From wind to loads: wind turbine site-specific load estimation with surrogate models trained on high-fidelity load databases

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

DTU Wind Energy, Technical University of Denmark
Risø Campus, 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 simplified load mapping functions (surrogate models), trained by means of a database with high-fidelity load simulations. The performance of five surrogate models is assessed by comparing site-specific lifetime fatigue load predictions at ten sites using an aeroelastic model of the DTU 10MW reference wind turbine. The surrogate methods are 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 provides 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.

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 for wind energy applications. Kashef and Winterstein (1999) , Manuel et al. (2001) use probabilistic expansions based on statistical moments. Simple multivariate regression models of first order are employed by Mouzakis et al. (1999), Stewart (2014), while in Toft et al. (2016) a second-order response surface is used. Another response surface approach using artificial neural networks is described in Müller et al.

(2017). Polynomial chaos expansion (PCE) is employed by Ganesh and Gupta (2013) for blade load prediction, albeit on a very simple structural representation. Teixeira et al. (2017) use a Kriging surrogate model to map the load variations with respect to offshore environmental conditions. Other relevant studies use some of the methodologies which represent specific analysis steps shown in the present work. These include Hübler et al. (2017) where variance-based sensitivity analysis is employed, Yang et al. (2015) where Kriging is used to enable efficient implementation of reliability-based design optimization, and Murcia et al. (2018) where polynomial chaos expansions are used to carry out uncertainty propagation. In the latter, the model training sample is generated using a Monte Carlo simulation with a quasi-random sequence, a technique which is also employed in Müller and Cheng (2018) and Graf et al. (2016). An alternative to the surrogate modelling approach discussed in this paper could be the load set reduction, as described in e.g. Häfele et al. (2018); Zwick and Muskulus (2016) which also reduces the number of simulations required. This approach however still requires carrying out high-fidelity simulations which leads to using more time for simulation set-up, computations and post-processing, while with a surrogate model the lifetime equivalent load computation takes typically less than a minute on a regular personal computer. The studies most in line with the scope of the present paper are those by Müller et al. (2017), Teixeira et al. (2017) and Toft et al. (2016). The former two employ advanced surrogate modelling techniques (artificial neural networks and Kriging respectively), however the experimental designs are relatively small and with limited range of variation for some of the variables, and the discussion does not focus on the practical problem of computing lifetime-equivalent site-specific loads. The computation of site-specific lifetime-equivalent design loads is the main focus in Toft et al. (2016), however with a limited number of variables, and using a low-order quadratic response surface. The vast majority of the studies employ a single surrogate modelling approach, meaning that it has not been possible to directly compare the performance of different approaches.

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 surrogate modelling approach for site feasibility assessment. The study aims at fulfilling the following four 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;

- demonstrate how different surrogate modelling approaches can be successfully employed in the problem, and compare their performance; 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) (Dimitrov et al., 2017; Bak et al., 2013). 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. The loads analysis is based on the DTU 10MW reference wind turbine (Bak et al., 2013) simulated using the Hawc2 software (Larsen and Hansen, 2012).

2 Definition of the surrogate load modelling procedure

2.1 Schematic description

5 Figure 1 shows a schematic overview of the procedure for site-specific load assessment using simplified load mapping functions (here referred to in general as surrogate models). The main advantage of this procedure is that the computationally expensive high-fidelity simulations are only carried out once, during the model training process (top of Figure 1). In the model deployment process (bottom of Figure 1), only the coefficients of the trained surrogate models are used, and a site-specific load evaluation typically takes less than a minute on a standard personal computer.
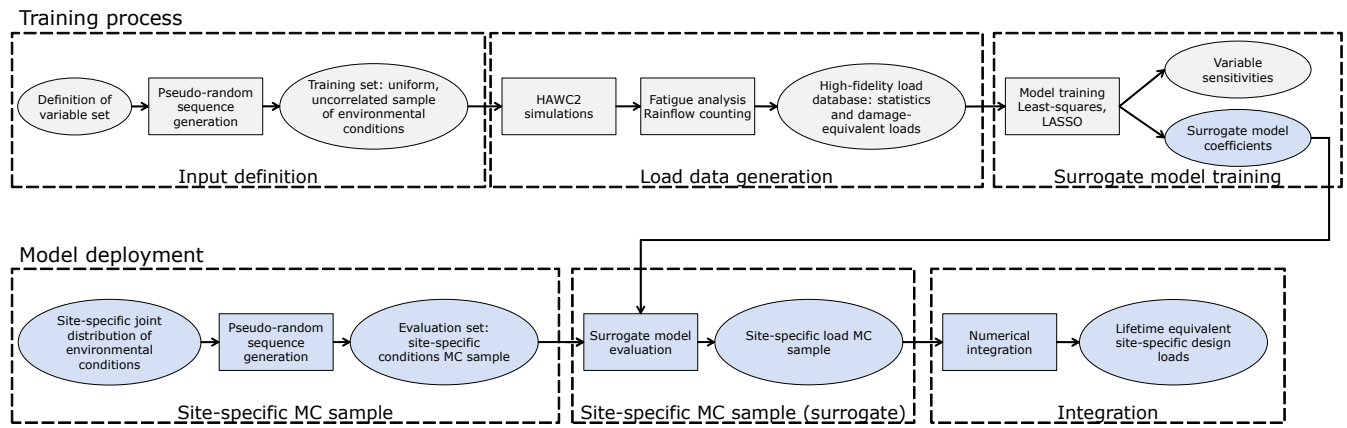
Figure 1. Schematic overview of the site-specific load analysis procedure

10 ### 2.2 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,

15 - 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 $\alpha\varepsilon^{2/3}$;

○ 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 the quantities referred to above are considered in terms of 10-minute average values. All variables, except the variables defining mean inflow direction, are probabilistic and site-dependent in nature. The mean inflow direction variables 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 IEC 61400-1 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 a given combination of 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 the 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 (see Table 1 for details)

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

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 field 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, because it allows more complex conditional dependencies than the Nataf transformation which implies linear correlation. The Rosenblatt transformation maps a vector of n dependent variables $\mathbf{X}$ into a vector of independent components $\mathbf{Y}$ based on conditional relations:

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

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

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

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.

2.3 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.

For the case of building a high-fidelity load database, all variables given in Table 1 except the wind speed are uniform, and only the distribution bounds are conditional on other variables as specified by the 2nd and 3rd columns of the table. The bounds of several variables are conditional on the wind speed, and as shown on Figure 2 they are wider at low wind speeds, meaning that more sample points are needed to cover the space evenly. This dictates that the choice of distribution for the wind speed should provide more samples at low wind speeds. In the present study we have selected a Beta distribution, but other choices as e.g. a truncated Weibull are also feasible.

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 $\sigma_{u,\mathrm{IEC}} +$

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	Distribution
U	$U \geq 4\text{m/s}$	$U \leq 25\text{m/s}$	Beta
σ_u	$\sigma_u \geq 0.025 \cdot U\,(\text{m/s})$	$\sigma_u \leq 0.18\left(6.8 + 0.75U + 3\left(\frac{10}{U}\right)^2\right)(\text{m/s})$	Uniform
α	$\alpha \geq \alpha_{ref,LB} - 0.23\left(\frac{U_{max}}{U}\right)\left(1 - \left(0.4\log\frac{R}{z}\right)^2\right)$	$\alpha \leq \alpha_{ref,UB} + 0.4\left(\frac{R}{z}\right)\left(\frac{U_{max}}{U}\right)$	Uniform
L	$L \geq \max\{7.5\text{m}, \quad (15\text{m}) \cdot |\alpha|^{-2/3}\}$	$L \leq 275\text{m}$	Uniform
Γ	$\Gamma \geq 1$	$\Gamma \leq 5$	Uniform
$\Delta\varphi_h$	$\Delta\varphi_h \geq -0.1D\left(\frac{5}{U}\right)$	$\Delta\varphi_h \leq \min\left\{60°\sin|\phi|, \quad 1.0D\left(\frac{5}{U}\right)^2\right\}$	Uniform
$\bar{\varphi}_h$	$\bar{\varphi}_h \geq -10°$	$\bar{\varphi}_h \leq 10°$	Uniform
$\bar{\varphi}_v$	$\bar{\varphi}_v \geq -10°$	$\bar{\varphi}_v \leq 10°$	Uniform
ρ	$\rho \geq 1.1\text{kg/m}^3$	$\rho \leq 1.35\text{kg/m}^3$	Uniform

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;

- $U_{max} = 25$m/s is the upper bound of the wind speed;

- ϕ is the reference latitude (here chosen as $50°$).

$I_{\text{ref,A}}U_{\text{cut-in}}[1 + (U_{\text{cut-out}}/U)]$ with $\{U_{\text{cut-in}}, U_{\text{cut-out}}\} = \{4,25\}$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

5   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 $L > \max\{7.5\text{m}, (15\text{m})|\alpha|^{-2/3}\}$ 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), as $L \simeq zI_u/\alpha$ (Kelly, 2018). The minimum scale (7.5 m) and proportionality constant (15 m) are set to allow a wide range of conditions (though most sites

10   will likely have a scaling factor 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 (and applicability of Taylor's hypothesis). 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

15   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 $\partial\varphi_h/\partial z$. For the remaining variables, $\bar{\varphi}_h$, $\bar{\varphi}_v$, and ρ, the bounds are chosen arbitrarily such that they are wide enough to encompass the values typically used in a design load basis.

2.4 Sampling procedure

5 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, i.e., surrogate models and response surfaces.

10 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-

15 discrepancy Halton sequence in a quasi-Monte Carlo approach (Caflisch, 1998). 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 quasi-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 quasi-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

20 (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 quasi-Monte Carlo samples should result in about the same error as 10^5 crude Monte Carlo samples. A disadvantage of the quasi-random sequences is that their properties typically deteriorate in high-dimensional problems, where periodicity and correlation between points in different dimensions may appear (Morokoff and Caflisch, 1995). However, such behaviour typically occurs when more than

25 20-25 dimensions are used. In the present problem the dimensionality is limited by the computational requirements of the surrogate models and the aeroelastic simulations used to train them. Therefore the behaviour of quasi-random sequences in high dimensions does not have implications for the present study. The Halton sequence is applied by taking consequentially all points in the quasi-random series without omission and without repetitions, starting from the second point. The first point in the sequence is discarded as it contains zeros (i.e., the lower bounds of the interval $[0, 1]$) in all dimensions, which corresponds

30 to zero joint probability for the input variables $\mathbf{X}$.
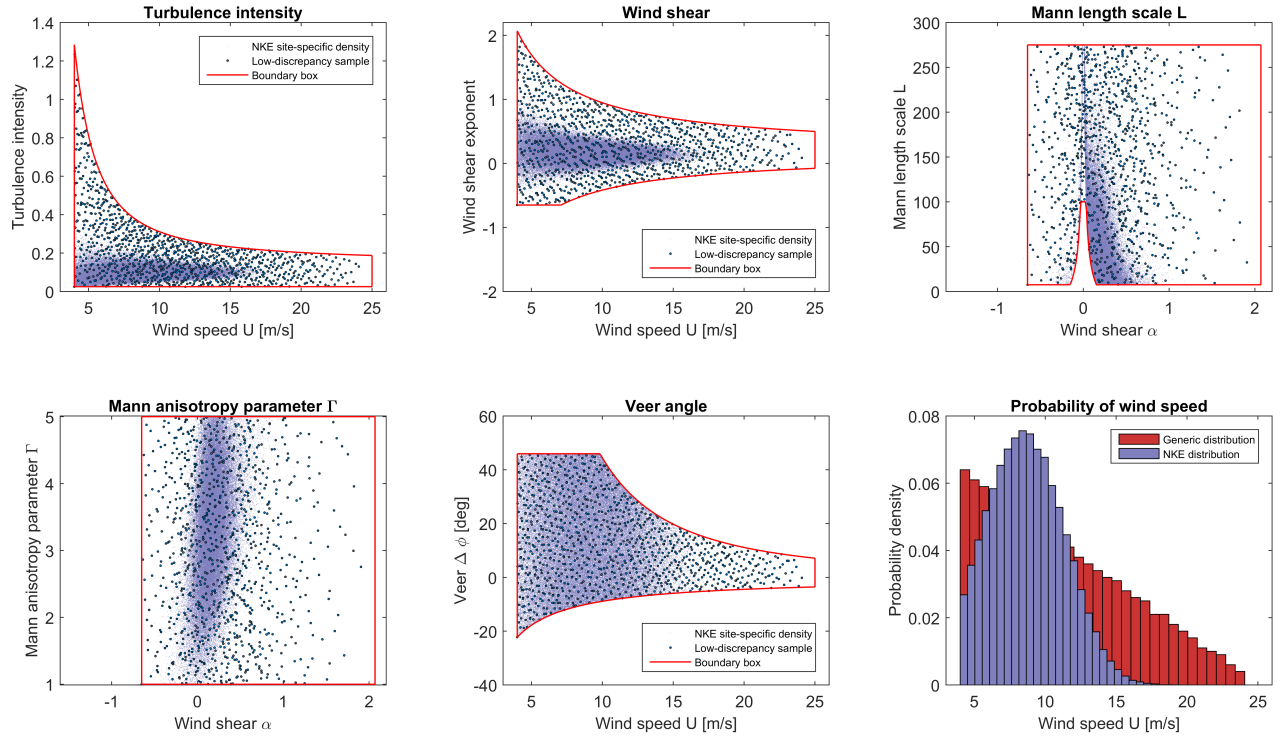
Figure 2. Sample distributions obtained using 1024 low-discrepancy points within a 6-dimensional variable space $\{U, I_u, \alpha, \Delta\phi_h, L, \Gamma\}$. Here U is Beta-distributed, while the other variables are uniformly distributed within their ranges. Solid lines show the sampling space bounds which are curved due to conditional dependencies. Blue shading shows the site-specific variable distribution for the NKE reference site (site 0, c.f. Table 5/Section 6.1).

2.5 Database specification

A large-scale generic load database is generated in order to serve as a training data set for the load mapping functions. The point sampling is done using a Halton low-discrepancy sequence within the 9-dimensional variable space defined in section 2.4 (Figure 2 shows the bounds for the first 6 variables). The database setup is the following:

- Up to 10^4 quasi-random MC sample points in the interval $[0, 1)$ are generated, following a low-discrepancy sequence for obtaining evenly distributed points within the parametric space.

- The physical values of the stochastic variables for all quasi-MC samples are obtained by applying a Rosenblatt transformation using the conditional distribution bounds given in Table 1 and using uniform distribution density, except for the wind speed for which a Beta distribution is used.

- For each sample point, eight simulations, with 3800s duration each, are carried out. The first 200s of the simulations are discarded in order to eliminate simulation run-in time transients, and the output is 3600s (1h) of load time series from each simulation.

- The Mann model simulation parameters (L, Γ, $\alpha\epsilon^{2/3}$) which determine the turbulence intensity are tuned to match the required 10-minute turbulence statistics (1 h statistics are slightly different due to longer sampling time).

- Each 1h time-series is split into six 10-minute series, which on average will have the required statistics. This leads to a total of 48 10-minute time series for each quasi-MC sample point.

- 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).

By choosing to run 1h simulations followed by splitting up of the time series instead of directly simulating 10-minute periods, we want to capture some of the low-frequency fluctuations generated by the Mann model turbulence, especially at larger turbulence length scales. When we generate a longer turbulence box, it includes more of these low-frequency variations, which in fact introduce some degree of non-stationarity when looking at 10-minute windows.

3 Post-processing and analysis

3.1 Time series postprocessing and cycle counting

The main quantities of interest from the load simulation output are the short-term (10-minute) 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 ($N_{\text{ref}} = 600$ for obtaining 1 Hz-equivalent DEL over a 10 min 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

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

3.2 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 eq. (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 ith 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 $p(\mathbf{x}_i) = 1/N$.

3.3 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. This uncertainty is normally taken into account by carrying out load simulations with multiple realizations (seeds) of turbulence inputs. Confidence intervals reflecting such an uncertainty can 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 (confidence interval) bounds for a confidence level c_ℓ are

$$\left\{ CI^-_{S_{\text{eq,lifetime}}}(c_\ell), \, CI^+_{S_{\text{eq,lifetime}}}(c_\ell) \right\} = \left\{ R_{[c_\ell M/2]}, \, R_{[(1-c_\ell/2)M]} \right\} \tag{7}$$

where the square brackets $[x]$ indicate the integer part of x, and $R_{[x]}$ means the value in R with rank equal to $[x]$. In the present study, bootstrapping is applied by generating independent bootstrap samples each with size equal to the entire data set. Both the sample points and the turbulence seed numbers are shuffled, meaning that the resulting confidence intervals should account for both the statistical uncertainty due to finite number of samples, and the uncertainty due to seed-to-seed variation. Note that these two uncertainty types are the only ones assumed, for the confidence intervals; reducing the CI by creating a large number of model realizations does not eliminate other model uncertainties, nor does it remove uncertainties in the input variables.

4 Load mapping functions

In this section we present five different approaches which can be used to map loads from a high-fidelity database to integrated site-specific design loads:

1) Importance sampling,

2) Nearest-neighbor interpolation,

3) Polynomial chaos expansion,

4) Universal Kriging, and

5) Quadratic response surface.

The first two methodologies carry out a direct numerical integration over the high-fidelity database presented in Section 2.5, while the latter three are machine learning models which are trained using the same database. Despite the different nature of the functions, they serve the same purpose and for brevity we will refer to all of them as "surrogate models".

4.1 Importance Sampling

Figure 2 showed the distributions of the first 6 input variables from our high-fidelity database (Section 2.5), along with the site-specific distributions for reference site 0 (c.f. Table 5 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 (Ditlevsen and Madsen, 1996). The IS method and various modifications of it are commonly used in wind energy-related applications (e.g. Choe et al., 2015; Graf et al., 2018). In the classical definition of IS, the integration (importance sampling) function for determining the expected value of a function $g(\mathbf{X})$ is given by

$$\mathbb{E}[g(\mathbf{X})] = \frac{1}{N} \sum_{i=1}^{N} g(\mathbf{X}) \frac{f(\mathbf{X}_i)}{h(\mathbf{X}_i)}, \tag{8}$$

where in our application

- $i = 1 \ldots N$ 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. This could be considered as a non-standard application of the IS approach, because typically the IS sample distribution is chosen to maximize the probability density of the integrand. In the present case, this objective can be satisfied only approximately, and only in cases where the number of IS samples is smaller than the total number of database samples ($N_{IS} < N$). Under these conditions, the importance sampling weights ($f(\mathbf{X_i})/h(\mathbf{X}_i)$ from Eq. 8) can be evaluated for all points in the database. However, the approach adopted in the present paper is to include only the N_{IS} points with the highest weights (as shown in Section 6.2).

4.2 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 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:

$$|x| = \sqrt{\mathbf{D}^T \mathbf{D}} \tag{9}$$

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 4.6).

4.3 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. For the present problem, using a Wiener-Askey Generalized PCE (Xiu and Karniadakis, 2002) employing Legendre polynomials is considered most suitable for any (scaled) variable $\xi \in [-1,1]$. Because Legendre polynomials $P_n(\xi)$ are orthogonal with respect to a uniform probability measure, the PCE can conveniently be applied to the cumulative distribution functions of the variables $\mathbf{X}$ which are defined in the interval $[0,1]$. Then

$$\xi_i = 2F(X_i) - 1, \tag{10}$$

where $F(X_i)$ is the cumulative distribution function of a variable $X_i \in \mathbf{X}$, $i = 1, \ldots, M$. The Legendre polynomial coefficients can be generated using the recurrence relation

$$(n+1)P_{n+1}(\xi) = (2n+1)\xi P_n(\xi) - nP_{n-1}(\xi) \tag{11}$$

where the first two entries, $P_0(\xi) = 1$ and $P_1(\xi) = \xi$, serve for initialization. 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

$$\tilde{S}(\boldsymbol{\xi}) = \sum_{j=0}^{N_p-1} S_j \Psi_{\boldsymbol{\gamma},j}(\boldsymbol{\xi}); \tag{12}$$

here $\Psi_{\boldsymbol{\gamma}}$ are multivariate orthogonal polynomials composed of the product of univariate polynomials having (nonnegative integer) orders defined by the vector $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_M]$, with the total of orders being constrained by the degree: $\sum_{i=1}^{M} \gamma_i \leq p$. The unknown coefficients $S_j \in \mathbf{S} = [S_1, \ldots, S_{Np}]$ need to be determined, and $\boldsymbol{\xi} = [\xi_1, \ldots \xi_M]$ are functions of $\mathbf{X}$ as defined in eq. (10). Training the PCE model amounts to determining the vector of coefficients, $\mathbf{S}$. For a more detailed explanation of the training process, as well as the basic PCE theory, the reader is referred to Appendix A (and further to Xiu and Karniadakis, 2002; Sudret, 2008, for yet more detail).

4.4 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 model is described (Sacks et al., 1989) by

$$Y(\mathbf{X}) = \mathbf{f}(\mathbf{X})^T \boldsymbol{\beta} + Z(\mathbf{X}), \tag{13}$$

where for N evaluation samples and an M-dimensional problem, $\mathbf{X}$ represents an $M \times N$ matrix of input variables and $Y(\mathbf{X})$ is the output vector. The term $\mathbf{f}(\mathbf{X})^T \boldsymbol{\beta}$ 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 $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_P]$, whereas $Z(\mathbf{X})$ is a *zero-mean* stationary Gaussian process. The (joint) probability distribution of the Gaussian process is characterized by its covariance; for two distinct

'points' $\mathbf{X}$ and $\mathbf{W}$ in the sample domain the covariance is defined as

$$V(\mathbf{W},\mathbf{X}) = \sigma^2 R(\mathbf{W},\mathbf{X},\boldsymbol{\theta}), \tag{14}$$

where σ^2 is the overall 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. correlation length scale(s). Since the mean

5 and variance of the Gaussian field can be expressed as functions of θ (this is shown in details in Appendix A), the calibration of the Kriging model amounts to determining the trend coefficients and obtaining an optimal solution for θ.

The functional form of the mean field $\mathbf{f}(\mathbf{X})^T\boldsymbol{\beta}$ is identical to the generalized PCE defined in eq. (A8), 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 4.3. A suitable approach for tuning the Gaussian field statistics is to find the values

10 of $\boldsymbol{\beta}$, σ^2 and $\boldsymbol{\theta}$ which maximize the likelihood of the training set variables $\mathbf{Y}$, i.e. minimize the model error in a least-squares sense (Lataniotis et al., 2015). This is described in Appendix A.

The main practical difference between regression- or expansion-type models such as regular PCE and the Kriging approach is *how* the training sample is used in the model: in pure regression-based approaches the training sample is used to only calibrate the regression coefficients, while in Kriging (and in other interpolation techniques) the training sample is retained and used in

15 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. 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

20 variations in turbulence) the inferred correlation length will tend to zero and the Kriging estimator will be represented by the trend function alone.

4.5 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, has been used for wind turbine load prediction (Toft et al., 2016). The procedure involves

25 fitting a quadratic polynomial regression ('response surface') to a normalized space of i.i.d. variables, which are derived from the physical variables using an isoprobabilistic transformation—such as the Rosenblatt transformation given in eq. (1) and (2). The design points used for calibrating the response surface in k dimensions form a combination of a central point, axial points a distance of $\sqrt{k}$ in each dimension, and a 2^k 'factorial design' set where there are two levels (points) per variable dimension located at unit distance from the origin; this is shown in Figure 3 for the case of $k = 2$ variables (dimensions). Due to the

30 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. The low order of the response surface also prohibits full characterization of the highly nonlinear 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^{-1} in
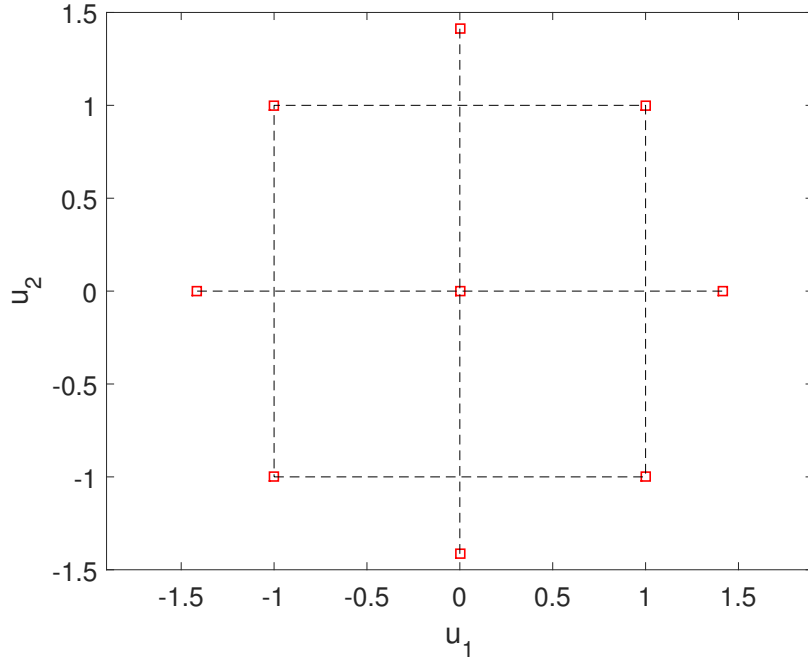
Figure 3. Example of a rotatable Central Composite Design (CCD) in a 2-dimensional standard normal space $[u_1, u_2]$. The CCD consists of a central point, a 2^k 'factorial design' with 2 levels and $k = 2$ dimensions, and axial points at distance $u = \sqrt{2}$, meaning that all the outer points lie on a circle.

$1\,\mathrm{m\,s^{-1}}$ steps. This approach may in principle be extended to include additional variables such as turbulence (σ_u), however doing so will reduce the practicality of the procedure as it will require multi-dimensional interpolation between large number of models and the uncertainty may increase. 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
5   to replace three of the variables with relatively low importance (yaw, tilt, and air density) with their mean values. The result is a 6-dimensional problem consisting of 22 different 5-dimensional response surfaces, which require $22 \cdot (1 + 2 \cdot 5 + 2^5) = 946$ design points in total. Analogous 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. (A8). For any sample point p in the central composite design, the corresponding row in the design
10   matrix is defined as

$$\mathbf{\Psi}_p = \left[\{1\}, \quad \{U_1, \ldots U_n\}, \quad \{U_1^2, \ldots U_n^2\}, \quad \{U_i \cdot U_j, \, i = 1 \ldots n, \, j = 1 \ldots (i-1)\} \right] \tag{15}$$

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

4.6 Sensitivity indices

We use the global Sobol indices, (Sobol, 2001), for evaluating the sensitivity of the response with respect to input variables. Having trained a surrogate model, the total Sobol indices can be computed efficiently by carrying out a Monte Carlo simulation on the surrogate. For number of dimensions equal to M (e.g. $M = 9$ in the present study) and for N (quasi) 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, $i = 1 \ldots M$, a third matrix $\mathbf{AB}_i$ is created by taking the ith column of $\mathbf{AB}_i$ equal to the ith column from $\mathbf{B}$, and all other columns taken from $\mathbf{A}$. The load surrogate 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 $i = 1 \ldots M$, simulation-based Sobol' sensitivity indices can be estimated as

$$SU_i = \frac{1}{N} \sum_{j=1}^{N} f(\mathbf{B})_j \left(f(\mathbf{AB}_i)_j - f(\mathbf{B})_j \right) \tag{16}$$

where $j = 1 \ldots N$ is the row index in the design matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{AB}_i$ (Saltelli et al., 2008). For the problem discussed in the present study, it was sufficient to use approximately 500 MC samples per variable dimension in order to compute the total Sobol indices.

4.7 Model reduction

For any polynomial-based regression model which includes dependence between variables, the problem grows steeply in size when the number of dimensions, M, and the maximum polynomial order, 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. For a PCE model, the objective function using a LASSO regularization is

$$\mathbf{S} = \min \left\{ \left\| \frac{1}{2N_e} \sum_{i=1}^{N_e} \left[g(\boldsymbol{\xi}^{(i)}) - \sum_{j=0}^{N_p-1} S_j \Psi_{\boldsymbol{\gamma},j}(\boldsymbol{\xi}^{(i)}) \right]^2 + \lambda \sum_{j=0}^{N_p-1} |S_j| \right\| \right\} \tag{17}$$

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. In the present study, the LASSO regularization is used on the PCE-based models to reduce the number of coefficients.

One useful corollary of the orthogonality in the PCE polynomial basis is that the contribution of each individual term to the total variance of the expansion (i.e. the individual Sobol indices) can be easily computed based on the coefficient values (see Appendix A). 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. (17), a model reduction of an order of magnitude or more can be achieved. For example, for the 9-dimensional PCE of order 6 discussed in Section 5.3, 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.

5 Model training and performance

5.1 Model convergence

5 We assess the convergence of PCE by calculating the normalized root-mean-square error (NRMSE) 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 $\mathbf{X}^{(i)}$ from the training sample defined in Section 2:

$$\varepsilon_{\mathrm{NRMS}} = \frac{1}{\mathrm{E}[\mathbf{y}]} \sqrt{\frac{\sum_{i=1}^{N}(\tilde{y}_i - y_i)^2}{N}} \tag{18}$$

where $\mathrm{E}[\mathbf{y}]$ is the expected value of the observed variable. Figure 4 shows the NRMSE for a non-truncated PCE of order 6
10 and with 6 dimensions, as function of the number of samples used to train the PCE, and the hours of load simulations (i.e. number of seeds) used for each sample point. The NRMSE shown on Figure 4 is calculated based on a set of 500 quasi-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. Each of the quasi-MC samples is the mean from 48 turbulent 10-minute simulations. To mimic the seed-
15 to-seed uncertainty, each of the PCE predictions is also evaluated as the mean of 48 normally distributed random realizations, with mean and standard deviation prescribed by the PCE model for mean and standard deviation of the blade flapwise DEL respectively.

Figure 4 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
20 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 NRMSE 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
25 not explained by the smooth PCE surface. Further increase in the number of training points or simulation length will only reduce this statistical uncertainty, but will not contribute significantly to changes in the model predictions as the flexibility of the model is limited by the maximum polynomial order. Therefore, the model performance achieved under these conditions can be considered near to the best possible for the given PCE order and number of dimensions. 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,
30 and higher order and more dimensions will require more than the approximately 3000 points which seem sufficient for a PCE of order 6 and dimension 6, as shown on Figure 4.
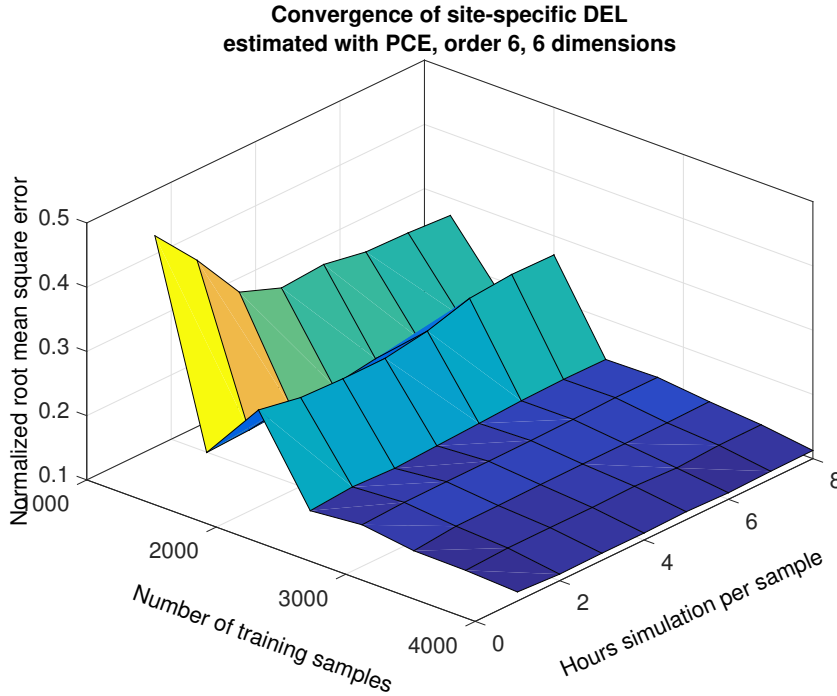
Figure 4. Convergence of a PCE of dimension 6 and order 6, as function of number of collocation points and hours of simulation per collocation point. The z-axis represents the NRMSE obtained from the difference between 500 site-specific quasi-MC samples of blade root flapwise DEL for reference site 0, and the corresponding predictions from PCE.

The IS procedure has relatively slow convergence compared to e.g. a quasi-MC simulation. Figure 5 shows an example of the convergence of an IS integration for reference site 0, based on computing the target (site-specific) distribution weights for all 10^4 points in a reference high-fidelity database. The confidence intervals are obtained by bootstrapping.

5.2 One-to-one comparison and mean squared error

5 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 error (NRMSE), defined in eq. (18) resulting from a point-by-point comparison of load predictions from a reduced-order model against actual reference

10 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 NRMSE for site 0 for Kriging, RS, and PCE-based load predictions are listed in Table 2. In addition, Figure 6
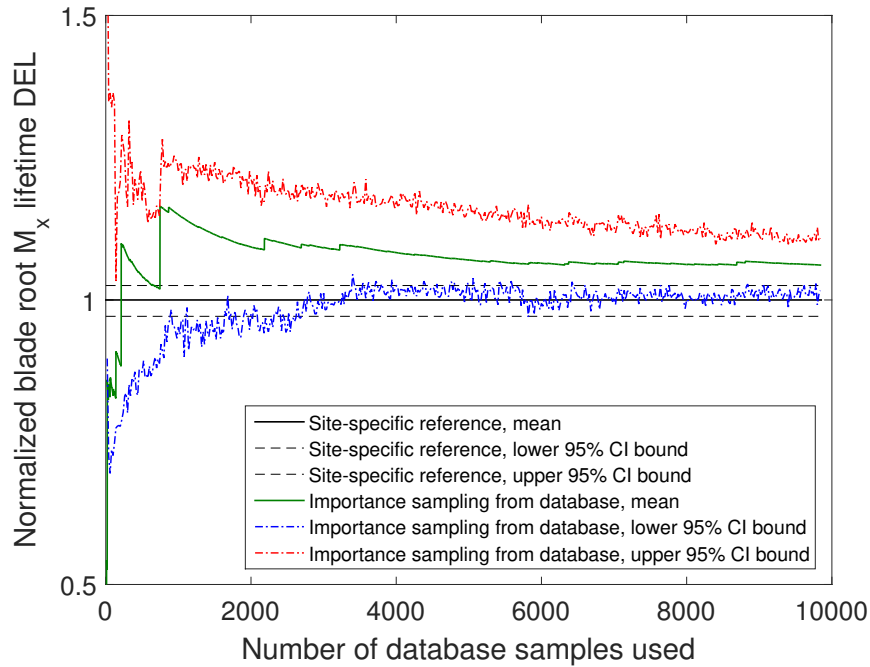
Figure 5. Convergence of an importance sampling (IS) calculation of the blade root moment from the hi-fi database, towards site-specific lifetime fatigue loads for reference site 0 (table 5).

presents a one-to-one comparison where for a set of 200 sample points the load estimates from the site-specific MC simulations are compared against the corresponding surrogate model predictions in terms of $y{-}y$ plots.

Table 2. Normalised 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
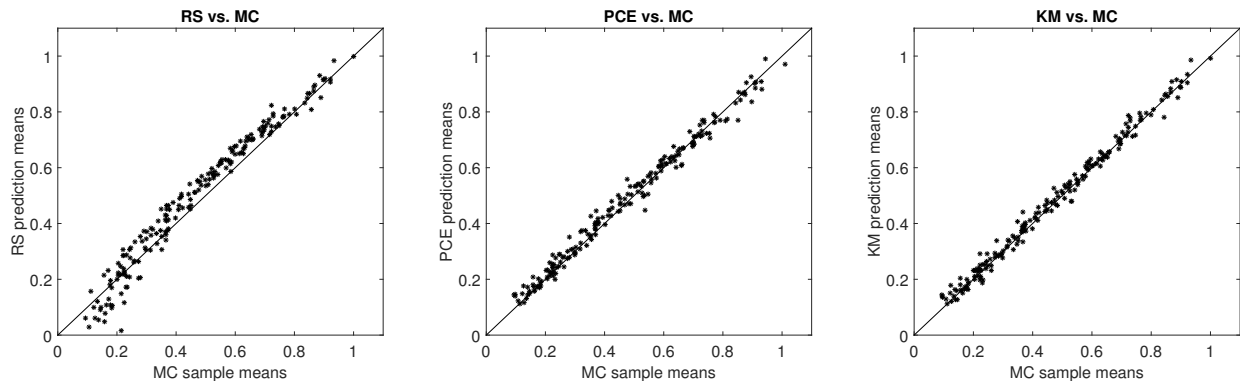
Figure 6. $y-y$ plots comparing the blade root flapwise 1Hz damage-equivalent load predictions for three load surrogate models - quadratic Response Surface, Polynomial Chaos expansion, and Kriging model, compared against site-specific Monte Carlo simulation. The x-axis represents the loads obtained using site-specific Monte Carlo simulations for reference site 0, and the y-axis represents the mean 1Hz-equivalent load estimated for the same sample points using a surrogate model. All values are normalized with the maximum equivalent load attained from the site-specific Monte Carlo simulation.

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.

5.3 Variable sensitivities

As described earlier in Section 4.6, 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 joint distribution used to generate the database with high-fidelity load simulations. The total Sobol indices for the high-fidelity load database variable range are computed directly from the PCE fitted to the database by evaluating the variance contributions from the expansion coefficients (see Appendix A) and are listed in Table 3. The indices for the site-specific distribution corresponding to reference site 0 are computed using the Monte-Carlo based method described in Section 4.6 as direct PCE indices are not available for this sample distribution. The resulting total Sobol indices for the 6 variables available at site 0 are listed in Table 4. 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.

Table 3. PCE-based Sobol sensitivity indices for the high-fidelity load database variable ranges.

FATIGUE LOAD SENSITIVITY INDICES									
	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

Table 4. Site-specific Sobol sensitivity indices derived for Site 0 using Monte Carlo simulation from a PCE.

FATIGUE LOAD SENSITIVITY INDICES					
	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

6 Site-specific calculations

6.1 Reference sites

5 The low-fidelity 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 virtual sites, based on real-world measurement data which cover most 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 virtual sites chosen are characterized with conditions that slightly exceed the standard conditions specified by a certain type-certification class. Exceptions are site 0, which has the most measured 10 variables available and is therefore chosen as a primary reference site, and the virtual "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 5.

Table 5. Reference virtual sites used for validation of the site-specific load estimation methods.

Site No.	Location	Terrain	Specific condition	Variables included	MC sample size
0	Denmark	Flat agricultural	-	$U, \sigma_u, \alpha, L, \Gamma, \Delta\varphi$	492
1	Denmark	Flat agricultural	IIIC exceedance	U, σ_u, α	823
2	Denmark	Forested	IIIB exceedance	U, σ_u, α	884
3	Denmark	Forested	IA exceedance	U, σ_u, α	949
4	Denmark	Forested	IIA exceedance	U, σ_u, α	871
5	Colorado, USA	Mountain foothills	Low-wind	U, σ_u, α	657
6	Colorado, USA	Mountain foothills	Low-wind	U, σ_u, α	853
IEC IA, NTM	-	-	Standard reference class	U	22
IEC IIB, NTM	-	-	Standard reference class	U	22
IEC IIIC, NTM	-	-	Standard reference class	U	22
IEC IIB, ETM	-	-	Standard reference class	U	22

Site 0 (also also referred to herein as the reference site) is located at the Nørrekær Enge wind farm in Northern Denmark (Bor-raccino et al., 2017), over flat, open agricultural terrain. Site 1 is a flat-terrain, near-coastal site at the the National Centre for Wind Turbines at Høvsøre, Denmark (Peña et al., 2016). 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). 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 virtual 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 (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.

For each site, the joint distributions of all variables are defined in terms of conditional dependencies, and generating simu-lations of site-specific conditions is carried out using the Rosenblatt transformation, eq. (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 rec-ommended by the IEC 61400-1 (ed. 3, 2005) standard. The wind shear exponent is dependent on the mean wind speed and on

the turbulence, and the distribution parameters of α are defined by the relationship (Kelly et al., 2014; Dimitrov et al., 2015)

$$\mu_{\alpha|I_c,u} = \alpha_{\text{ref}} + \frac{I_{c,\text{ref}} - I_c(U)}{I_c(U) \cdot c_\alpha}$$
$$\sigma_\alpha = 1/U$$

(19)

where μ_α and σ_α are the mean and standard deviation of α, respectively; $I_c(U) = (\sigma_u/U | F(\sigma_u) = Q)$ is a characteristic turbulence intensity based on a turbulence quantile Q, as a function of the wind speed U. Here $I_{c,\text{ref}} = I_c(U = 15\,\text{m s}^{-1})$ is the reference characteristic turbulence intensity at $U = 15\,\text{m s}^{-1}$ and $\alpha_{\text{ref}} = \alpha | (I_c = I_{c,\text{ref}}, U = 15\,\text{m s}^{-1})$ is a reference wind shear exponent, with α_c being an empirically determined constant. Since the turbulence quantities $I_c(u)$ and $I_{c,\text{ref}}$ are defined by the conditional relationship between wind speed and turbulence, the only site-specific parameters required for characterizing the wind shear are α_{ref} and c_α. For each of the physical sites, the wind speed, turbulence and wind shear distribution parameters are defined based on anemometer measurements at the sites. The results are listed in Table 6. In addition, high-frequency 3D sonic measurements were available at site 0 for the entire measurement period, which allowed for estimating Mann turbulence model parameters using the approach described in (Dimitrov et al., 2017).

Table 6. Parameters defining the conditional distribution relationships used in computing joint distributions of the environmental conditions for the test sites/conditions.

Site	Weibull A	Weibull K	μ_{σ_u} relationship	σ_{σ_u} relationship	α_{ref}	c_α
0	9.44	3.36	$0.106U - 0.0973$	$0.0041U + 0.194$	0.142	4
1	8.24	1.78	$0.109U + 0.0624$	$0.021U + 0.154$	0.188	4
2	8.51	2.35	$0.148U - 0.248$	$0.0061U + 0.330$	0.294	4
3	10.25	2.47	$0.149U - 0.185$	$0.0021U + 0.329$	0.230	4
4	9.03	2.43	$0.175U - 0.497$	$0.009U + 0.298$	0.407	4
5	6.12	1.46	$0.0637U + 0.915$	$-0.0113U + 0.807$	0.108	4
6	9.03	1.80	$0.0972U + 0.437$	$0.0014U + 0.377$	0.121	4
7 (IEC IA, NTM)	11.28	2	$0.12U + 0.608$	$0U + 0.224$	0.2	0
8 (IEC IIB, NTM)	9.59	2	$0.105U + 0.532$	$0U + 0.196$	0.2	0
9 (IEC IIIC, NTM)	8.46	2	$0.09U + 0.456$	$0U + 0.168$	0.2	0

With this procedure, 1000 quasi-Monte Carlo samples of the environmental conditions at each site are generated from the respective joint distribution. All realizations where the wind speed is between the DTU 10MW wind turbine cut-in and cut-out wind speed are fed as input to load simulations. The actual number of load simulations for each site are given in Table 6. Similarly to the load database simulations, 8 simulations with 1h duration are carried out at each site-specific MC sample point. The resulting reference lifetime equivalent loads are then defined by applying eq. (6) on the Monte Carlo simulations using equal weights $p(\mathbf{X}_i) = 1/N$, while the uncertainty in the lifetime loads is estimated using bootstrapping on the entire MC sample.

6.2 Lifetime fatigue loads

The lifetime damage-equivalent (DEL) loads are computed for all reference sites in Table 5, using the five load surrogate models defined above: 1) quadratic response surface; 2) polynomial chaos expansion, 3) importance sampling, 4) nearest-neighbor interpolation; and 5) Kriging with the mean defined by polynomial chaos basis functions. Methods 2–5 are based on data from the high-fidelity load database described in Section 2. In addition to the surrogate model computations, a full MC reference simulation is carried out for each validation site. The load predictions with the Monte Carlo approach are obtained by carrying out Hawc2 aeroelastic simulations on the same DTU 10MW reference wind turbine model used for training the load surrogate models. A total of $N_{MC} = 1000$ quasi-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 of the quasi-MC samples where the wind speed is between cut-in and cut-out. 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 quasi-MC simulation of the respective model with the same sample set of environmental inputs used in the reference MC simulations. The load predictions with Importance Sampling are based on the probability-weighted contributions from the samples in a high-fidelity load database. For each site-specific distribution, the database samples are ordered according to their weights, and only a number of points, N_{IS}, with the highest weights are used in the estimation. For the sake of easier comparison between different methods, it is chosen that $N_{IS} = N_{MC}$. Based on computations from PCE with different number of dimensions and 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 7 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 generating a number of realizations for each sampled combination of input variables, and subsequently computing confidence intervals by bootstrapping (eq. (7)). In this way, the bootstrapping is done simultaneously for 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 the same bootstrapping approach is 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 uncertainties and confidence intervals as defined in Section 3.3 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 8 and 9, where the performance of all site-specific load estimation methods is compared for reference site 0, for 8 load channels
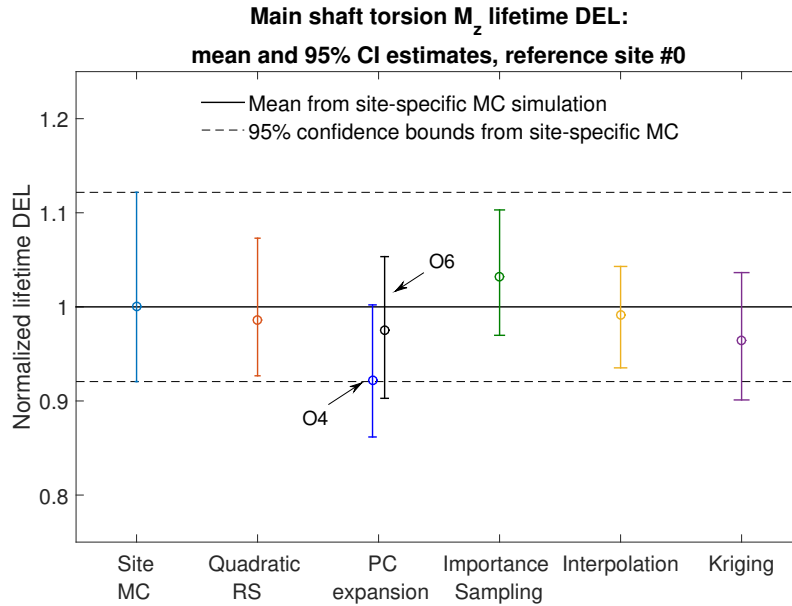
Figure 7. Comparison of predictions of the lifetime damage-equivalent loads for six different estimation approaches. All values are normalized with respect to the mean estimate from a site-specific Monte Carlo simulation, and the error bars represent the bounds of the 95% confidence intervals. Results from two PCEs are shown: the blue bar corresponds to the output of a 4th order PCE, while the black bar corresponds to a 6th order PCE.

in total, with number of samples as listed in Table 8. Figure 8 shows results for tower base and tower top fore-aft and side-side bending moments, and Figure 9 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
5 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 summarized site-specific results for all surrogate-based load estimation methods are shown in Table 7. In order to compute these values, the load estimates for each site and load channel are normalized to the results obtained with the direct site-specific Monte Carlo simulations. The values given in Table 7 are
10 averaged over all reference sites. The results for individual sites and load channels are depicted as bar plots in Figures 10 and 11, for tower load and rotor load channels respectively. 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 ~ 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
15 be attributed to significantly smaller number of samples used for the IEC-based sites (22 samples where only the wind speed
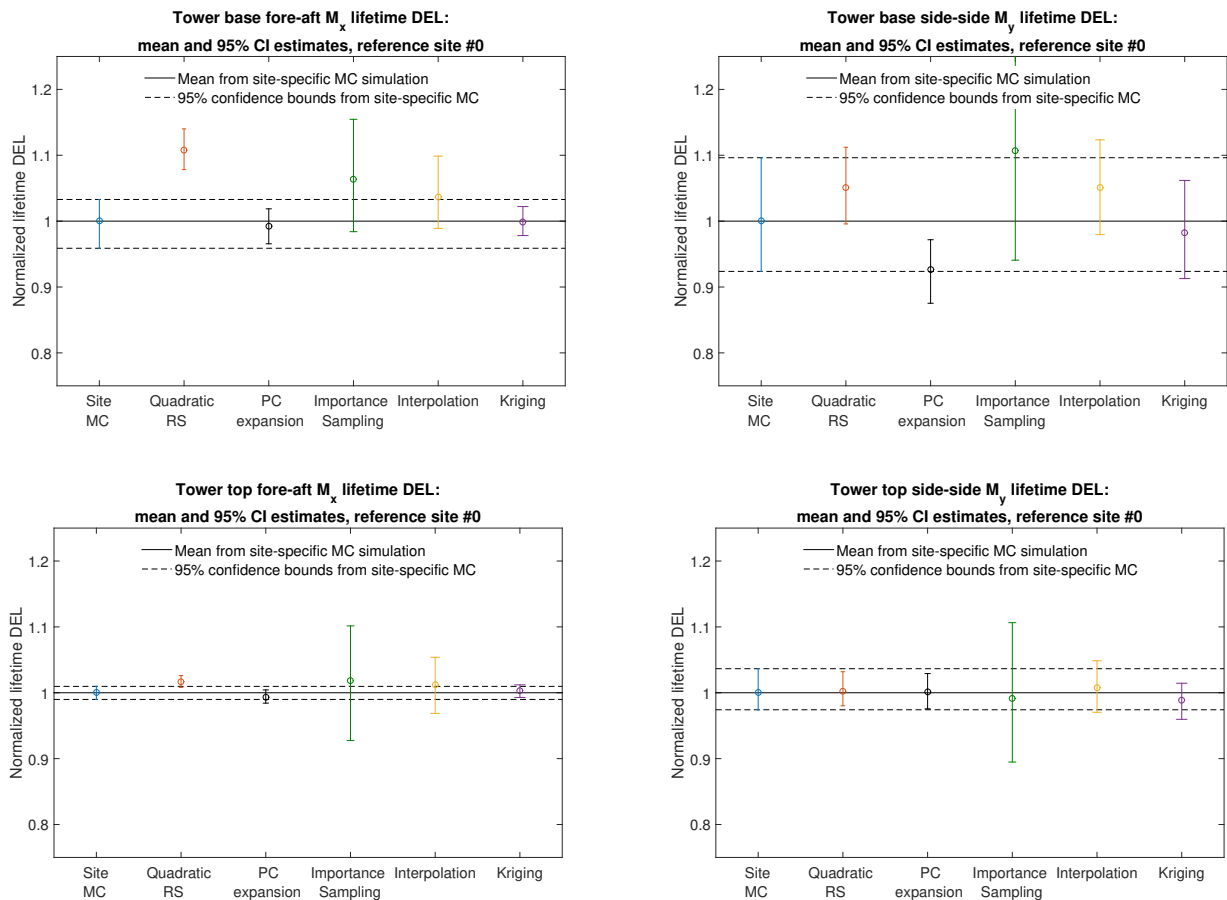
Figure 8. Comparison of predictions of the lifetime damage-equivalent loads for six different estimation approaches. All values are normalized with respect to the mean estimate from a site-specific Monte Carlo simulation.

is varied in 1m/s steps from 4 to 25m/s). 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 4.1, the uncertainty of the IS model can increase when the site-specific distribution has fewer dimensions than the model because fewer points from the high-fidelity database will have high probabilities with respect to the site-specific distribution. It can be expected that this effect

5 is strongest for the IEC class-based sites, as for them only a single variable — the wind speed — is considered stochastic.

Another important aspect to consider when comparing the performance of the surrogate models is the model execution speed, and whether there is a tradeoff between speed and accuracy. A comparison of the model evaluation times for the site-specific lifetime load computation for site 0 is given in Table 8. Noticeably the Kriging model requires significantly longer execution time than other approaches, which is mainly due to the requirement of populating a cross-correlation matrix.
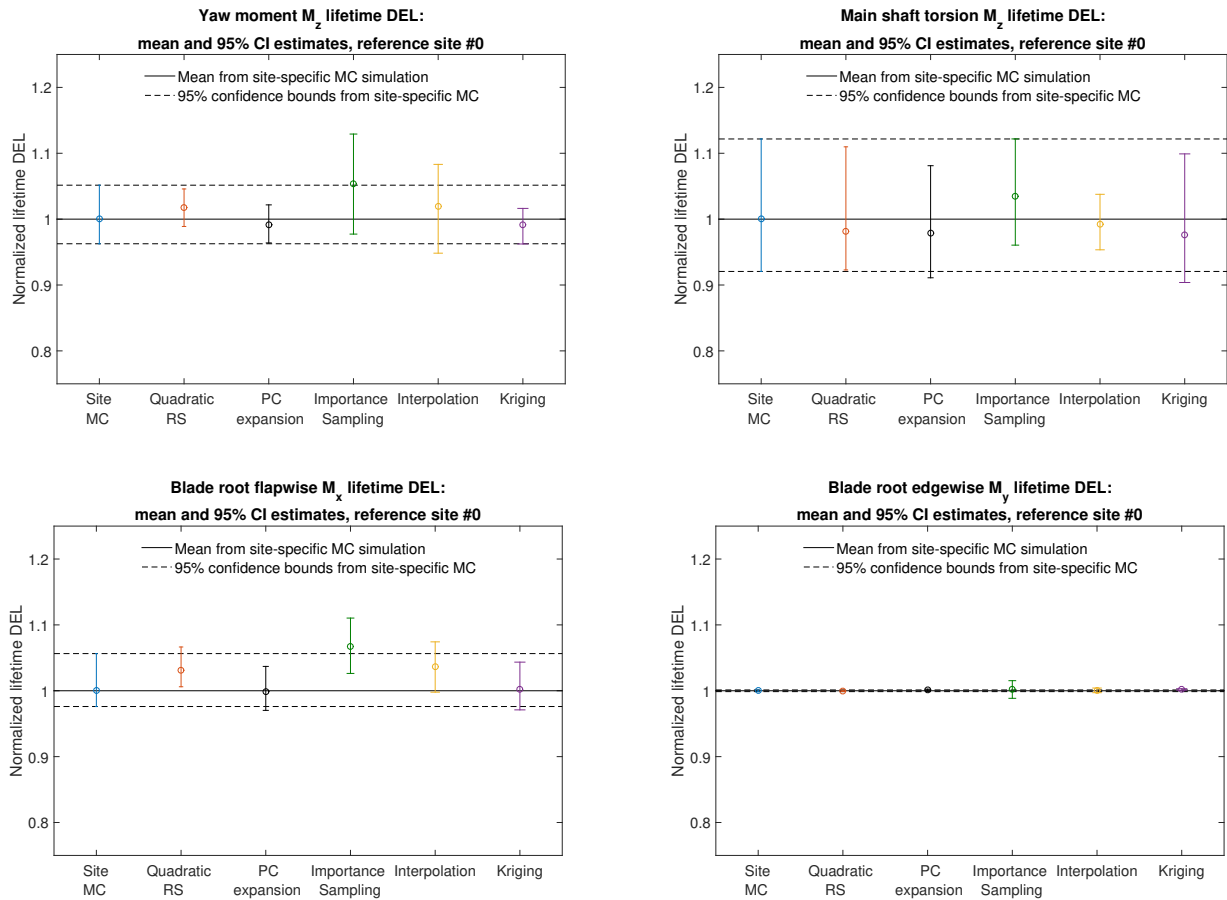
Figure 9. Comparison of predictions of the lifetime damage-equivalent loads for six different estimation approaches. All values are normalized with respect to the mean estimate from a site-specific Monte Carlo simulation.

7 Discussion and conclusions

7.1 Discussion

The previous sections of this paper described a procedure for estimating site-specific lifetime damage-equivalent loads, using several simplified model techniques applied to ten different sites and conditions. Based on the site-specific lifetime DEL

5 comparisons, for quick site-specific load estimation the three models based on machine learning showed to be most viable (sufficiently accurate over the majority of the sampling space): polynomial chaos expansion, Kriging, and the quadratic response surface. 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

10 of the quadratic RS, which means that the model accuracy decreases in the sampling space away from the calibration points,
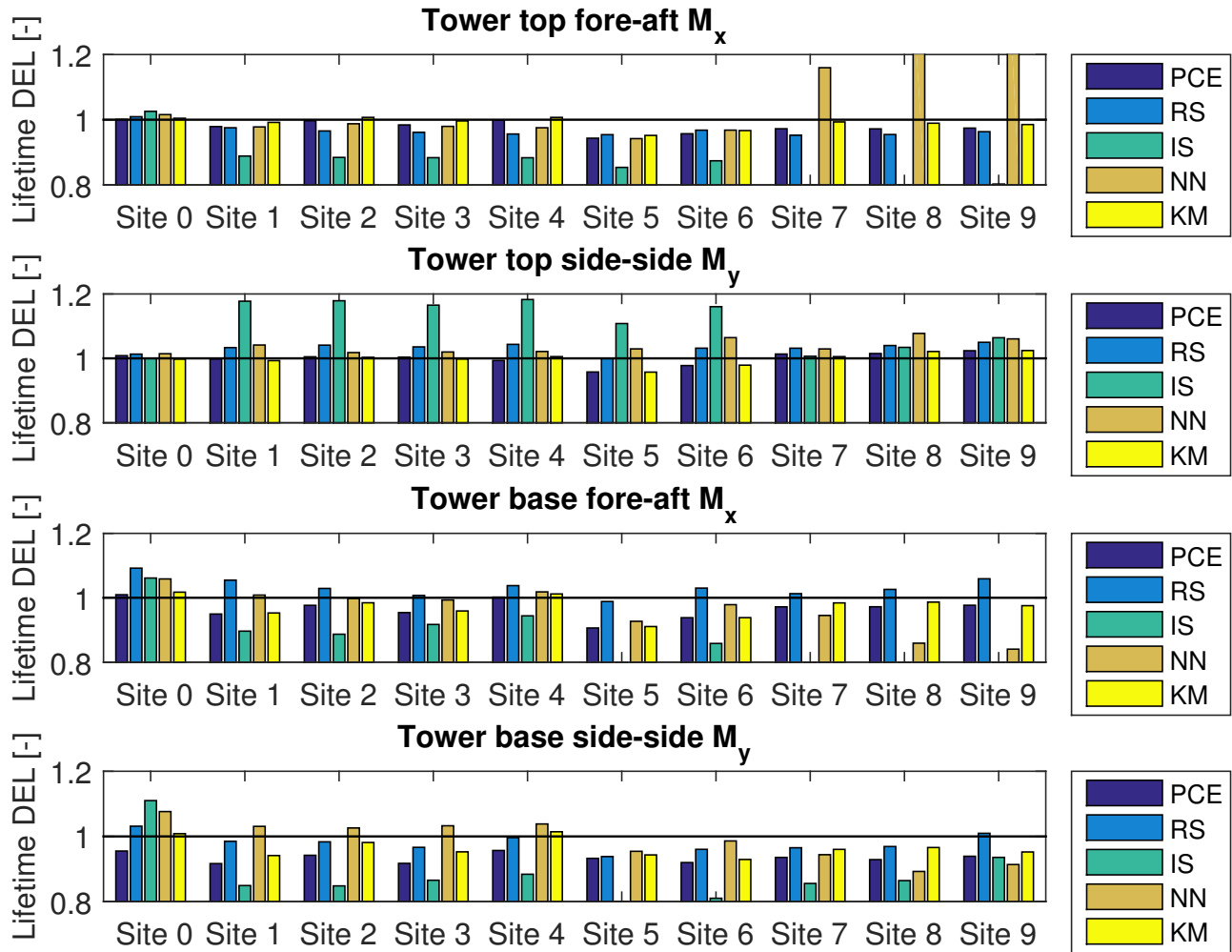
Figure 10. Predictions of lifetime damage-equivalent tower loads for five different estimation approaches and four load channels for the different sites (0–6) and IEC conditions (virtual sites 7–9). All values are normalized with respect to the mean estimate from a site-specific Monte Carlo simulation. The abbreviations refer to: PCE: Polynomial Chaos Expansion; RS: quadratic Response Surface; IS: Importance Sampling; NN: Nearest-neighbor Interpolation; KM: Universal Kriging Model.
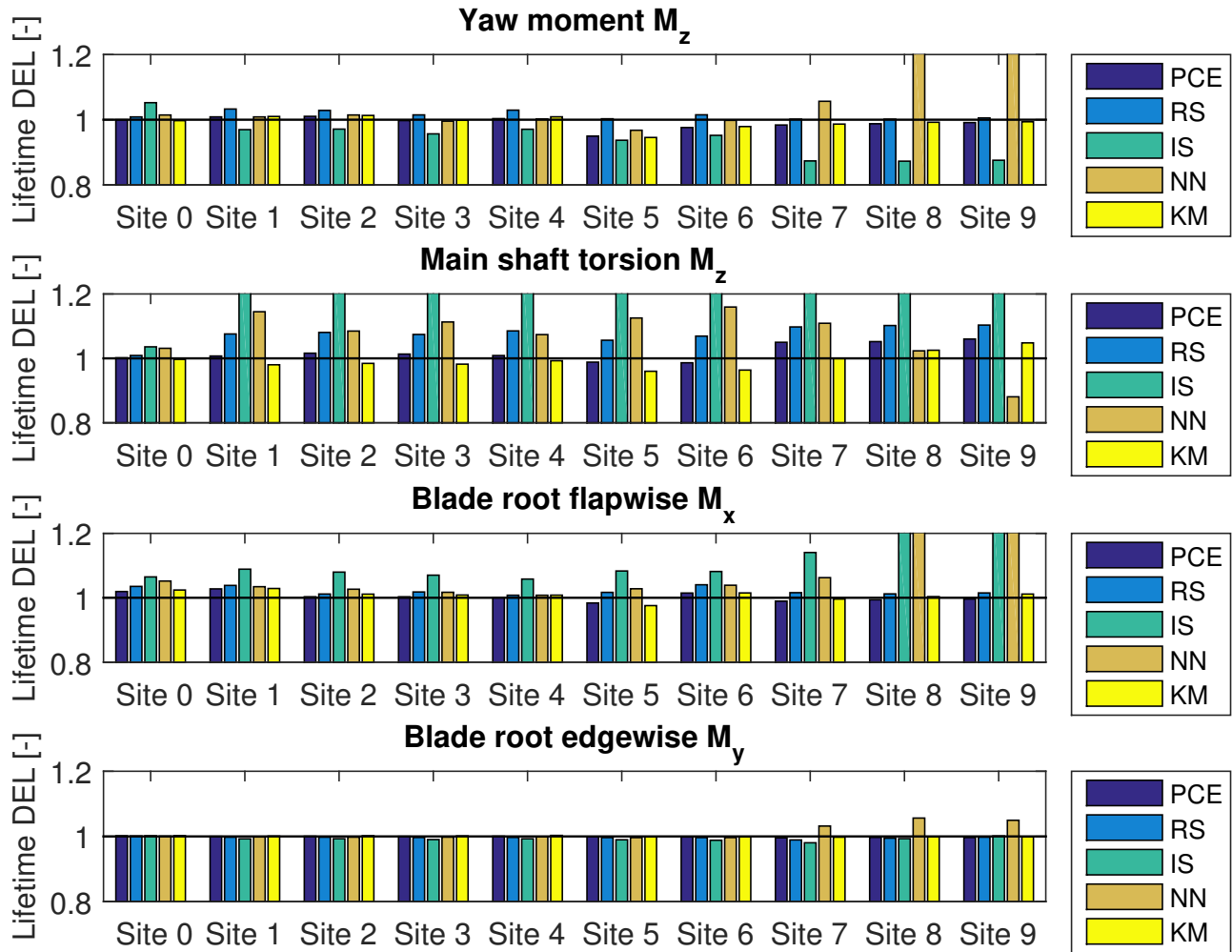
Figure 11. Predictions of lifetime damage-equivalent loads (yaw, shaft torsion, blade-root) for five different estimation approaches and four load channels. All values are normalized with respect to the mean estimate from a site-specific Monte Carlo simulation. The abbreviations refer to: PCE: Polynomial Chaos Expansion; RS: quadratic Response Surface; IS: Importance Sampling; NN: Nearest-neighbor Interpolation; KM: Universal Kriging Model.

Table 7. Lifetime-equivalent load predictions normalized with respect to Monte Carlo simulations and averaged over 10 reference sites. 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.

		Load channels							
		TB M_x	TB M_y	TT M_x	TT M_y	TT M_z	MS M_z	BR M_x	BR M_y
POLYNOMIAL CHAOS EXPANSION									
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
UNIVERSAL KRIGING									
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
QUADRATIC RESPONSE SURFACE									
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
IMPORTANCE SAMPLING									
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
NEAREST-NEIGHBOR INTERPOLATION									
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. Model execution times for the lifetime damage-equivalent fatigue load computations for site 0

Surrogate model	Training set size	Evaluation set size	Evaluation time
MC	-	492	22.7s
PCE	10000	492	8.2s
RS	946	492	2.2s
IS	10000	492	4.6s
NN	10000	492	4.4s
Kriging	2048	492	217.6s

especially if there is any extrapolation. This inaccuracy is reflected in the NRMSE-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 smallest overall uncertainty, both in sample-to-sample comparisons and in lifetime DEL computations. This is to be expected, since the Kriging employs a well-performing model (the PCE) and combines it with an interpolation scheme that subsequently 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 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 (the β term in Eq. 13) is the primary 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, for some other calculations it may be more practical to choose other bound definitions: e.g. for the extreme turbulence models prescribed by the IEC 61400-1, the currently suggested bounds do not include ETM class A.

For the more advanced methods like PCE and Kriging, there is a practical limitation of the number of training points to be used in 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–$2{\cdot}10^4$ points on a typical desktop computer (as of 2018). 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 most applications is less than for the PCE. However, as it was shown in Sections 4.3 and 6, 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, such as orthogonality facilitating creation of sparse models through variance-based sensitivity analysis, we consider the PCE as the most useful method overall.

In addition to the load-mapping approaches presented in this paper, Artificial Neural Networks (ANNs) are interesting alternative candidates. ANN (see Goodfellow et al., 2016) are machine learning models that have gained popularity due to their flexibility and history of successful application to many different problems. It is very likely that a sufficiently large Neural Network model can provide similar output quality and performance as the methods described in the present study. This is therefore a matter worth of future research. However, the PCE-based models may sometimes have a practical advantage over ANNs, due to the "white-box" features—such as being able to track separate contributions to variance (and uncertainty), as well as 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 about 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 (e.g. uncertainty in the site conditions or the turbine operating strategy) can lead to even larger errors. The procedures demonstrated in this study are thus very suitable for carrying out quick site feasibility assessments; the latter can help to decide in a timely fashion whether to discard a given site as unfeasible, or to 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.

7.2 Summary and conclusions

In the present work we defined and demonstrated a procedure for quick assessment of site-specific lifetime fatigue loads using load 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 by training the surrogate models on a database with aeroelastic load simulations of the DTU 10MW reference wind turbine. 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 (standard deviation of wind speed fluctuations) are the factors having the highest influence on fatigue loads. The wind shear and the Mann turbulence length scale were also found to have an appreciable influence, with the effect of wind shear being more pronounced for rotating components such as blades. Within the 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 employing polynomial chaos expansion 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 conditional on a specific wind turbine model.

Competing interests. No competing interests are present

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.

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.

Choe, Y., Byon, E., and Chen, N.: Importance Sampling for Reliability Evaluation With Stochastic Simulation Models, Technometrics, 57, 351–361, https://doi.org/10.1080/00401706.2014.1001523, 2015.

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.

Ganesh, N. and Gupta, S.: Estimating the Rain-Flow Fatigue Damage in Wind Turbine Blades Using Polynomial Chaos, in: Proceedings of the International Symposium on Engineering under Uncertainty: Safety Assessment and Management (ISEUSAM - 2012), https://doi.org/10.1007/978-81-322-0757-3, 2013.

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.

Graf, P., Dykes, K., Damiani, R., Jonkman, J., and Veers, P.: Adaptive stratified importance sampling: hybridization of extrapolation and importance sampling Monte Carlo methods for estimation of wind turbine extreme loads, Wind Energy Science, 3, 475–487, https://doi.org/10.5194/wes-3-475-2018, 2018.

Graf, P. A., Stewart, G., Lackner, M., Dykes, K., and Veers, P.: High-throughput computation and the applicability of Monte Carlo integration in fatigue load estimation of floating offshore wind turbines, Wind Energy, 19, 861–872, https://doi.org/10.1002/we.1870, http://onlinelibrary.wiley.com/doi/10.1002/we.1608/fullhttp://doi.wiley.com/10.1002/we.1870, 2016.

Häfele, J., Hübler, C., Gebhardt, C. G., and Rolfes, R.: A comprehensive fatigue load set reduction study for offshore wind turbines with jacket substructures, Renewable Energy, 118, 99–112, https://doi.org/10.1016/j.renene.2017.10.097, 2018.

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.

Hübler, C., Gebhardt, C. G., and Rolfes, R.: Hierarchical four-step global sensitivity analysis of offshore wind turbines based on aeroelastic time domain simulations, Renewable Energy, 111, 878–891, https://doi.org/10.1016/j.renene.2017.05.013, http://dx.doi.org/10.1016/j.renene.2017.05.013, 2017.

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

5 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, 533–543, https://doi.org/10.5194/wes-3-533-2018, https://www.wind-energ-sci.net/3/533/2018/, 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.

10 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.

15 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.

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.

20 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.

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,
25 Wind Engineering and Industrial Aerodynamics, 82, 69–88, 1999.

Müller, K. and Cheng, P. W.: Application of a Monte Carlo procedure for probabilistic fatigue design of floating offshore wind turbines, Wind Energy Science, 3, 149–162, https://doi.org/10.5194/wes-3-149-2018, 2018.

Müller, K., Dazer, M., and Cheng, P. W.: Damage Assessment of Floating Offshore Wind Turbines Using Response Surface Modeling, Energy Procedia, 137, 119–133, https://doi.org/10.1016/j.egypro.2017.10.339, https://doi.org/10.1016/j.egypro.2017.10.339, 2017.

30 Murcia, J. P., Réthoré, P.-E., Dimitrov, N., Natarajan, A., Sørensen, J. D., Graf, P., and Kim, T.: Uncertainty propagation through an aeroelastic wind turbine model using polynomial surrogates, Renewable Energy, 119, 910–922, https://doi.org/https://doi.org/10.1016/j.renene.2017.07.070, 2018.

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,
35 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.

Sobol, I.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Mathematics and Computers in Simulation, 55, 271–280, 2001.

Stewart, G. M.: Design Load Analysis of Two Floating Offshore Wind Turbine Concepts, Ph.D. thesis, University of Massachusetts - Amherst, http://scholarworks.umass.edu/dissertations{_}2, 2014.

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

Teixeira, R., O'Connor, A., Nogal, M., Krishnan, N., and Nichols, J.: Analysis of the design of experiments of offshore wind turbine fatigue reliability design with Kriging surfaces, Procedia Structural Integrity, 5, 951–958, https://doi.org/10.1016/j.prostr.2017.07.132, http://linkinghub.elsevier.com/retrieve/pii/S2452321617302445, 2017.

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.

Yang, H., Zhu, Y., Lu, Q., and Zhang, J.: Dynamic reliability based design optimization of the tripod sub-structure of offshore wind turbines, Renewable Energy, 78, 16–25, https://doi.org/10.1016/j.renene.2014.12.061, 2015.

Zwick, D. and Muskulus, M.: Simplified fatigue load assessment in offshore wind turbine structural analysis, Wind Energy, 19, 265–278, https://doi.org/10.1002/we.1831, http://onlinelibrary.wiley.com/doi/10.1002/we.1608/fullhttp://doi.wiley.com/10.1002/we.1831, 2016.

Appendix A: Reduced-order models: background and theory

A1 Polynomial chaos expansion

Polynomial Chaos Expansion (PCE) is a popular method for approximating stochastic functions 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 $(-\infty, \infty)$.

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

$$\xi_i = 2F(X_i) - 1, \tag{A1}$$

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

5 distributed ('i.i.d.'). Note that eq. (10) and the evaluation of the cumulative distribution in general does not account for dependence between variables - this has to be addressed by applying an appropriate transformation. In the present case where the joint probability distribution of input variables is defined in terms of conditional dependencies, 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

10 other procedures, e.g. the quadratic 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

$$15 \quad P_{\alpha_i}(\xi), \qquad \text{where} \quad i = 1, \ldots, M, \quad \alpha_i \in \mathbb{N}, \quad \sum_{i=1}^{M} \alpha_i \leq p. \tag{A2}$$

The multivariate polynomial of dimension M is then defined as

$$\Psi_{\boldsymbol{\alpha}} = \prod_{i=1}^{M} P_{\alpha_i}(\xi_i) \tag{A3}$$

With the above, each multivariate polynomial is built as the product of M univariate polynomial terms, and $\boldsymbol{\alpha}$ vector stores the orders for each univariate polynomial term. The total number of polynomials of this type is (Sudret, 2008):

$$20 \quad N_p = \begin{pmatrix} (M+p) \\ p \end{pmatrix} \frac{(M+p)!}{M!p!} \tag{A4}$$

The aim of using PCE is to represent a scalar quantity $S = g(\boldsymbol{\xi}(\mathbf{X}))$ in terms of a truncated sequence $\tilde{S}(\boldsymbol{\xi}(\mathbf{X})) + \varepsilon$ 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

$$\tilde{S}(\boldsymbol{\xi}) = \sum_{j=0}^{N_p-1} S_j \Psi_{\boldsymbol{\alpha},j}(\boldsymbol{\xi}) \tag{A5}$$

25 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 eq. (10). The most straightforward way of determining $\mathbf{S}$ is minimizing the variance of the residual ε using a least-squares regression approach:

$$\mathbf{S} = \min \left\{ \frac{1}{N_e} \left| \sum_{i=1}^{N_e} \left[g(\boldsymbol{\xi}^{(i)}) - \sum_{j=0}^{N_p-1} S_j \Psi_{\boldsymbol{\alpha},j}(\boldsymbol{\xi}^{(i)}) \right]^2 \right| \right\} \tag{A6}$$

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 $\mathbf{\Psi}$ needs to be constructed:

$$\mathbf{\Psi}_{ij} = \Psi_{\boldsymbol{\alpha},j}(\xi^{(i)}); \qquad i = 1,\ldots,N_e, \qquad j = 1,\ldots,N_p. \tag{A7}$$

Plugging the definition of $\mathbf{\Psi}$ in eq. (A5), the PCE can be expressed as $\mathbf{y} = \mathbf{\Psi S}$. Under the condition that the residuals ϵ are
5 (approximately) normally-distributed, the solution for $\mathbf{S}$ which minimizes the sum of residuals is given by

$$\mathbf{S} = (\mathbf{\Psi}^{\mathrm{T}}\mathbf{\Psi})^{-1} \cdot \mathbf{\Psi}^{\mathrm{T}} \cdot \mathbf{y}, \tag{A8}$$

with $\mathbf{y} = g(\boldsymbol{\xi}^{(i)})$ 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. (A8) requires that the so-called information matrix $(\mathbf{\Psi}^{\mathrm{T}}\mathbf{\Psi})$ is well-conditioned, which normally requires
10 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 model by penalizing the sum of the absolute value of model coefficients:

$$\mathbf{S} = \min\left\{\left|\left|\frac{1}{2N_e}\sum_{i=1}^{N_e}\left[g(\boldsymbol{\xi}^{(i)}) - \sum_{j=0}^{N_p-1} S_j \Psi_{\boldsymbol{\alpha},j}(\boldsymbol{\xi}^{(i)})\right]^2 + \lambda\sum_{j=0}^{N_p-1}|S_j|\right|\right|\right\} \tag{A9}$$

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

A2 Kriging

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

20 $$Y(\mathbf{X}) = \mathbf{f}(\mathbf{X})^T\boldsymbol{\beta} + Z(\mathbf{X}), \tag{A10}$$

where $\mathbf{X}$ represents the input variables, and $Y(\mathbf{X})$ is the output. The term $\mathbf{f}(\mathbf{X})^T\boldsymbol{\beta}$ 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 $\boldsymbol{\beta} = [\beta_1,\ldots,\beta_P]$; $Z(\mathbf{X})$ is a stationary, zero-mean Gaussian process. The probability distribution of the Gaussian process is characterized by its covariance, which for two distinct points in the domain, $\mathbf{x}$ and $\mathbf{w}$ is

25 $$V(\mathbf{w},\mathbf{x}) = \sigma^2 R(\mathbf{w},\mathbf{x},\boldsymbol{\theta}) \tag{A11}$$

where σ^2 is the overall 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

$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_N]$ where the true function values $\mathbf{y} = Y(\mathbf{X})$ are known, the aim is to obtain a model prediction at a new point, $\mathbf{x}'$. Based on Gaussian theory, the N known values $Y(\mathbf{X})$ and the Kriging predictor $\hat{Y}(\mathbf{x}')$ will be jointly Gaussian distributed:

$$\left\{ \begin{array}{c} Y(\mathbf{x}') \\ Y(\mathbf{X}) \end{array} \right\} \sim \mathcal{N}_{N+1} \left(\left[\begin{array}{c} \mathbf{f}(\mathbf{x}')^T \boldsymbol{\beta} \\ \boldsymbol{\Psi}\boldsymbol{\beta} \end{array} \right], \sigma^2 \left[\begin{array}{cc} 1 & \mathbf{r}^T(\mathbf{x}') \\ \mathbf{r}(\mathbf{x}') & \mathbf{R} \end{array} \right] \right) \tag{A12}$$

5 Here

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

$\Psi_{ij} = f_j(\mathbf{x}_i)$ **for** $i = 1 \ldots N$ **and** $j = 1 \ldots P$

where N is the number of samples and P is the total number of terms output from the basis functions - which may be different than the number of dimensions M as a basis function (e.g. a higher-order polynomial) can return more than one

10 term per variable;

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

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

$R_{ij} = R(\mathbf{x}_i, \mathbf{x}_j, \theta)$ for $i, j = 1, \ldots, N$.

It follows that the model prediction $\hat{Y}(\mathbf{x}')$ will have the following mean and variance (Santher et al., 2003):

$$\mu_{\hat{Y}}(\mathbf{x}') = \mathbf{f}(\mathbf{x}')^T \boldsymbol{\beta} + \mathbf{r}(\mathbf{x}')^T \mathbf{R}^{-1}(\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\beta})$$
$$\sigma_{\hat{Y}}^2(\mathbf{x}') = \sigma^2(1 - \mathbf{r}(\mathbf{x}')^T \mathbf{R}^{-1}\mathbf{r}(\mathbf{x}') + u(\mathbf{x}')^T [\boldsymbol{\Psi}^T \mathbf{R}^{-1} \boldsymbol{\Psi}]^{-1} u(\mathbf{x}')]. \tag{A13}$$

where $u(\mathbf{x}') = \boldsymbol{\Psi}^T \mathbf{R}^{-1}\mathbf{r}(\mathbf{x}') - \mathbf{f}(\mathbf{x}')$. 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 $\mathbf{y}$, (Lataniotis et al., 2015):

$$\mathcal{L}(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}) = \frac{\det(\mathbf{R})^{-1/2}}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\beta})^T \mathbf{R}^{-1}(\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\beta}) \right]. \tag{A14}$$

20 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, by solving the equation $\mathrm{d}(-\log \mathcal{L})/\mathrm{d}\boldsymbol{\beta} = 0$ for $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} = \beta(\boldsymbol{\theta}) = (\boldsymbol{\Psi}^T \mathbf{R}^{-1} \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \mathbf{R}^{-1} \mathbf{y}. \tag{A15}$$

Similarly, by solving $\mathrm{d}(-\log \mathcal{L})/\mathrm{d}\sigma^2 = 0$ for σ^2, the field variance is obtained as

$$\sigma^2 = \sigma^2(\boldsymbol{\theta}) = \frac{1}{N}(\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\beta})^T \mathbf{R}^{-1}(\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\beta}). \tag{A16}$$

From eq. (A15) and (A16) 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. A14–A16 this leads to the optimization problem

$$\boldsymbol{\theta} = \arg\min_{\theta} \left(\frac{1}{2}\log\left(\det(\mathbf{R})\right) + \frac{N}{2}\log\left(2\pi\sigma^2\right) + \frac{N}{2} \right). \tag{A17}$$

5   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,

$$R(\mathbf{x}, \mathbf{x}', \boldsymbol{\theta}) = \prod_{i=1}^{M} R(x_i, x_i', \theta_i). \tag{A18}$$

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

$$R(h, \theta) = \exp\left(-\frac{|h|}{\theta}\right). \tag{A19}$$

One of the possibilities for tuning the performance of a Kriging model is selecting different trend functions. If the trend function is replaced by a constant (i.e. the mean of the field) the resulting model is referred to as simple Kriging. A linear trend is denoted

15 as ordinary Kriging, while with any other more advanced function the model is called universal Kriging. In universal Kriging, the functional form of the mean field $\mathbf{f}(\mathbf{x})^T \boldsymbol{\beta}$ is identical to the generalized PCE defined in eq. (A8), 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 4.3.

     The main practical difference between regression- or expansion-type models such as regular PCE and the Kriging approach

20 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

25 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

30 variations in turbulence) the inferred correlation length will tend to zero and the Kriging estimator will be represented by the trend function alone.

A3 Sobol indices from the PCE

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),

$$\text{Var}\left[\tilde{S}(\boldsymbol{\xi})\right] = \text{Var}\left[\sum_{j=0}^{N_p-1} S_j \Psi_{\gamma,j}(\boldsymbol{\xi})\right] = \sum_{j=1}^{N_p-1} S_j^2 \text{E}\left[\Psi_{\gamma,j}^2(\boldsymbol{\xi})\right]. \tag{A20}$$

5   Each of the terms in the sum in eq. (A20) represents the contribution of the variables contained in the respective multivariate polynomials $\Psi_{\gamma,j}$ where $j = 0 \ldots N_p - 1$. 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. (17), 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

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

15   $$SU_{i_1,\ldots,i_s} = \left(\sum_{j \in \mathscr{F}_{i_1,\ldots,i_s}} S_j^2 \text{E}\left[\Psi_j^2(\boldsymbol{\xi})\right]\right) \cdot \left(\text{Var}\left[\tilde{S}(\boldsymbol{\xi})\right]\right)^{-1}. \tag{A21}$$

Based on eq. (A21) it is also straightforward to obtain the total Sobol indices for a given variable γ by summing all $SU_{i_1,\ldots,i_s}$ where $\gamma \in (i_1,\ldots,i_s)$. 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.

20   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.