	1.003	1.009	1.001	1.001
Site 5	0.906	0.933	0.943	0.958	0.950	0.988	0.984	1.000
Site 6	0.938	0.920	0.957	0.978	0.976	0.986	1.015	0.998
IEC IA	0.972	0.936	0.972	1.013	0.984	1.050	0.990	0.996
IEC IIB	0.972	0.929	0.972	1.015	0.987	1.052	0.994	0.997
IEC IIIC	0.977	0.939	0.974	1.024	0.991	1.060	0.996	0.997
Mean	0.966	0.934	0.978	1.000	0.991	1.018	1.003	0.999
Std.Dev	0.030	0.014	0.019	0.019	0.018	0.026	0.014	0.002

5.3 Variable sensitivities

As described earlier in Section 4.1.2, we consider variable sensitivities (i.e. the influence of input variables on the variance of the outcome) in terms of Sobol indices. By definition the Sobol indices will be dependent on the variance of input variables, meaning that for one and the same model, the Sobol indices will be varying under different (site-specific) input variable distributions. Taking this into account, we evaluate the Sobol indices for the two types of joint variable distributions used in this study - 1) a site-specific distribution, and 2) the uniform, bounded joint distribution used to generate the database with high-fidelity load simulations. The Sobol indices for the high-fidelity load database variable range are listed in Table 9, while the indices for the site-specific distribution corresponding to reference site 0 are listed in Table 10. The two tables show similar results - the mean wind speed and the turbulence are the most important factors affecting both fatigue and extreme loads. Other two variables which show a smaller but still noticeable influence are the wind shear α, and the Mann model turbulence length scale L. The effect of wind shear is pronounced mainly for blade root loads which can be explained by the rotation of the blades which, if subjected to wind shear, will experience cyclic changes in wind velocity. The effect of Mann model Γ, veer, yaw, tilt, and air density within the chosen variable ranges seems to be minimal, especially for fatigue loads.

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Table 4. Lifetime damage-equivalent fatigue load predictions obtained using the Universal Kriging approach with a 6-dimensional Polynomial Chaos Expansion of order 6 used as basis functions. The results are compared against site-specific Monte Carlo simulations. Load channel abbreviations are the following: TB: tower base; TT: tower top; MS: main shaft; BR: blade root. Loading directions consist of M_x: fore-aft (flapwise) bending, M_y: side-side (edgewise) bending, and M_z: torsion.

UNIVERSAL KRIGING								
	Load channels							
Site number	TB M_x	TB M_y	TT M_x	TT M_y	TT M_z	MS M_z	BR M_x	BR M_y
Site 0	1.017	1.008	1.004	0.998	0.997	0.997	1.024	1.002
Site 1	0.953	0.941	0.992	0.993	1.010	0.980	1.029	1.001
Site 2	0.984	0.981	1.007	1.003	1.013	0.984	1.011	1.002
Site 3	0.959	0.953	0.996	0.998	0.998	0.982	1.009	1.001
Site 4	1.012	1.014	1.007	1.005	1.009	0.993	1.008	1.003
Site 5	0.911	0.943	0.952	0.957	0.946	0.960	0.976	0.999
Site 6	0.938	0.929	0.967	0.979	0.979	0.964	1.015	0.999
IEC IA	0.984	0.960	0.993	1.005	0.987	1.000	0.996	0.998
IEC IIB	0.987	0.966	0.989	1.021	0.992	1.025	1.004	0.999
IEC IIIC	0.976	0.952	0.985	1.024	0.994	1.048	1.012	0.999
Mean	0.972	0.965	0.989	0.998	0.992	0.993	1.008	1.000
Std.Dev	0.033	0.028	0.018	0.019	0.020	0.027	0.015	0.002

5.4 Mean extreme loads

Fatigue loads are the main quantity of interest in the present study, however the methods discussed can also be applied to other types of statistics as long as the load simulations carry the necessary information and cover the input variable space. For example, the conditions specified in the extreme-turbulence load case scenario DLC1.3 in IEC61400-1 class C and B are within the envelope of the high-fidelity database. DLC1.3 requires taking the mean of the extremes from all seeds simulated at a specific condition - meaning that the requirements for the reduced-order models are very similar to the case with fatigue loads. As a demonstration for the feasibility of these calculations, Figure 16 shows prediction of the mean of extremes for the DLC1.3 load case for IEC class IB, carried out by calibrating a PCE and a quadratic RS according to the same procedures as already defined for the fatigue DEL.

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Table 5. Lifetime damage-equivalent fatigue load predictions obtained using 6-dimensional Quadratic Response Surface, relative to predictions from site-specific Monte Carlo simulations. Load channel abbreviations are the following: TB: tower base; TT: tower top; MS: main shaft; BR: blade root. Loading directions consist of M_x: fore-aft (flapwise) bending, M_y: side-side (edgewise) bending, and M_z: torsion.

QUADRATIC RESPONSE SURFACE								
	Load channels							
Site number	TB M_x	TB M_y	TT M_x	TT M_y	TT M_z	MS M_z	BR M_x	BR M_y
Site 0	1.092	1.031	1.009	1.013	1.009	1.009	1.035	1.001
Site 1	1.055	0.985	0.975	1.034	1.033	1.076	1.038	0.998
Site 2	1.029	0.983	0.966	1.041	1.028	1.081	1.012	0.998
Site 3	1.007	0.967	0.961	1.036	1.015	1.074	1.018	0.997
Site 4	1.038	0.996	0.956	1.044	1.029	1.085	1.008	0.997
Site 5	0.988	0.938	0.954	1.001	1.003	1.056	1.017	0.996
Site 6	1.030	0.960	0.968	1.032	1.015	1.069	1.040	0.997
IEC IA	1.013	0.965	0.952	1.032	1.002	1.098	1.016	0.989
IEC IIB	1.026	0.969	0.955	1.040	1.002	1.102	1.012	0.995
IEC IIIC	1.059	1.010	0.963	1.050	1.006	1.103	1.015	0.998
Mean	1.034	0.980	0.966	1.032	1.014	1.075	1.021	0.996
Std.Dev	0.029	0.027	0.017	0.015	0.012	0.028	0.012	0.003

6 Discussion and conclusions

6.1 Discussion

The previous sections of this paper described the procedure for estimating site-specific lifetime damage-equivalent loads using several simplified model techniques which were applied for ten sites. Based on the site-specific lifetime DEL comparisons, three models showed to be viable (robust and sufficiently accurate) choices for quick site-specific load estimation: the PC expansion, Kriging, and the quadratic RS. When estimating lifetime DEL, these methods showed approximately equal levels of uncertainty. However, in the one-to-one comparisons the quadratic RS model showed larger error, especially for sample points corresponding to more extreme combinations of environmental conditions. This is due to the lower order and the relatively small number of calibration points of the quadratic RS, which means that the model accuracy decreases in the sampling space away from the calibration points, especially if there is any extrapolation. This inaccuracy is reflected in the RMS-error from one-to-one comparisons, but is less obvious in the lifetime fatigue load computations which average out errors with zero mean. The universal Kriging model demonstrated the overall lowest uncertainty both in sample-to-sample comparisons and in lifetime DEL computations. This is to be expected since the Kriging employs an already well-performing model (the PCE), and combines it with an interpolation scheme which reduces the uncertainty even further. However, in most cases the observed improvement over a pure PCE is not significant. This indicates that the sources of the remaining uncertainty are outside the

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Table 6. Lifetime damage-equivalent fatigue load predictions obtained using Importance Sampling of points from a 6-dimensional high-fidelity load database, relative to predictions from site-specific Monte Carlo simulations. Load channel abbreviations are the following: TB: tower base; TT: tower top; MS: main shaft; BR: blade root. Loading directions consist of M_x: fore-aft (flapwise) bending, M_y: side-side (edgewise) bending, and M_z: torsion.

IMPORTANCE SAMPLING								
	Load channels							
Site number	TB M_x	TB M_y	TT M_x	TT M_y	TT M_z	MS M_z	BR M_x	BR M_y
Site 0	1.059	1.098	1.015	1.004	1.039	1.042	1.050	1.002
Site 1	0.893	0.849	0.889	1.174	0.974	1.333	1.089	0.992
Site 2	0.884	0.846	0.883	1.176	0.972	1.296	1.079	0.993
Site 3	0.914	0.867	0.884	1.166	0.958	1.315	1.069	0.990
Site 4	0.944	0.884	0.882	1.182	0.970	1.316	1.060	0.992
Site 5	0.762	0.771	0.806	1.065	0.838	1.178	0.965	0.985
Site 6	0.856	0.810	0.867	1.155	0.944	1.303	1.077	0.988
IEC IA	0.779	0.856	0.797	1.007	0.873	1.221	1.141	0.980
IEC IIB	0.753	0.864	0.797	1.032	0.874	1.238	1.199	0.993
IEC IIIC	0.746	0.932	0.802	1.064	0.875	1.268	1.277	1.001
Mean	0.859	0.878	0.862	1.102	0.932	1.251	1.100	0.992
Std.Dev	0.101	0.088	0.067	0.075	0.063	0.088	0.086	0.007

models—e.g. the seed-to-seed turbulence variations: the models being calibrated with turbulence realizations different from the ones used to compute the reference site-specific loads. As a result the trend function has the main contribution to the Kriging estimator and the influence of the Gaussian-field interpolation is minimal. A drawback of the Kriging model with respect to the other techniques is the larger computational demands due to the need of computing correlation matrices and the use of the training sample for each new evaluation.

For all site-specific load assessment methods discussed, the estimations are trustworthy only within the bounds of the variable space used for model calibration - extrapolation is either not possible, or may lead to unpredictable results. It is therefore important to ensure that the site-specific distributions used for load assessment are not outside the bounds of validity of the load estimation model.

The variable bounds presented in this paper are based on a certain degree of consideration of atmospheric physics employed in the relationships between wind speed, turbulence, wind shear, wind veer and turbulence length scale. The primary scope is to encompass the ranges of conditions relevant for fatigue load analysis, and the currently suggested variable bounds include all normal-turbulence (NTM) classes. However, in some situations it may be more practical to choose other bound definitions - e.g. for consistency with the IEC 61400-1 definitions of the extreme turbulence models, where the currently suggested bounds do not include ETM class A.

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Table 7. Lifetime damage-equivalent fatigue load predictions obtained using nearest-neighbor interpolation in a 6-dimensional high-fidelity load database, relative to predictions from site-specific Monte Carlo simulations. Load channel abbreviations are the following: TB: tower base; TT: tower top; MS: main shaft; BR: blade root. Loading directions consist of M_x: fore-aft (flapwise) bending, M_y: side-side (edgewise) bending, and M_z: torsion.

NEAREST-NEIGHBOR INTERPOLATION								
	Load channels							
Site number	TB M_x	TB M_y	TT M_x	TT M_y	TT M_z	MS M_z	BR M_x	BR M_y
Site 0	1.077	1.093	1.018	1.029	1.036	1.052	1.056	1.000
Site 1	1.002	1.019	0.987	1.039	1.012	1.141	1.043	0.998
Site 2	0.993	1.022	0.987	1.013	1.019	1.080	1.032	0.998
Site 3	0.983	1.030	0.985	1.019	1.000	1.106	1.021	0.998
Site 4	1.004	1.031	0.968	1.010	1.008	1.067	1.014	0.997
Site 5	0.927	0.953	0.922	1.025	0.903	1.097	0.945	0.994
Site 6	0.981	0.985	0.960	1.063	0.992	1.166	1.037	0.997
IEC IA	0.832	0.900	0.895	0.906	0.901	1.042	0.951	0.985
IEC IIB	0.839	0.934	0.894	0.892	0.929	0.987	0.962	0.988
IEC IIIC	0.867	0.963	0.898	0.898	0.921	0.927	0.953	0.990
Mean	0.951	0.993	0.951	0.989	0.972	1.066	1.001	0.994
Std.Dev	0.081	0.057	0.045	0.064	0.052	0.070	0.044	0.005

Table 8. Root mean square error characterizing the difference between aeroelastic simulations and reduced-order models. Load channel abbreviations are the following: TB: tower base; TT: tower top; MS: main shaft; BR: blade root. Loading directions consist of M_x: fore-aft (flapwise) bending, M_y: side-side (edgewise) bending, and M_z: torsion.

NORMALIZED RMS ERROR - SITE 0								
	Load channels							
Prediction model	TB M_x	TB M_y	TT M_x	TT M_y	TT M_z	MS M_z	BR M_x	BR M_y
Quadratic RS	0.0452	0.1404	0.1981	0.2612	0.0644	0.2280	0.1504	0.0098
PC expansion	0.0362	0.0955	0.1019	0.2089	0.0362	0.1530	0.0620	0.0084
Kriging	0.0334	0.0706	0.0837	0.1761	0.0368	0.1072	0.0519	0.0083

For the more advanced methods like PCE and Kriging, there is a practical limitation of the number of training points to be used on a single-computer setup. For a PCE the practical limit is mainly subject to memory availability when assembling and inverting the information matrix, and for a PCE of order 6 and with 9 dimensions this limit is on the order of 1–2·10^4 points on a typical desktop computer. For Kriging, the computing time also plays a role: although a similar number of training points could be stored in memory as for the PCE, the computational time is much longer and the practical limit of training points for

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

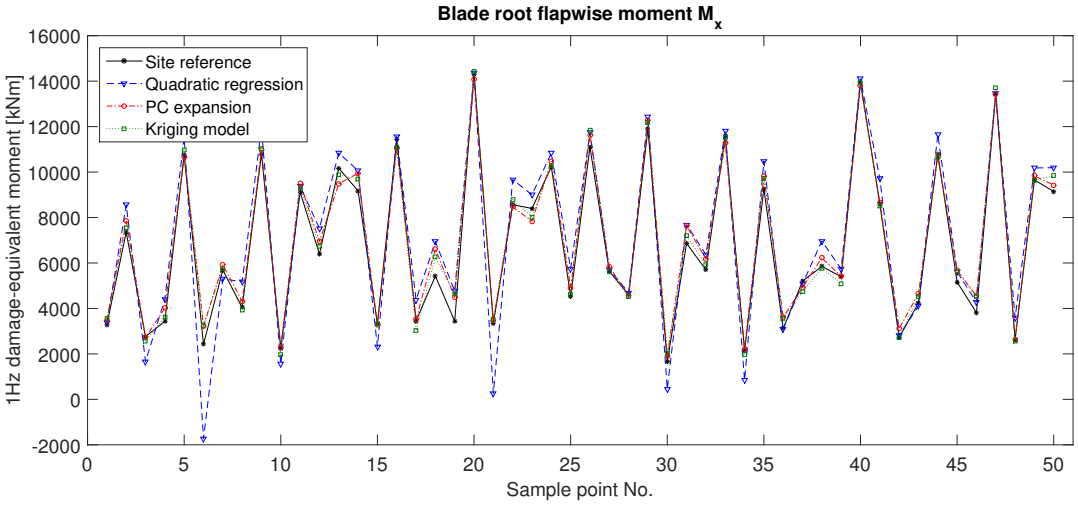

Figure 15. Sample-by-sample comparison of the 1Hz damage-equivalent load predictions, for two load prediction methods - quadratic Response Surface, and Polynomial Chaos expansion, compared against site-specific Monte Carlo simulation. Each point on the x-axis represents a sample point from the site-specific distribution of input variables for reference site 0, and the y-axis represents the estimated mean 1Hz-equivalent load.

most applications is less than for the PCE. However, as it was shown in Sections 4.1 and 5, a training sample of 10^4 points or even half of that should be sufficient for most applications in site-specific load prediction.

Considering the overall merits of the load prediction methods analyzed, the PCE provided an accurate and robust performance. The Kriging approach showed slightly better accuracy but at the expense of increased computational demands. Taking this together with the other useful properties of the PCE - e.g. the orthogonality allowing for creating sparse models or doing variance-based sensitivity analysis, we consider the PCE as the most useful method overall.

In addition to the surrogate models presented in this paper, Artificial Neural Networks (ANNs) are interesting alternative candidates. ANN (see Goodfellow et al., 2016) are machine learning models which are gaining large popularity due to their flexibility and history of successful applications in many problems. It is very likely that a deep Neural Network model can provide similar output quality and performance than the methods described in the present study. This is therefore a matter worth of future research. However, the PCE-based models may have a practical advantage over ANNs due to the "white-box" features such known variance contributions and the possibility of obtaining analytical derivatives which is important for applications to optimization problems.

The results from the site validations showed that for the majority of sites and load channels, the simplified load assessment techniques can predict the site-specific lifetime fatigue loads to within $\approx 5\%$ accuracy. However, it should be noted that this accuracy is relative to full-fidelity load simulations, and not necessarily to the actual site conditions, where additional uncertainties can lead to larger errors. The procedures demonstrated in this study are thus very suitable for carrying out quick site

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

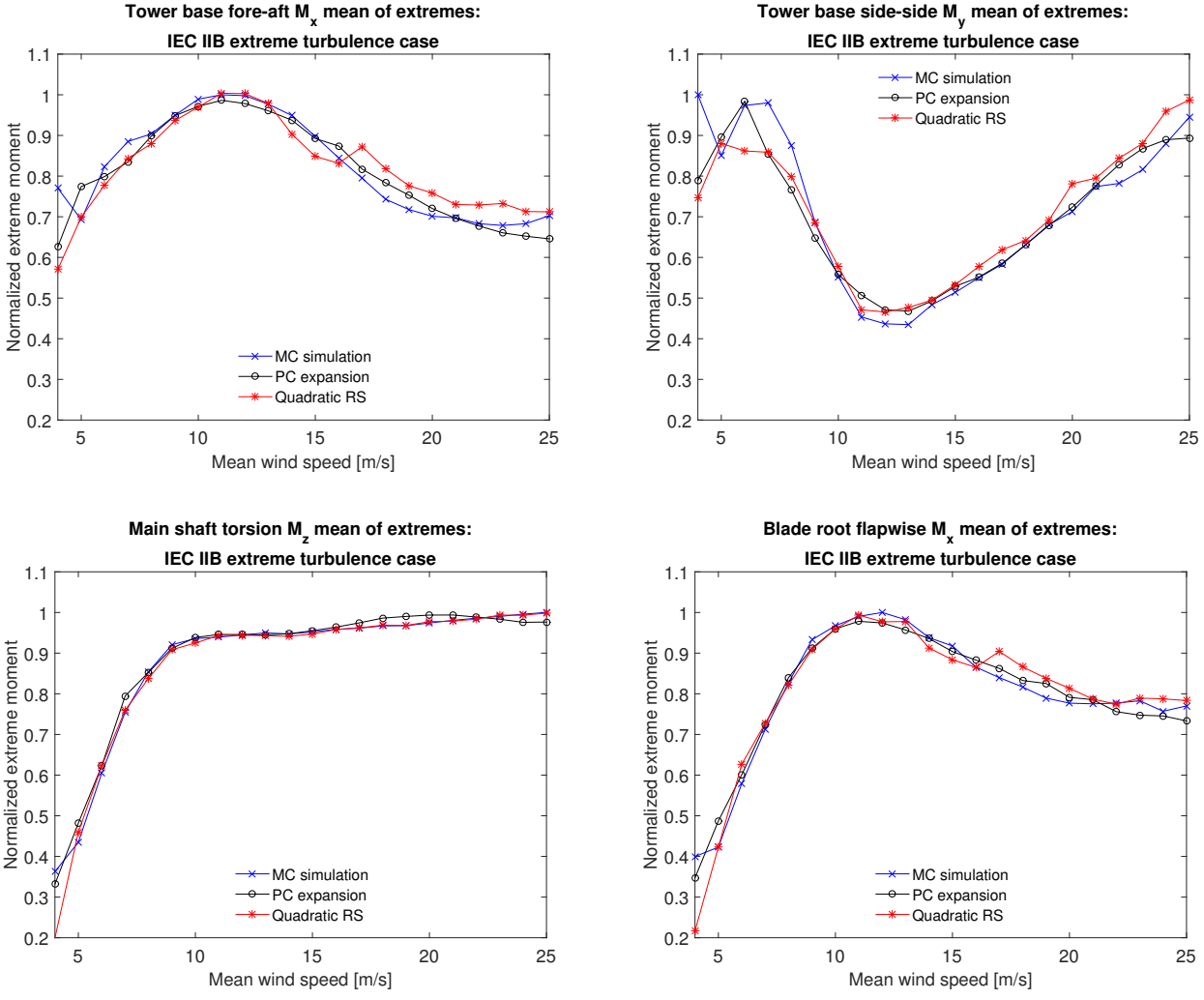

Figure 16. Comparison of predictions of the mean of extremes obtained for the extreme turbulence load case DLC1.3 for IEC turbulence class B conditions. MC simulation (blue lines with crosses) denotes the reference values obtained with full-fidelity aeroelastic simulations.

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Table 9. PCE-based Sobol sensitivity indices for the high-fidelity load database variable ranges.

FATIGUE									
	Variables								
Load channel	U	σ_u	α	L	Γ	$\Delta\varphi_h$	$\bar{\varphi}_h$	$\bar{\varphi}_v$	ρ
Tower base fore-aft moment M_x	0.42	0.65	0.01	0.03	0.02	0.01	0.00	0.00	0.01
Tower base side-side moment M_y	0.62	0.42	0.05	0.04	0.04	0.02	0.02	0.02	0.02
Blade root flapwise moment M_x	0.20	0.64	0.19	0.02	0.01	0.00	0.01	0.00	0.02
Blade root edgewise moment M_y	0.22	0.54	0.25	0.05	0.03	0.01	0.01	0.03	0.01
Tower top yaw moment M_z	0.14	0.85	0.00	0.02	0.01	0.00	0.00	0.00	0.01
Main shaft torsion M_z	0.48	0.53	0.01	0.06	0.01	0.01	0.01	0.01	0.01
EXTREMES									
	Variables								
Load channel	U	σ_u	α	L	Γ	$\Delta\varphi_h$	$\bar{\varphi}_h$	$\bar{\varphi}_v$	ρ
Tower base fore-aft moment M_x	0.52	0.51	0.04	0.04	0.04	0.03	0.03	0.03	0.05
Tower base side-side moment M_y	0.46	0.55	0.13	0.07	0.07	0.05	0.04	0.04	0.04
Blade root flapwise moment M_x	0.59	0.38	0.07	0.04	0.03	0.02	0.02	0.02	0.04
Blade root edgewise moment M_y	0.65	0.46	0.10	0.07	0.08	0.06	0.06	0.06	0.06
Tower top yaw moment M_z	0.18	0.71	0.16	0.02	0.02	0.01	0.02	0.02	0.02
Main shaft torsion M_z	0.86	0.20	0.02	0.03	0.02	0.02	0.02	0.02	0.02

feasibility assessments, which can in a timely fashion aid the decision on whether to discard a given site as unfeasible, or e.g. make additional high-fidelity computations or more measurements of site conditions. The same procedure, but with additional variables (e.g. 3 variables for wake-induced effects as in (Galinos et al., 2016)) may also be useful as objective function or constraint in farm optimization problems.

6.2 Conclusions

In the present work we defined and demonstrated a procedure for quick assessment of site-specific lifetime fatigue loads using surrogate models calibrated by means of a database with high-fidelity load simulations. The performance of polynomial chaos expansion, quadratic response surface, universal Kriging, importance sampling, and nearest-neighbor interpolation in predicting site-specific lifetime fatigue loads was assessed. Practical bounds of variation were defined for nine environmental variables and their effect on the lifetime fatigue loads was studied. The study led to the following main conclusions:

- The variable sensitivity analysis showed that mean wind speed and turbulence are the factors having the highest influence on fatigue and extreme loads. The Mann turbulence length scale and the wind shear exponent were also found to have an influence, and for the wind shear the effect is more pronounced for rotating components such as blades. Within the

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Table 10. Site-specific Sobol sensitivity indices derived for Site 0 using Monte Carlo simulation from a PCE.

FATIGUE					
	Variables				
Load channel	U	σ_u	α	L	Γ
Tower base fore-aft moment M_x	0.08	1.32	0.07	0.18	0.09
Tower base side-side moment M_y	0.90	0.09	0.07	0.23	0.13
Blade root flapwise moment M_x	0.42	0.38	0.05	0.01	0.01
Blade root edgewise moment M_y	0.43	0.18	0.26	0.22	0.11
Tower top yaw moment M_z	0.23	0.53	0.01	0.08	0.01
Main shaft torsion M_z	0.47	0.36	0.06	0.03	0.07
EXTREMES					
	Variables				
Load channel	U	σ_u	α	L	Γ
Tower base fore-aft moment M_x	0.68	0.44	0.03	0.02	0.01
Tower base side-side moment M_y	0.52	0.40	0.07	0.22	0.17
Blade root flapwise moment M_x	0.76	0.21	0.02	0.02	0.01
Blade root edgewise moment M_y	0.62	0.05	0.10	0.05	0.05
Tower top yaw moment M_z	0.15	0.45	0.08	0.08	0.02
Main shaft torsion M_z	0.81	0.25	0.04	0.04	0.00

studied ranges of variation, the Mann turbulence parameter Γ, wind veer, yaw angle, tilt angle, and air density, were found to have small or negligible effect on the loads.

- The best performing models had errors of less than 5% for most sites and load channels, which is in the same order of magnitude as the variations due to realization-to-realization uncertainty.

- A universal Kriging model with a polynomial chaos expansion used as a trend function achieved the most accurate predictions, but also required the longest computing times.

- A polynomial chaos expansion with Legendre basis polynomials was concluded to be the approach with best overall performance.

- The procedures demonstrated in this study are well suited for carrying out quick site feasibility assessments with a particular wind turbine model.

Competing interests. No competing interests are present

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

Acknowledgements. The work reported in this manuscript was carried out as part of the Wind2Loads internal project at the Technical University of Denmark, Department of Wind Energy. The authors thank their colleagues for the valuable inputs and support.

Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2018-18
Manuscript under review for journal Wind Energ. Sci.
Discussion started: 2 March 2018
© Author(s) 2018. CC BY 4.0 License.

References

Bak, C., Zahle, F., Bitsche, R., Kim, T., Yde, A., Henriksen, L. C., Natarajan, A., and Hansen, M.: Description of the DTU 10MW reference wind turbine, Tech. Rep. I-0092, Technical University of Denmark, Department of Wind Energy, 2013.

Berg, J., Natarajan, A., Mann, J., and Patton, E.: Gaussian vs non-Gaussian turbulence: impact on wind turbine loads, Wind Energy, 19, 1975–1989, 2016.

Borraccino, A., Schlipf, D., Haizmann, F., and Wagner, R.: Wind field reconstruction from nacelle-mounted lidar short-range measurements, Wind Energy Science, 2, 269–283, https://doi.org/10.5194/wes-2-269-2017, https://www.wind-energ-sci.net/2/269/2017/, 2017.

Caflisch, R. E.: Monte Carlo and Quasi-Monte Carlo methods, Acta Numerica, pp. 1–49, 1998.

Clifton, A., Schreck, S., Scott, G., Kelley, N., and Lundquist, J.: Turbine Inflow Characterization at the National Wind Technology Center, J. of Solar Energy Engineering, 135, 031 017, https://doi.org/doi:10.1115/1.4024068, 2013.

Dimitrov, N., Natarajan, A., and Kelly, M.: Model of wind shear conditional on turbulence and its impact on wind turbine loads, Wind Energy, 18, 1917–1931, 2015.

Dimitrov, N., Natarajan, A., and Mann, J.: Effect of Normal and Extreme turbulence spectral parameters on wind turbine loads, Renewable Energy, 101, 1180–1193, 2017.

Ditlevsen, O. and Madsen, H. O.: Structural Reliability Methods, Wiley, 1996.

Efron, B.: Bootstrap methods: another look at the jackknife, Annals of Statistics, 7, 1–26, 1979.

Galinos, C., Dimitrov, N., Larsen, T. J., Natarajan, A., and Hansen, K. S.: Mapping Wind Farm Loads and Power Production - A Case Study on Horns Rev 1, Journal of Physics: Conference Series, 753, 032 010, 2016.

Ghanem, R. G. and Spanos, P. D.: Stochastic finite elements - a spectral approach, Springer, Berlin, 1991.

Goodfellow, I., Bengio, Y., and Courville, A.: Deep Learning, MIT Press, http://www.deeplearningbook.org, 2016.

Hansen, B. O., Courtney, M., and Mortensen, N. G.: Wind Resource Assessment – Østerild National Test Centre for Large Wind Turbines, Tech. Rep. E-0052, Technical University of Denmark, Department of Wind Energy, 2014.

Hansen, M. H. and Henriksen, L. C.: Basic DTU Wind Energy Controller, Tech. Rep. E-0028, Technical University of Denmark, Department of Wind Energy, 2013.

IEC: International Standard IEC61400-1: Wind Turbines - Part 1: Design Guidelines, 2005.

Kashef, T. and Winterstein, S. R.: Relating turbulence to wind turbine blade loads: parametric study with multiple regression analysis, Solar Energy Engineering, 121, 156–161, 1999.

Kelly, M.: From standard wind measurements to spectral characterization: turbulence length scale and distribution, Wind Energy Science, 3, xxx–yyy, 2018.

Kelly, M. and van der Laan, P.: Veer and shear relations: beyond Ekman, towards statistical characterization, Quarterly Journal of the Royal Meteorological Society, p. in preparation, 2018.

Kelly, M., Larsen, G., Natarajan, A., and Dimitrov, N.: Probabilistic meteorological characterization for turbine loads, Journal of Physics: Conference Series (online), 524, 012 076, 2014.

Larsen, T. J. and Hansen, A. M.: How to HAWC2, the user's manual, Tech. Rep. R-1597, Technical University of Denmark, Department of Wind Energy, 2012.

Lataniotis, C., Marelli, S., and Sudret, B.: UQLab user manual – Kriging (Gaussian process modelling), Tech. rep., Chair of Risk, Safety & Uncertainty Quantification, ETH Zurich, report UQLab-V0.9-105, 2015.

© Author(s) 2018. CC BY 4.0 License.

Liu, P. L. and Der Kiureghian, A.: Multivariate distribution models with prescribed marginals and covariances, Probabilistic Engineering Mechanics, 1, 105–112, 1986.

Mann, J.: The spatial structure of neutral atmospheric surface-layer turbulence, Journal of Fluid Mechanics, 273, 141–168, 1994.

Mann, J.: The spectral velocity tensor in moderately complex terrain, J. Wind Eng., 88, 153–169, 2000.

Manuel, L., Veers, P. S., and Winterstein, S. R.: Parametric models for estimating wind turbine fatigue loads for design, Solar Energy Engineering, 123, 346–355, 2001.

Mckay, M. D., Beckman, R. J., and Conover, W. J.: A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output From a Computer Code, Technometrics, 42, 55–61, 2000.

Morokoff, W. J. and Caflisch, R. E.: Quasi-Monte Carlo Integration, Computational Physics, 122, 218–230, 1995.

Mouzakis, F., Morfiadakis, E., and Dellaportas, P.: Fatigue loading parameter identification of a wind turbine operating in complex terrain, Wind Engineering and Industrial Aerodynamics, 82, 69–88, 1999.

Murcia, H. P., Natarajan, A., Dimitrov, N., Sørensen, J. D., Graf, P., and Kim, T.: Uncertainty propagation through an aeroelastic wind turbine model using polynomial surrogates, Submitted to Renewable Energy, 2016.

Peña, A., Floors, R., Sathe, A., Gryning, S.-E., Wagner, R., Courtney, M. S., Larsén, X. G., Hahmann, A. N., and Hasager, C. B.: Ten Years of Boundary-Layer and Wind-Power Meteorology at Høvsøre, Denmark, Boundary-Layer Meteorology, 158, 1–26, https://doi.org/10.1007/s10546-015-0079-8, https://doi.org/10.1007/s10546-015-0079-8, 2016.

Rosenblatt, M.: Remarks on a multivariate transformation, Annals of Mathematical Statistics, 23, 470–472, 1952.

Rychlik, I.: A New Definition of the Rainflow Cycle Counting Method, International Journal of Fatigue, 9, 119–121, 1987.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P.: Design and analysis of computer experiments, Statistical Science, 4, 409–423, 1989.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S.: Global Sensitivity Analysis: The Primer, Wiley, 2008.

Santher, T. J., Williams, B. J., and Notz, W. I.: The design and analysis of computer experiments, Springer-Verlag, 2003.

Sudret, B.: Global sensitivity analysis using polynomial chaos expansions, Reliability Engineering and Systems Safety, 93, 964–979, 2008.

Tibshirani, R.: Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society, Series B, 58, 267–288, 1996.

Toft, H. S., Svenningsen, L., Moser, W., Sørensen, J. D., and Thørgensen, M. L.: Assessment of Wind Turbine Structural Integrity using Response Surface Methodology, Engineering Structures, 106, 471–483, 2016.

Xiu, D. and Karniadakis, G. E.: The Wiener-Askey polynomial chaos for stochastic differential equations, Journal of Scientific Computing, 191, 4927–4948, 2002.