Efficient Bayesian calibration of aerodynamic wind turbine models using surrogate modeling

Benjamin Sanderse[1], Vinit V. Dighe[1,2], Koen Boorsma[3], and Gerard Schepers[3]

[1]Centrum Wiskunde & Informatica, Amsterdam, the Netherlands
[2]Delft University of Technology, Delft, the Netherlands
[3]TNO, Petten, the Netherlands

Correspondence: Vinit Dighe (v.v.dighe@tudelft.nl)

Abstract. This paper presents an efficient strategy for the Bayesian calibration of parameters of aerodynamic wind turbine models. The strategy relies on constructing a surrogate model (based on adaptive polynomial chaos expansions), which is used to perform both parameter selection using global sensitivity analysis and parameter calibration with Bayesian inference. The effectiveness of this approach is shown in two test cases: calibration of airfoil polars based on the measurements from the DanAero MW experiments, and calibration of five yaw model parameters based on measurements on the New MEXICO turbine in yawed conditions. In both cases, the calibrated models yield results much closer to the measurement data, and in addition they are equipped with an estimate of the uncertainty in the predictions.

1 Introduction

Aeroelastic wind turbine models based on blade element momentum theory (BEM) are used extensively within the wind energy community for simulating rotor characteristics such as aerodynamic loads, power and thrust. They are indispensable tools for the design and optimization of wind turbines. However, in several situations, the accuracy of such models can be unsatisfactory when comparing the results of the model predictions with experiments (Buhl and Manjock, 2006). For instance, the 'blind comparison' study organized by NREL (Simms et al., 2001) revealed large differences when comparing the predictions of different aeroelastic models with experimental measurements. In some cases, differences exceeded 200%, even when simple operating conditions were being considered (i.e. uniform wind-speed, fixed blade pitch and zero yaw angle). The differences were attributed to the several empirical correction factors or tuning parameters integrated into the aeroelastic models that are used to improve the unsteady aerodynamic and aeroelastic force predictions. More recent results show better agreements, at least for simple wind tunnel conditions, but many challenges exist, for example in dynamic wake prediction and yaw, especially in the context of upscaling (see Schepers et al. (2021), Chapter 12). Examples are dynamic wake correction factors or dynamic stall model parameters (Wang et al., 2016). These empirical correction factors suffer from inherent uncertainties. As explained by Leishman (2002) and Sørensen and Toft (2010), a major challenge is to identify the uncertainties associated with wind turbine aerodynamics in order to develop more rigorous models suitable for a wider range of operating conditions, and to better integrate and validate these models with reference to good quality experimental measurements. Similarly, Abdallah et al. (2015) concluded that the uncertainties in the model parameters used in aeroelastic models have a significant impact on the

25 accuracy of model predictions. In other words, in order to build robust aeroelastic wind turbine models with a quantified level of uncertainty, it is important to calibrate these models in a framework that includes uncertainty estimates (Murcia, 2016).

A common approach to calibrate aerodynamic models is via parameter tuning, in which one assumes that the form of the model is in principle correct, and that the errors in the model outcomes can be reduced by properly choosing the value of one or more parameters. These parameter values are preferably independent of the model inputs, i.e. they should lead to accurate

30 predictions for a wide range of operating conditions. Examples of parametric model calibration in wind energy applications can be found in Bottasso et al. (2014), Murcia et al. (2018) and van Beek et al. (2021). In these calibration studies, either least-squares methods or maximimum likelihood estimation (MLE) methods are used. MLE determines the model parameters such that it maximises the likelihood that describes the (presumed) relation between model and measurement data (Severini, 2000). However, a major drawback of least-squares and MLE methods is that prior information is not naturally included (Smith,

35 2013). Using prior information is especially relevant when few measurement data is available, which is a common situation in wind turbine model calibration (van Kuik et al., 2016). Consequently, the MLE method can exhibit large uncertainty in the estimation of the parameters and, as a result, in the model predictions. Furthermore, in least-squares or MLE methods, parameters are typically considered deterministic (fixed but unknown) so that a point estimate (plus confidence intervals) results, which does not provide details regarding the full probability distribution of the calibrated parameters (Smith, 2013).

40 In order to address these issues, the goal of this paper is to set-up a framework for calibrating aerodynamic wind turbine models that also works in case of limited measurement data and gives full uncertainty estimates (in terms of probability density functions) of the calibrated parameters. We propose a rigorous approach to the calibration problem by recasting it in a probabilistic setting using a Bayesian framework (Kennedy and O'Hagan, 2001). Within this framework, the model parameters are posed as random variables, and it is possible to include prior knowledge by specifying a prior distribution, thus allowing

45 model calibration even when small sample sizes are available. Bayes' theorem (Bayes, 1763) is then used to calculate the posterior distribution of the model parameters conditioned on the given measurement data. The posterior distribution gives more information than MLE about the calibrated parameters, i.e. it gives the entire posterior probability density function, from which point estimates such as the posterior mean and the standard deviation can be calculated (if required). Furthermore, the calibration can be verified by computing the posterior predictive distribution (Gelman et al., 2013). Since the expression for the

50 posterior distribution is generally not available in an analytically tractable form (Gelman et al., 2013), we will resort to Markov Chain Monte Carlo (MCMC) methods to sample from the posterior distribution (Papageorgiou and Traub, 1996; Andrieu et al., 2003). The main downside of the Bayesian approach, associated to the MCMC sampling step, is its high computational expense. We will alleviate this issue by constructing a surrogate model of the full aerodynamic model (Sudret, 2008), and perform the MCMC sampling with the surrogate model in lieu of the full model. In this work, polynomial chaos expansions

55 (PCE) (Laloy et al., 2013) will be used, which can be constructed using a relatively small number of aerodynamic model runs.

In addition, the cost of the Bayesian calibration can be reduced by eliminating non-influential parameters. These can be determined by performing a sensitivity analysis (Oakley and O'Hagan, 2004). In this study we will employ a variance-based global sensitivity analysis using Sobol' indices (Sobol, 1993), based on our earlier work (Kumar et al., 2020). Determining Sobol' indices is straightforward once the PCE surrogate model has been determined.

60   The novelty of this work lies in the construction of a Bayesian framework for aerodynamic wind turbine model calibration. In addition to that, two realistic calibration studies based on the DanAero MW experiments (Madsen et al., 2010) and the New MEXICO experimental data (Boorsma and Schepers, 2016). The former dataset will be used to calibrate airfoil polars, while the latter will be used to calibrate yaw model parameters. The DanAero experiments were supplied by DTU within the framework of IEA Task 29. Although extensive comparisons between results from a large variety of codes (including the

65   AeroModule employed in this study) were performed on DanAero and NewMexico in Task 29 (Schepers et al., 2018, 2021), no thorough uncertainty analysis and calibration was performed yet. We stress that, even though these studies show a realistic application of our method with actual data, they correspond to idealized situations, and the main purpose of this paper is to demonstrate the calibration framework and its potential for application to a wide variety of wind engineering problems.

    The outline of this paper is as follows: Section 2 discusses the two experimental datasets considered in this study (DanAero

70   and New MEXICO). Section 3 describes the aerodynamic code used in this work (the so-called Aero-Module), plus parameterization of its inputs and outputs. The Bayesian calibration methodology, which is accelerated by constructing a PCE-based surrogate model, is detailed in section 4. Finally, the results of the calibration and discussion are presented in section 5 followed by conclusions drawn in section 6.

2   Experimental data description

75   In order to demonstrate the proposed Bayesian calibration framework, measurements from two experiments constitute the basis for the analysis, which are explained in sections 2.1 and 2.2.

2.1   DanAero MW experiment

The objective of the DanAero MW experiment was to provide an experimental basis that can improve the understanding of the fundamental aerodynamic and aeroacoustic phenomena using a full scale wind turbine model (Madsen et al., 2010). A

80   2.3MW NM80 turbine located at the Tjæreborg Enge site and a nearby met mast were both instrumented with various sensors. A LM38.8m test blade (schematic) was instrumented with pressure taps at four blade sections is shown in Figure 1 (left). The data acquisition rate was 35 Hz and in total about 275 ten minute time series were acquired between July and September 2009, which were made available for the present analysis.

    In the current study, we aim to calibrate airfoil polars (to be described later), and only a subset of the entire DanAero dataset

85   will be used. To be precise, data from 'Run 14' (a single 10-minute time series) on the first measurement day (16 July 2009) is used. This corresponds to a case with little yaw and shear and with little turbulence (roughly constant inflow conditions) under normal operation. The inflow velocity for this case is around 6 m/s and the rotational speed is 12 rpm. Within this particular 10-minute series, the data corresponding to $200s < t < 450s$ was used, in which almost constant wind and rotor speed were observed. For more details, we refer to Madsen et al. (2018). As a result, we have normal force measurements $y^{(\varrho)}(t)$ at the

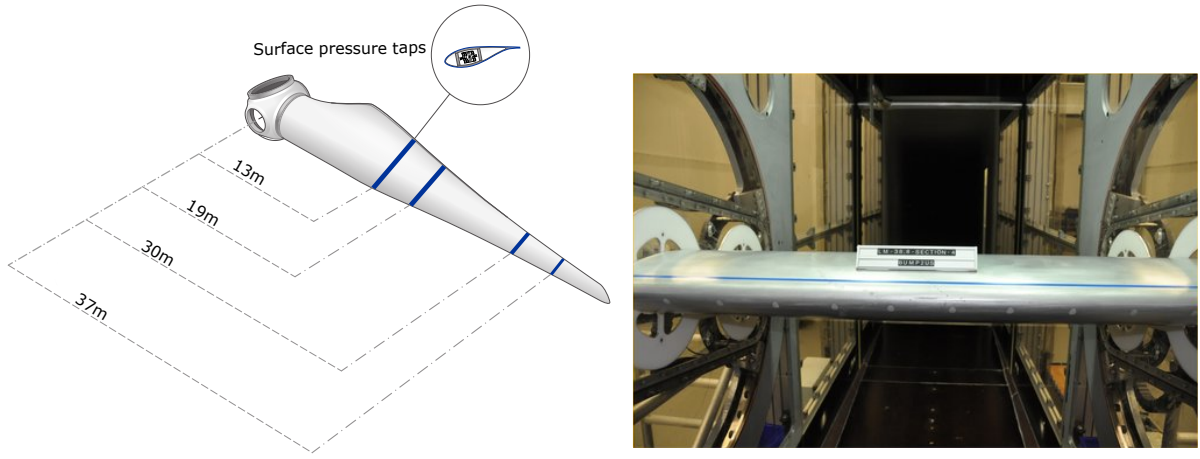



Figure 1. Schematic of NM80 wind turbine blade implemented with surface pressure taps (left) and NACA 63-418 airfoil blade section tested in the LM Wind Power wind tunnel in Lunderskov (right) (Özçakmak et al., 2018).

four blade sections ($\varrho = 1, 2, 3, 4$) at $N_t = 8750$ discrete time steps, gathered in the data matrix y:

$$
y = \begin{pmatrix} y^{(1)}(t_1) & \cdots & y^{(1)}(t_{N_t}) \\ \vdots & & \vdots \\ y^{(4)}(t_1) & \cdots & y^{(4)}(t_{N_t}) \end{pmatrix}, \tag{1}
$$

where $t_j = j\tau/N_t$, $\tau = 250$ seconds. The radial positions (measured from the center of the hub) corresponding to these sections are $r = (13.116, 19.06, 30.216, 36.775)$m, see Figure 1 (left). To obtain the distance from the blade root, as used in the Aero-Module calculations, we subtract the distance from the blade root to the hub centre which is 1.24m. Although using a single 10-minute time series corresponding to a single operating condition is generally insufficient to perform accurate BEM model calibration, this experiment is merely used as a first demonstration of our framework. A more advanced calibration run involving multiple operating conditions will be performed with the data from the New MEXICO experiment described in the next section.

Besides the data obtained directly from the Tjæreborg Enge site, airfoil polars were obtained from several wind tunnel tests, such as those on the NACA 63-418 airfoil cross-section in the LM Wind Power wind tunnel (Madsen et al., 2010); see Figure 1 (right). These airfoil polars consist of lift, drag, and moment coefficients as a function of angle of attack. Four polars are used in this study, whose location roughly (but not exactly) corresponds to the measurement positions mentioned above. These polars will form the inputs to our BEM code (see section 3.1) and are to be calibrated in this study.

2.2 New MEXICO experiment

The main objective of the New MEXICO experimental campaign was to create a database of detailed aerodynamic and load measurements on an experimental wind turbine to be used for computational models validation and improvement (Boorsma

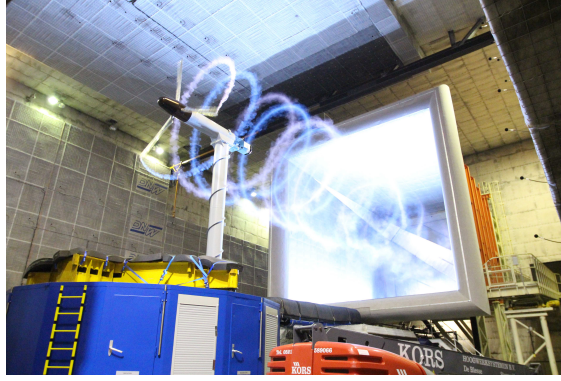



Figure 2. New MEXICO rotor tested in the DNW-LLF wind tunnel (Boorsma and Schepers, 2014, 2016).

and Schepers, 2014, 2016). To this aim, a three bladed 4.5 m diameter wind turbine model was built and tested in the large low speed facility of the German-Dutch wind tunnel (DNW-LLF) during a campaign in June-July 2014 (see Figure 2); a detailed description of the experiment is available in Boorsma and Schepers (2014). The data acquisition system consists of dynamic pressure sensors divided over five sections and distributed over three blades: at 25% and 35% (blade 1), 60% (blade 2), and 82% and 92% (blade 3) radial position, respectively. These were post-processed to obtain (amongst others) sectional normal forces, whose variations with azimuth and yaw angle will be considered in this study.

The corresponding operating conditions ('scenarios') are described by a vector $\mathcal{S}_i$:

$$\mathcal{S}_i = (\rho_\infty, V_\infty, \beta, \phi), \qquad i = 1, \ldots, N_S, \tag{2}$$

where ρ_∞ is the density, V_∞ is the inflow velocity, β is the yaw angle, ϕ is the pitch angle, and N_S the number of operating conditions. For the yawed flow case, 29 runs were performed in total, but in this study we restrict ourselves to $N_S = 3$, corresponding to IDs 935, 939 and 948 (see Table 1). These conditions are such that a significant induced velocity is expected (so that the yaw model will have a significant effect), while at the same time there is little dynamic stall occurring. For each operating condition, there is a corresponding dataset containing the normal force as a function of azimuth at five radial sections $\varrho = 1, \ldots, 5$ (the tangential force is also available, but not used here due to the large uncertainty associated with the measurements). Similar to equation (1), this will be denoted by the data matrix y, but now with an additional subscript i to indicate that there is a data matrix corresponding to each different operating condition:

$$y_i = \begin{pmatrix} y^{(1)}(t_1) & \cdots & y^{(1)}(t_{N_t}) \\ \vdots & & \vdots \\ y^{(5)}(t_1) & \cdots & y^{(5)}(t_{N_t}) \end{pmatrix}_i, \qquad i = 1, \ldots, N_S. \tag{3}$$

Table 1. New MEXICO operating conditions considered for yaw model calibration.

scenario	run ID	ρ_∞ (kg/m^3)	V_∞ (m/s)	β (deg)	ϕ (deg)
$\mathcal{S}_1$	935	1.20395	10.04	15.01	-2.3
$\mathcal{S}_2$	939	1.20395	9.98	30.01	-2.3
$\mathcal{S}_3$	948	1.20453	9.98	45.01	-2.3

3 Aerodynamic wind-turbine model

3.1 Aero-Module description and uncertainties

TNO (Netherlands Organisation for Applied Scientific Research) is the developer of a state-of-the-art aerodynamic model based on a BEM formulation, called the Aero-Module (Boorsma and Grasso, 2015) (formerly developed by ECN, now part of TNO). The model simulates the aerodynamic behavior of wind turbines by combining the concept of momentum conservation of the flow (BEM theory) and can be coupled to an aeroelastic model that solves the equations of motion for the structure, possibly extended with the hydrodynamics of the sea and control algorithms. In this work, we concentrate on the first aspect, namely the prediction of flow and blade forces as given by the BEM method. All calculations are done for a rigid construction, since the effects from flexibilities are considered small: in New Mexico a small rigid rotor was used, and for DanAero the elastic effects were found to be small (Schepers et al., 2021). A detailed description of the BEM approach within the Aero-Module is beyond the scope of the current discussion, and can be found in Boorsma et al. (2012). Important for the current discussion is to distinguish between different types of inputs in the Aero-Module. The first type of inputs consists of *external (operating) conditions*, such as wind speed and air density. The second type consists of *turbine specifications*, such as the blade geometry. The third type consists of *model parameters* inherent to the BEM formulation, such as lift- and drag-polars, tip correction factors, yaw model parameters, etc. For the case of a rigid turbine, with a uniform inflow field, the main uncertainties in this third type (the BEM model parameters) mainly arise from (Abdallah et al., 2015):

- Airfoil aerodynamics: The static airfoil data from wind tunnel experiments or from 2D airfoil codes, utilized as an input for the BEM simulations, have significant uncertainties and can be inaccurate (Bak et al., 2010).

- Empirical models: Several empirical models such as dynamic stall models, 3D correction models, and Prandtl correction models are used to include unsteady and 3D effects in BEM models (Wang et al., 2016; Schepers, 2012). It is often the choice of a designer to select between different empirical models, which can suffer from modelling uncertainty.

In the current study we will focus on calibrating this third type of input parameters, i.e. the model parameters, in particular static airfoil data (lift, drag and moment polars) and yaw model parameters. However, we stress that the calibration framework that is presented here can be directly applied to the first and second type of uncertainties as well.

In mathematical notation, these uncertainties will be captured in a vector of model parameters $\theta_M = (\theta_{M,1}, ..., \theta_{M,N_\theta}) \in \mathbb{R}^{N_\theta}$, to be described in more detail in section 3.2. The Aero-Module for a certain wind turbine is denoted by $\mathcal{M}$ and returns a

150 vector of outputs Y, depending on the (uncertain) model parameters θ_M and on the (given) value of the operating conditions $\mathcal{S}_i$:

$$Y_i = Y_i(\theta_M) = \mathcal{M}(\theta_M, \mathcal{S}_i), \qquad i = 1, \ldots, N_S, \tag{4}$$

where N_S is the number of operating conditions. The output Y contains amongst others forces, moments, power, etc, which are generally time-dependent. Typically only a subset of the entire Aero-Module output, indicated as the quantity of interest Q,

155 will be used to perform sensitivity analysis and calibration of the model. This will be further described in section 3.3. It should be stressed that θ_M, Y and Q are random vectors, each of which is associated with a joint probability density function.

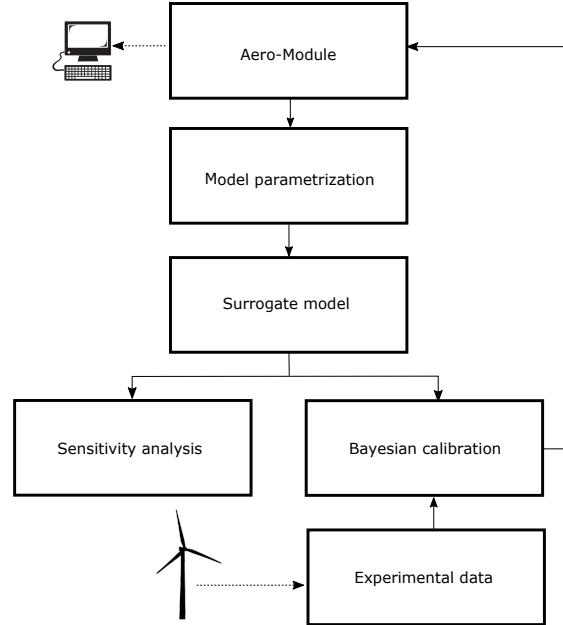



Figure 3. Flowchart for our Bayesian calibration framework.

The analysis carried out in this study is based on the procedure shown in Figure 3, including the steps that will be separately described in the following sub-sections. The proposed Bayesian calibration approach is performed using the UQLab uncertainty quantification software (Marelli and Sudret, 2014), especially the recently developed Bayesian inversion module (Wagner et al.,

160 2019).

3.2 Input parametrization

3.2.1 DanAero case: uncertainty in polars

For the DanAero case, we will consider the uncertainties associated with the airfoil aerodynamics: lift coefficient (C_l), drag coefficient (C_d) and moment coefficient (C_m). These coefficients are functions of both angle of attack α and radial position

165 r along the blade (also Reynolds number and Mach number, but this dependence is not studied here); this give rise to a very large number of uncertain parameters. In order to reduce this number, we parametrize these uncertainties as a function of angle of attack and radial position.

The parametrization as a function of radial position is automatically accounted for within the Aero-Module code: the user has to provide the lift, drag and moment polars only for a few airfoil sections along the radius of the blade, e.g. $C_{l,j}(\alpha)$ for

170 $j = 1 \dots N_{\text{sec}}$; N_{sec} being the number of airfoil sections. The Aero-Module interpolates these polars to other radial positions based on the relative airfoil thickness.

The parametrization as a function of angle of attack is performed as follows. Given a reference polar, e.g. $C_{l,\text{ref},j}(\alpha)$ for the lift coefficient at airfoil section j, a perturbed polar $\tilde{C}_{l,j}$ is obtained by scaling the reference curve as follows:

$$\tilde{C}_{l,j}(\alpha) = \begin{cases} C_{l,\text{ref},j}(\alpha) \cdot (1 + \Delta C_{l,j}) & \alpha_{\min,j} < \alpha < \alpha_{\max,j}, \\ C_{l,\text{ref},j}(\alpha) & \text{otherwise.} \end{cases} \tag{5}$$

175 The same equation is used for the drag and moment coefficients. The value of $\Delta C_{l,j}$ determines how much the reference curve is scaled. The bounds $\alpha_{\min,j}$ and $\alpha_{\max,j}$ indicate for each airfoil section j which part of the polar is perturbed. The unperturbed and perturbed parts of the polar are combined via a NURBS curve. A similar equation holds for the drag and moment coefficients. Example curves obtained with different realizations of ΔC_l, ΔC_d and ΔC_m are shown in Figure 4.

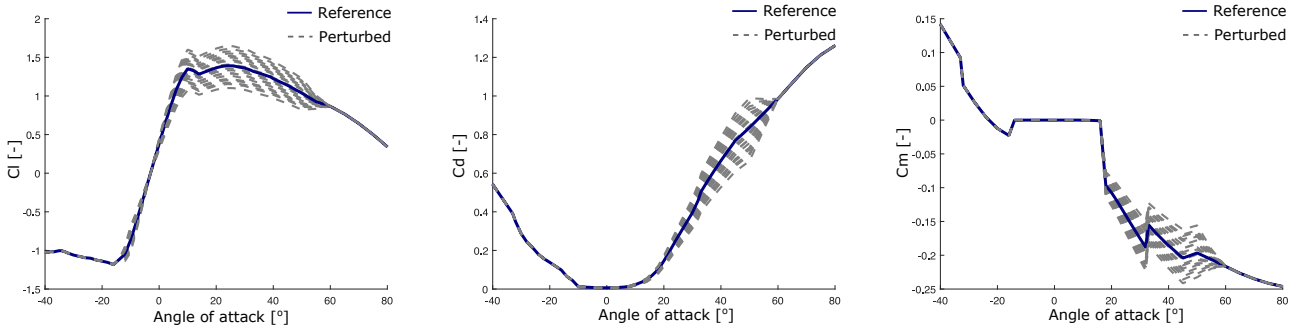



Figure 4. Examples of perturbed C_l, C_d and C_m polars at section 2 as described by equation (5), with $\alpha_{\min} = -10°$ and $\alpha_{\max} = 50°$.

For the DanAero case, where the number of airfoil sections is $N_{\text{sec}} = 4$, the parameterization of lift, drag and moment

180 coefficients leads to the following $N_\theta = 12$-dimensional parameter vector θ_M:

$$\theta_M = (\Delta C_{l,1}, \dots, \Delta C_{l,4}, \Delta C_{d,1}, \dots, \Delta C_{d,4}, \Delta C_{m,1}, \dots, \Delta C_{m,4}). \tag{6}$$

One advantage of the multiplicative-type of parameterization (5) is that the uncertainty becomes largest when the magnitude of the reference curve is large; this is physically meaningful, as lift curves tend to be most uncertain around the region of maximum lift (and/or at high angles of attack, but these are not considered here). However, other types of parameterization

185 could be considered. For example, Bottasso et al. (2014) considered an additive-type of correction (i.e. $C_{l,\text{ref}} + \Delta C_l$), with ΔC_l

expressed in terms of shape functions and coefficients, and applied a decorrelation procedure to improve the identifiability of the drag coefficients. Matthäus et al. (2017) obtained a perturbed lift curve by interpolating between two reference lift curves corresponding to clean and rough states. In Abdallah et al. (2015), a number of typical points along the $C_l(\alpha)$ curve (e.g. maximum lift, separation point) was used to construct a parametric spline approximation to the lift curve. Since the focus of this article lies in showing how the combination of surrogate models, sensitivity analysis and Bayesian inference can be used for efficient calibration, we have not considered such more advanced polar parameterizations, but we note that it is possible to use any of them within our calibration framework.

3.2.2 New MEXICO case: uncertainty in yaw model

With the New MEXICO experiments, as described in section 2.2, the goal is to calibrate a set of parameters that determine the yaw model of the Aero-Module. This yaw model is described in Schepers (2012) and consists of 10 amplitude coefficients denoted by AM_{kl} and 10 phase coefficients denoted by PH_{kl} ($k = 1, 2$; $l = 1 \ldots 5$), which are used in an equation for the induced velocity in yawed conditions (see equation (C1) in appendix C). We will (as a proof of concept) calibrate only the first five parameters of this model, i.e. we take

$$\theta_M = (AM_{11}, AM_{12}, AM_{13}, AM_{14}, AM_{15}). \tag{7}$$

The nominal values for these coefficients can be found in Appendix B in Schepers (2012) and are repeated in appendix C. The other yaw model parameters are associated to the phase shift of the induced velocity and to higher order harmonics and will be left at their nominal values, since the amount of experimental data considered here is too limited to perform a sensible calibration.

3.3 Output parameterization and quantity of interest

The Aero-Module predictions given by equation (4) involve a large set of time-dependent quantities, making the dimension of the output effectively very high-dimensional. For the purpose of sensitivity analysis and model calibration, it is highly desirable to reduce the dimensionality of the output. As a first step (both for DanAero and New MEXICO simulations), out of all possible outputs (forces, moments, power, etc.) we restrict ourselves to the normal forces F_N ('normal' indicating normal to the chord), interpolated to the radial positions corresponding to the measurement positions.

3.3.1 DanAero case: time-independent results

For the DanAero case, the inflow conditions in the Aero-Module are assumed constant in time, and given that there is no shear or yaw, this results in normal force predictions that are steady state (time-independent). The experimental data are, on the other hand, only approximately steady state since they were performed in atmospheric conditions. The question then arises how to perform the comparison between simulation and experiment in order to perform the desired calibration. The most natural possibility would probably be to average the experimental data in time. However, since we have only a single 10-minute time series at this condition, this would effectively reduce the number of measurement points to just a single point, which would be

too little to perform any sensible calibration. As a compromise, we decided to split the time series into a number of subsets (10, 50, 100 and 200) of the 8750 datapoints mentioned in equation (1), at regularly spaced intervals. The mean of each subset is then considered to be an independent measurement point that will be used in the calibration process (as if they were corresponding to different measurements). This approach allows us to clearly show the effect of increasing the number of measurement points on the convergence of the posterior distribution of the airfoil polars. This will be further detailed in section 5.1.

3.3.2 New Mexico case: time-dependent results

For the New MEXICO case, the operating conditions lead to results that are periodic in time. There are five measurement positions, and after radial interpolation of the Aero-Module output to these positions, we effectively have for each operating condition $\mathcal{S}_i$ an output matrix of the form:

$$
Y_i(\theta_M) = \begin{pmatrix} F_N^{(1)}(t_1)\dots F_N^{(1)}(t_{N_t}) \\ F_N^{(2)}(t_1)\dots F_N^{(2)}(t_{N_t}) \\ F_N^{(3)}(t_1)\dots F_N^{(3)}(t_{N_t}) \\ F_N^{(4)}(t_1)\dots F_N^{(4)}(t_{N_t}) \\ F_N^{(5)}(t_1)\dots F_N^{(5)}(t_{N_t}) \end{pmatrix} \in \mathbb{R}^{5\times N_t}. \tag{8}
$$

Here we write for notational simplicity that the time instances $t_1 \dots t_{N_t}$ of the simulation are the same as the measurement time instances; see equation (1). This is in general not the case, but is not very important, because of the dimensionality reduction technique that we use to compress the output data, which is described next.

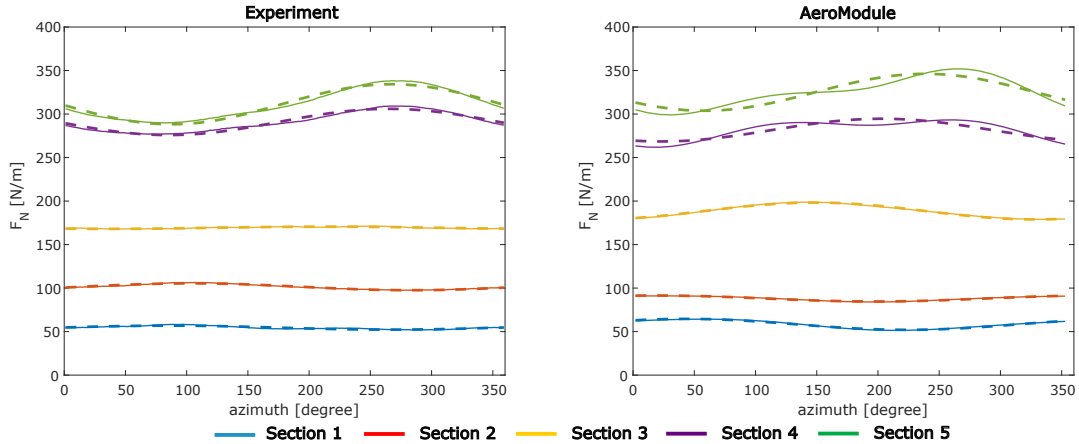



Figure 5. Representation of measurements (left, solid lines) and Aero-Module output (right, solid lines) using three Fourier coefficients (dashed) at different radial sections, for scenario $\mathcal{S}_1$ (run 935) of the New MEXICO experiment (last revolution shown).

Since the normal force is relatively smooth in time, the solutions at different time instances are highly correlated, and dimensionality reduction techniques can be effectively applied. Commonly used techniques are based on principal component

analysis (PCA) or the related singular value decomposition (SVD), see for example (Bottasso et al., 2014) and (Wagner et al., 2020). In this work, the normal forces are periodic in time and a suitable reduction technique is to decompose the output signal into Fourier modes via a discrete Fourier transform:

$$\hat{F}_N^{(\varrho)}(k) = \frac{1}{N_t} \sum_{j=0}^{N_t-1} F_N^{(\varrho)}(t_j) \exp^{-i2\pi jk/N_t}, \qquad k = 0, 1, \ldots, N_k - 1, \tag{9}$$

where $i = \sqrt{-1}$ and the resulting coefficients $\hat{F}_N^{(\varrho)}(k)$ are complex valued. Note that both F_N and $\hat{F}_N$ effectively depend on the parameters θ_M, but this dependence is omitted here to keep the notation concise. The normal force at a given section (ϱ) is then approximated by keeping N_k Fourier coefficients (ordered as $k_1, k_2, \ldots$) that correspond to the modes that have the largest power spectral density, plus the mean of the signal ($k = 0$). The selection of the PSD peaks was easy to automate since the peaks are easily distinguishable from any background noise, and the signals are well represented in terms of a few Fourier coefficients. We expect this to be true also for different operating conditions, although we recommend as a best practice to plot the original output alongside the Fourier representation when moving to new test cases.

An example of the Fourier representation of the normal force with three coefficients is shown in Figure 5 (right), together with the experimental results and their Fourier representation (also with three coefficients) on the left. The physics of the yaw model (Schepers, 2012) is such that its parameters are meant to change the amplitude and the phase shift of the normal forces (via the induced velocities), and *not* their mean value. Therefore, the mean of the signal will be left out from the calibration. Furthermore, as we focus on calibrating the parameters of the yaw model that are *amplitude* coefficients (see equation (7)), we will use the amplitude of the first mode (and not the phase shift).

3.3.3 Summary

The two previous sections can be summarized by introducing the quantity of interest Q_i for a certain operating condition $\mathcal{S}_i$ and model parameters θ_M as

$$Q_i(\theta_M) := \mathcal{Q}\big(Y_i(\theta_M)\big) \in \mathbb{R}^{N_{\mathrm{sec}} \times N_k}, \qquad i = 1, \ldots, N_S. \tag{10}$$

For the DanAero case we have

$$\text{DanAero:} \qquad Q = \begin{pmatrix} \hat{F}_N^{(1)}(0) \\ \hat{F}_N^{(2)}(0) \\ \hat{F}_N^{(3)}(0) \\ \hat{F}_N^{(4)}(0) \end{pmatrix}, \tag{11}$$

while for the New MEXICO case we have

$$\text{New MEXICO:} \qquad Q_i = \begin{pmatrix} |\hat{F}_N^{(1)}(k_1)| \\ |\hat{F}_N^{(2)}(k_1)| \\ |\hat{F}_N^{(3)}(k_1)| \\ |\hat{F}_N^{(4)}(k_1)| \\ |\hat{F}_N^{(5)}(k_1)| \end{pmatrix}, \qquad i = 1, 2, 3. \tag{12}$$

4 Methodology

4.1 PCE-based surrogate model

In order to perform parameter sensitivity and parameter calibration, typically a high number of computationally expensive Aero-Module runs $\mathcal{M}(\theta_M)$ are required. To reduce the computational time, a surrogate model or emulator is constructed (for the quantity of interest) and used in lieu of the full model. Examples of some popular surrogate models include Kriging (Gaussian process regression), polynomial chaos expansion (PCE), support vector machines (SVM) and radial basis functions (RBF) (Schöbi, 2019). In this study, a PCE-based surrogate model will be used because PCE has been found to be an efficient method in computing the stochastic responses of complex computational models (Soize and Ghanem, 2004; Guo et al., 2018; Dutta et al., 2018).

The PCE surrogate model is constructed to approximate the quantity of interest as predicted by the aerodynamic model:

$$\mathcal{Q}_i^{PC}(\theta_M) \approx \mathcal{Q}_i(\theta_M), \qquad i = 1, \ldots, N_S, \tag{13}$$

where the subscript i indicates that a different surrogate model is built for each operating condition $\mathcal{S}_i$, which depends on the same parameter set θ_M. This subscript will be left out in what follows if no confusion can arise. A PCE approximation $\mathcal{Q}^{PC}(\theta_M)$ of the aerodynamic model $\mathcal{Q}(\theta_M)$ can be defined as a weighted sum of multivariate polynomials in θ_M (Marelli and Sudret, 2019; Smith, 2013):

$$\mathcal{Q}^{PC}(\theta_M) = \sum_{\mathbf{k} \in \mathcal{K}} w_{\mathbf{k}} \Psi_{\mathbf{k}}(\theta_M), \tag{14}$$

where $\Psi(\theta_M)$ is the multivariate polynomial basis, and $w_{\mathbf{k}}$ is the coefficient corresponding to basis function $\Psi_{\mathbf{k}}$. $\mathbf{k}$ is the multi-index and $\mathcal{K}$ is the set of multi-indices describing which polynomial basis functions are used. The set $\mathcal{K}$ in (14) depends on the truncation scheme; in this work, a hyperbolic truncation scheme is used with truncation parameter equal to 0.75 (Blatman, 2009). Furthermore, an adaptive strategy is followed in which sparse PCE expansions are pursued, by favouring low rank truncation schemes (i.e. penalizing the norm $\|\mathbf{w}\|_1$). To achieve this, the coefficients $w_{\mathbf{k}}$ are computed from the following adapted least-squares minimization problem (Marelli and Sudret, 2019):

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathbb{E}\left[\left(\mathbf{w}^{\mathbf{T}}\Psi(\theta_M) - \mathcal{Q}(\theta_M)\right)^2\right] + \lambda\|\mathbf{w}\|_1. \tag{15}$$

This equation is solved with the Least-Angle Regression (LARS) algorithm (Efron et al., 2004), given that a set of N samples of θ_M have been provided, which we denote by $\theta_M^{(n)}$, with $n = 1 \ldots N$. We will use Latin Hypercube Sampling to obtain these samples.

The LARS algorithm in the context of PCE starts with all PCE coefficients set to zero, and then iteratively selects polynomials based on the correlation with the current residual. After every iteration an a-posteriori error, namely the Leave-One-Out (LOO) cross-validation error ϵ_{LOO}, is computed (Marelli and Sudret, 2019):

$$\epsilon_{LOO} = \frac{1}{N}\sum_{n=1}^{N}\left(\mathcal{Q}(\theta_M^{(n)}) - \mathcal{Q}^{PC \setminus (n)}(\theta_M^{(n)})\right)^2, \tag{16}$$

where $\mathcal{Q}^{PC\backslash(n)}$ denotes the PCE surrogate model trained by leaving the n-th sample out. The surrogate model with the smallest ϵ_{LOO} is then chosen as the best PCE model.

4.2 Sensitivity analysis

Sensitivity analysis aims at finding which input parameters θ_M of the Aero-Module explain at best the uncertainties or variations in the model predictions. Sensitivity analysis aids in identifying non-influential parameters that can subsequently be fixed at their nominal values in the calibration process. In this work, a so-called global sensitivity analysis using a variance-based Sobol' decomposition technique is performed. For the sake of conciseness, we describe this technique briefly; a more detailed description in the same context of aerodynamic wind turbine models is available elsewhere Kumar et al. (2020).

The idea of a variance-based analysis is to relate the variance in the model inputs to the variance in the model output. The Sobol' indices are defined as a ratio of variances. An important advantage of using PCE as surrogate model is that once the PCE coefficients are determined, the first order and the total order Sobol' indices can be obtained directly without any additional model evaluations. In this work the total order Sobol' index S_i^T, corresponding to $\theta_{M,i}$, is used and is given by:

$$S_i^T = \frac{1}{D} \sum_{\mathbf{k} \in \mathcal{K}_i} w_{\mathbf{k}}^2, \qquad i = 1, \ldots, N_\theta, \tag{17}$$

where $\mathcal{K}_i$ is a subset of $\mathcal{K}$ which consists of the set of multivariate polynomials that are non-constant in the i-th input parameter $\theta_{M,i}$, and $D = \mathrm{Var}[\mathcal{Q}^{PC}(\theta_M)]$ is the variance of the PCE. The total sensitivity indices can be interpreted as an importance measure for the parameter $\theta_{M,i}$: a large S_i^T implies, roughly speaking, that $\theta_{M,i}$ has a strong influence on Y. These total indices include possible interaction effects between the parameters, which can be excluded by looking at the first order indices. For the New MEXICO test case such an interaction is indeed present, but since it does not change the conclusions from the analysis, this will not be further reported here. Note that the sensitivity analysis is performed without taking into account any measurement data, it is purely model-based. Furthermore, it should be noted that the analysis assumes that the parameters $\theta_{M,i}$ are independent.

4.3 Bayesian calibration

A widely used Bayesian calibration framework has been introduced by Kennedy and O'Hagan (2001). The framework can be used to predict the 'true' behavior of a computational model by calibrating model parameters θ_M to make the model predictions Y most likely to represent measurements y. We assume that the discrepancy E between the PCE approximation to the Aero-Module prediction, $\mathcal{Q}^{PC}$, and the measurement data y is of additive type, so that we can write:

$$Q_i^d = Q_i^{PC}(\theta_M) + E, \qquad i = 1, \ldots, N_S. \tag{18}$$

The subscript i corresponds again to operating condition $\mathcal{S}_i$. Q^{PC} is the PCE approximation to the quantity of interest given by a few Fourier coefficients computed from the Aero-Module output Y; see equation (10). Q^d is the quantity of interest for the measurement data y, which is determined in a similar fashion. The discrepancy term E accounts for both model error and

measurement errors, and is assumed to be a normally distributed random vector, written as:

$$E \sim \mathcal{N}\big(0, \Sigma(\theta_E)\big), \tag{19}$$

where $\mathcal{N}\big(0, \Sigma(\theta_E)\big)$ denotes the multivariate normal distribution with zero mean value and diagonal covariance matrix Σ parameterized by a set of variance parameters $\theta_E \in \mathbb{R}^{N_{\mathrm{sec}} \cdot N_k}$. E will typically have a dependence on the operating condition, but for sake of simplicity this is not considered in our test cases. We furthermore note that for the sake of simplicity, and also due to the lack of knowledge of the model bias term, the discrepancy term is assumed to have a zero mean. This is a commonly used approach in Bayesian model calibration, meaning that on average, we believe the model is able to reproduce the data. More advanced approaches are possible (e.g. using a Gaussian process to model the discrepancy), also in the context of UQLab (by providing a user-defined likelihood function).

The parameters θ_E are known as hyperparameters and will be calibrated together with the model parameters θ_M. The combined parameter vector $\theta = (\theta_M, \theta_E)$ is assumed to be distributed according to a so-called prior distribution $\pi(\theta)$:

$$\pi(\theta) = \pi(\theta_M)\pi(\theta_E), \tag{20}$$

where we have assumed that the prior on the model parameters and on the hyperparameters are independent. The Gaussian discrepancy model from equation (19), induces the following likelihood function:

$$\mathcal{L}(\theta; Q^d) = \mathcal{N}\big(Q^d | Q^{PC}(\theta_M), \Sigma(\theta_E)\big). \tag{21}$$

The expression for the posterior distribution of the parameters θ then follows from Bayes' theorem (Gelman et al., 2013):

$$\pi(\theta | Q^d) = \frac{\mathcal{L}(\theta; Q^d)\pi(\theta)}{Z} \qquad \text{with} \qquad Z = \int \mathcal{L}(\theta; Q^d)\pi(\theta)\, d\theta, \tag{22}$$

where $\pi(\theta | Q^d)$ is the posterior distribution and Z is the normalizing factor called the evidence (the integration is over the domain of θ). The posterior distribution $\pi(\theta | Q^d)$ in equation (22) can be interpreted as degree of belief about the parameters θ given the measurement data Q^d. Commonly reported point estimates derived from the posterior are the mean and the Maximum a Posteriori (MAP) estimate, defined as the value where the posterior distribution is maximum, i.e. $\theta_{\mathrm{MAP}} = \arg\max_\theta \pi(\theta | Q^d)$. We will also report the posterior predictive, which is obtained by propagating the posterior distribution, given by (22), through the model:

$$\pi(\hat{Q}^d | Q^d) = \int \mathcal{L}(\theta; \hat{Q}^d)\pi(\theta | Q^d)\, d\theta. \tag{23}$$

Here $\hat{Q}^d$ represent future observations of the quantity of interest, so the posterior predictive expresses the probability of observing new data $\hat{Q}^d$ given existing data Q^d. The posterior predictive is computed by using the samples of the posterior and evaluating the likelihood from the PCE model evaluations (adding an independently sampled discrepancy term) (Wagner et al., 2019).

The computation of the high-dimensional integral Z in equation (22) is not tractable for a general model $\mathcal{Q}^{PC}(\theta_M)$. The computation of Z can be circumvented by using Markov chain Monte Carlo (MCMC) methods, which avoid the need to

compute Z. MCMC techniques construct Markov chains to produce samples distributed according to the posterior distribution. With these samples, the posterior characteristics can be evaluated. In this work, we will use the so-called affine-invariant ensemble sampler (AIES) (invariant to affine transformations of the target distribution), which requires little tuning and is suitable for cases where strong correlations exist between the parameters (Goodman and Weare, 2010). However, each posterior sample still requires an evaluation of the likelihood (cf. equation (22)), so even with AIES thousands of model evaluations are still needed to obtain an accurate posterior. The PCE-based surrogate model $\mathcal{Q}^{PC}$ will therefore be used in place of the full Aero-Module $\mathcal{Q}$.

5 Results and discussion

In this section, the framework presented in section 4 is applied to (i) calibrate the sectional lift polars that are input to the Aero-Module using the DanAero MW experiment, and (ii) to calibrate the yaw model parameters of the Aero-Module using the New MEXICO experimental data set.

5.1 Lift polar calibration with DanAero data

As described in section 2.1, the NM80 turbine at a mean wind speed at hub height set to 6.1 m/s will be considered. The turbine rotational speed is set at 12.3 rpm, the pitch angle to 0.15 degree and the yaw angle is set to zero.

5.1.1 Sensitivity analysis

In order to build the PCE surrogate model (14), the Aero-Module is evaluated at a number of random samples of the parameter vector θ_M (given by equation (6)). We specify a normal distribution for each component of θ_M:

$$\theta_{M,i} \sim \mathcal{N}(0, \sigma_C), \qquad i = 1, \ldots, N_\theta, \tag{24}$$

where C stands for either ΔC_l, ΔC_m or ΔC_d. We take $\sigma_C = 0.125$ (for all sections, and for both lift, drag, and moment coefficient). This choice is such that the original polar is perturbed around its mean and that 95% of the samples will fall within $\pm 25\%$ of the unperturbed value. It also encodes our belief that very large perturbations from the original polar are less likely than small perturbations. To avoid unphysical realizations (unlikely but not impossible), the normal distribution is truncated to have a bounded support of $[-0.5, 0.5]$. The resulting perturbed polars follow from equation (5), and examples are shown in Figure 4.

$N = 32$ model evaluations were sufficient to achieve an LOO-error (equation (16)) smaller than 10^{-3} (for details we refer to appendix B). The total order Sobol' indices S_i^T following equation (17) are computed with the PCE surrogate model, as explained in sections 4.1-4.2. The resulting Sobol' indices expressing the sensitivity of the mean normal force at each radial section towards the perturbation in the lift, drag, and moment coefficients are shown in Figure 6. Note that we report here the sensitivity indices for the sectional normal forces, which is an extension compared to our sensitivity analysis (Kumar et al., 2020), where we considered the total normal force. Figure 6 indicates that the variation in the normal forces can be completely

attributed to the variation in the lift coefficients. This conclusion is in line with what has been reported for the total normal force (Kumar et al., 2020). However, we should note that it was not trivial to obtain these results, because there was an inconsistency between the provided ('planform') thickness distribution of the blade and the provided thickness of the four airfoil sections. We have therefore changed the airfoil thicknesses as is explained in Appendix B2.

With the corrected thickness distribution the sensitivity analysis confirms what we know from BEM theory, namely that the sectional normal force $\mathrm{d}F_N$ depends on C_l and C_d via

$$\mathrm{d}F_N = \mathrm{d}L(\alpha)\cos\alpha + \mathrm{d}D(\alpha)\sin\alpha \qquad \text{where} \qquad \begin{cases} \mathrm{d}L(\alpha) & = C_l(\alpha)\frac{1}{2}\rho V^2 c\,\mathrm{d}r, \\ \mathrm{d}D(\alpha) & = C_d(\alpha)\frac{1}{2}\rho V^2 c\,\mathrm{d}r. \end{cases} \tag{25}$$

Here $\mathrm{d}L$ and $\mathrm{d}D$ are the sectional lift and drag forces respectively, α is the local angle of attack, V is the relative velocity, dr indicates a spanwise section, and c is the local chord length. Since the angle of attack is only a few degrees at the measurement stations under consideration, and the drag coefficient at small α is much smaller than the lift coefficient at small α, the normal force is dominated by the lift coefficient. Note that since we consider the relation between normal force (normal to the chord) and lift, the twist or pitch angle of the blade does not enter in equation (25). The moment coefficient does not influence the normal force, which is consistent with aerodynamic theory.

Given that the sensitivity of the normal force is dominated by the lift coefficients, we will consider only the calibration of the lift coefficients in the next section. Calibrating the drag or the moment coefficients would require one to either use different measurement data (e.g. tangential force measurements) or to use a more advanced technique to deal with the low identifiability of the drag coefficients, such as the SVD-based decorrelation technique proposed by Bottasso et al. (2014). Since the purpose of this test case is mainly to show an example of our methodology (sensitivity analysis and Bayesian inference with surrogate models), and not to accurately calibrate the airfoil polars over several operating conditions, this is not considered here.

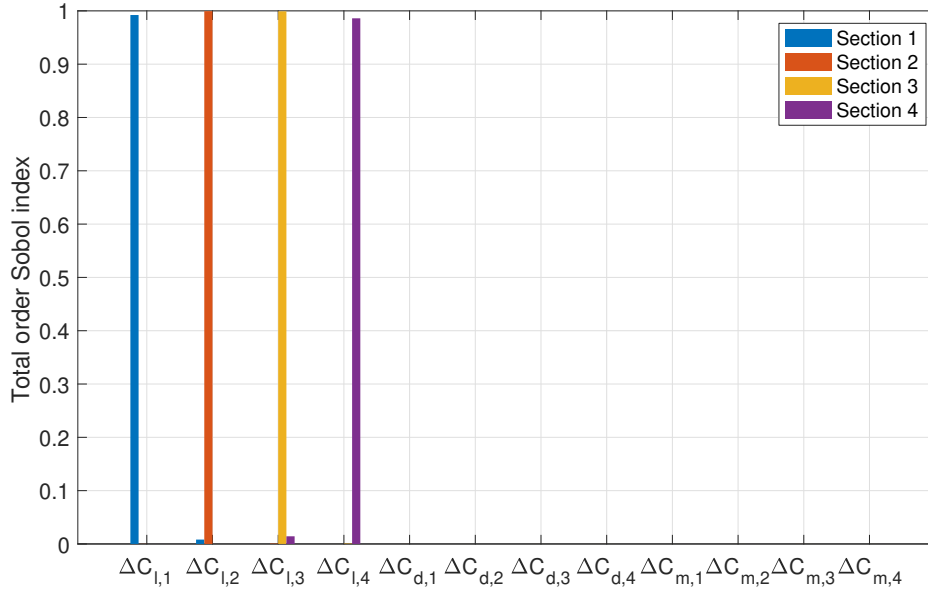



Figure 6. Sensitivity of sectional normal force with respect to perturbations in airfoil polars with adapted thickness.

5.1.2 Calibration

Following the sensitivity analysis, Bayesian calibration was performed for 8 parameters, namely the 4 model parameters $\Delta C_{l,1}\ldots\Delta C_{l,4}$ and the 4 discrepancy parameters $\theta_{E,1}\ldots\theta_{E,4}$, using the normal force measurements obtained from the DanAero MW experiment. The surrogate model constructed in section 5.1.1 is retrained, without the drag and moment co-efficients as uncertain parameters, again using $N = 32$ runs of the Aero-Module. Since the sampling method is random (LHS), the surrogate model used for the calibration can in principle be somewhat different from the one used for the sensitivity analysis. In practice the surrogate model for calibration will be even more accurate, since it involves fewer parameters.

In the Bayesian analysis, the prior on the model parameters is taken the same as in equation (24). The prior on the discrepancy parameters is taken as a uniform distribution:

$$\theta_{E,i} \sim \mathcal{U}(0, \sigma_E^2), \qquad i = 1, \ldots, 4. \tag{26}$$

Note that θ_E and σ_E^2 model the variance between model and data, and are therefore positive quantities. We take $\sigma_E^2 = 5 \cdot 10^4$ (so that $\sigma_E \approx 223$ N/m), which we determined by considering the standard deviation in the measurement data (around 100 N/m) and doubling this value to get a sufficiently broad prior. One could argue that this error should depend on the radial position along the blade, but this was not assumed in our prior specification. Instead, we did not want to introduce too much (possibly wrong or biased) a priori knowledge about the radial dependence but let the calibration process 'do the job'.

The PCE-based surrogate model for the quantity of interest, $Q^{PC}(\theta_M)$, is used in place of the Aero-Module throughout the analyses. The AIES algorithm with 10^2 parallel chains and 10^3 steps is deployed (in total 10^5 MCMC iterations and

concomitant surrogate model evaluations). Convergence is assessed based on the Gelman-Rubin diagnostic (Wagner et al., 2019) and visual inspection of the MCMC trace plots (see Figure B4 in the appendix), and a burn-in of 50% is used. With the full Aero-Module, this would take several weeks to compute on a desktop computer. By using the surrogate model instead, this is reduced to less than an hour.

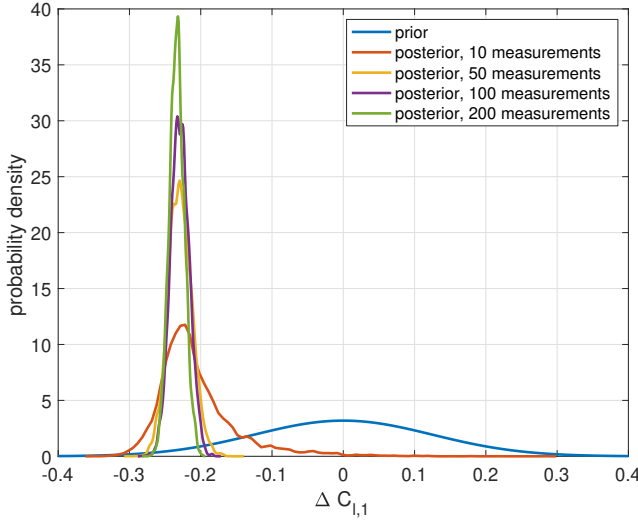
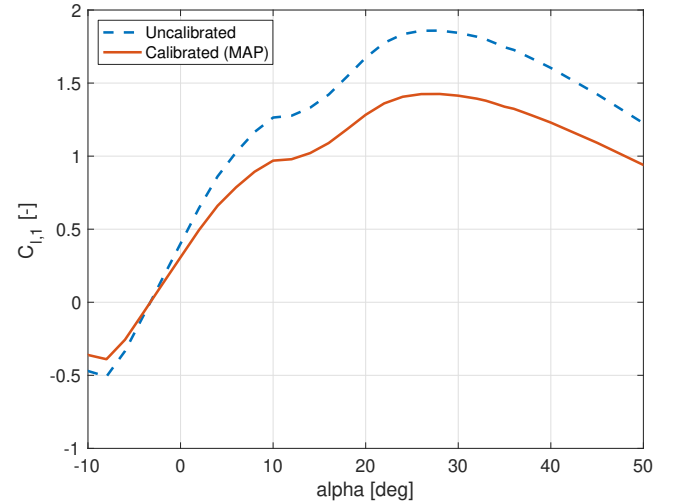
As discussed in section 3.3.1, the number of 'measurements' is varied to illustrate the effect on the posterior distribution of the parameters. As an example, Figure 7a illustrates how the prior on $\Delta C_{l,1}$ (truncated normal) becomes more and more dominated by the data when the number of measurement points increases. The marginal posteriors for the other parameters show similar behavior. If additional datapoints would be included, the posterior would become even stronger peaked (almost independent of the prior distribution).

(a) Posterior of $\Delta C_{l,1}$ with increasing number of measurements.

(b) The calibrated $C_{l,1}$ polar evaluated at the MAP of the posterior (obtained with 200 measurements), compared to the uncalibrated polar.

Figure 7. Calibration results for the lift coefficient at the first radial section.

In what follows, we will focus on the case where 200 measurement points are used for the calibration. Figure 8 shows the resulting samples of the posterior distribution for all parameters. The ellipsoidal form of the two-dimensional scatterplots indicates that the different parameters are to good approximation uncorrelated. This is consistent with the outcome of the sensitivity analysis. Note that, if the original thickness distribution would have been used, a strong correlation between the lift coefficients at section 3 and 4 would show up.

A summary of the posterior marginals displayed in Figure 8 is compiled in Table 2 in terms of the MAP and the standard deviation. Based on the MAP values, an example of a calibrated C_l polar, compared with the reference C_l polar, is shown in Figure 7b. The MAP values of the lift coefficients all lie around -0.2, meaning that the original lift coefficients need to be corrected by about 20% in order to match the experimental results (we will comment on this relatively large change in the next

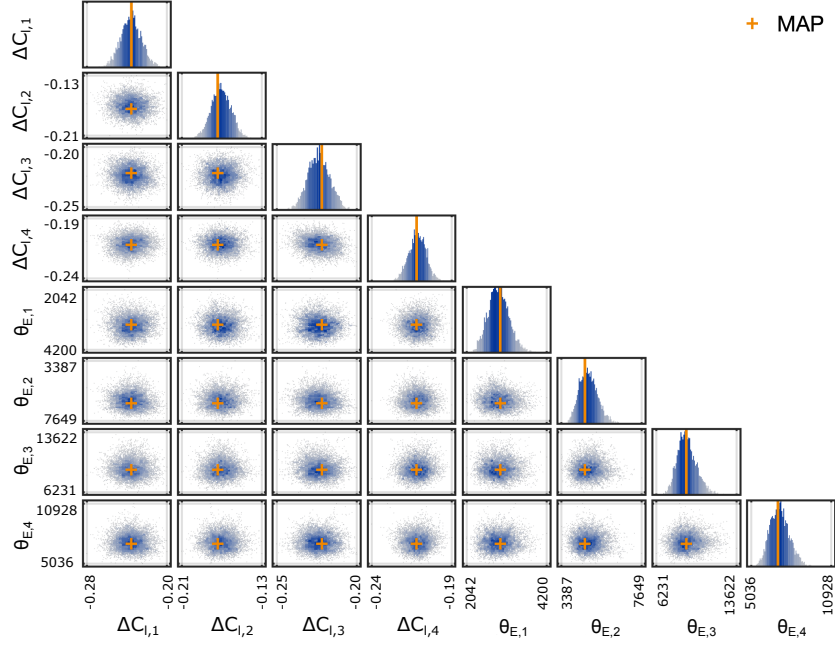



Figure 8. Samples of the posterior distribution of model parameters $\theta_M = (\Delta C_{l,1} \ldots \Delta C_{l,4})$ and discrepancy parameters $\theta_E = (\theta_{E,1} \ldots \theta_{E,4})$ for the DanAero test case with 200 measurements. The orange crosses correspond to the MAP estimates.

Table 2. Summary of prior and posterior distribution for DanAero calibration with 200 measurements. ΔC_l is dimensionless, θ_E has dimensions N^2/m^2.

Parameter	Prior	Posterior - MAP	Posterior - Standard deviation
$\Delta C_{l,1}$	$\mathcal{N}(0, 0.125)$	-0.23	$1.1 \cdot 10^{-2}$
$\Delta C_{l,2}$	$\mathcal{N}(0, 0.125)$	-0.17	$1.1 \cdot 10^{-2}$
$\Delta C_{l,3}$	$\mathcal{N}(0, 0.125)$	-0.22	$7.1 \cdot 10^{-3}$
$\Delta C_{l,4}$	$\mathcal{N}(0, 0.125)$	-0.21	$5.6 \cdot 10^{-3}$
$\theta_{E,1}$	$\mathcal{U}(0, 50000)$	$2.9 \cdot 10^3$	$2.9 \cdot 10^2$
$\theta_{E,2}$	$\mathcal{U}(0, 50000)$	$4.6 \cdot 10^3$	$5.1 \cdot 10^2$
$\theta_{E,3}$	$\mathcal{U}(0, 50000)$	$9.0 \cdot 10^3$	$9.4 \cdot 10^2$
$\theta_{E,4}$	$\mathcal{U}(0, 50000)$	$7.0 \cdot 10^3$	$7.3 \cdot 10^2$

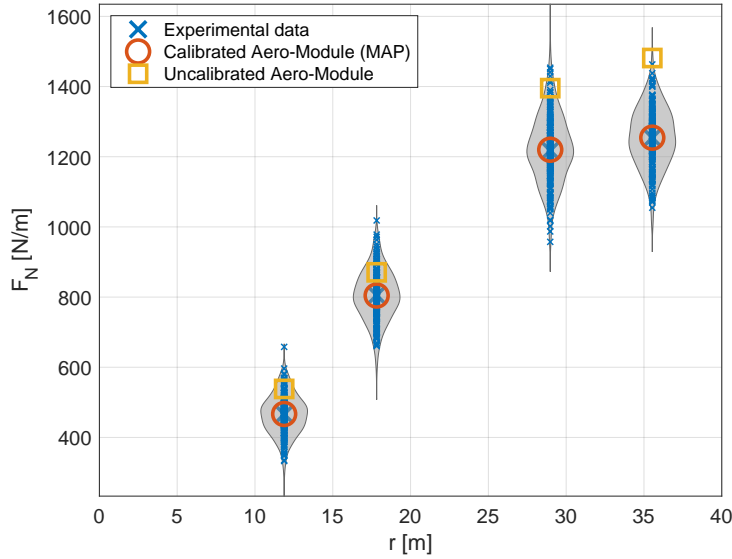



Figure 9. Comparison of uncalibrated and calibrated Aero-Module predictions of sectional normal forces with the DanAero measurements. The grey shaded areas indicate the posterior predictive distributions. The large blue crosses indicate the mean of the measurement data (at each section), the small crosses indicate the 200 measurements (at each section). The Aero-Module predictions are obtained from the surrogate model and are essentially indistinguishable from the full Aero-Module results, which have therefore been omitted for the sake of clarity.

paragraph). Table 2 also lists the standard deviation associated to the ΔC_l parameters, which is in all cases small, confirming the observation of Figure 7a that the posterior is sharply peaked when sufficient measurement points are included.

Given the samples of the posterior distribution, the posterior predictive distribution is computed following equation (23) and plotted along with the measurement data, the uncalibrated model results, and the model evaluated at the MAP in Figure 9.

435 Clearly, the calibrated Aero-Module (MAP) is overlapping with the mean of the experimental data. Furthermore, the posterior predictive (which expresses the probability of observing new data *given* the calibrated lift polars), centers nicely around the MAP and encapsulates the experimental data well. The results of the uncalibrated Aero-Module at the third and fourth radial section are very unlikely given the calibrated lift polars.

We note that in order to obtain the posterior predictive of other possible quantities of interest not considered in this work 440 (such as the power output or the blade bending moment), one would preferably add these quantities to the model output list *before* the surrogate model is being trained, so that the posterior predictive can be efficiently evaluated without requiring full model runs. Alternatively, one could use the full Aero-Module with the calibrated parameters and use these to determine the posterior predictive for the power, but that would be computationally very expensive.

It is important to note that using the obtained lift polars and hyperparameters in a predictive setting for a different set of 445 operating conditions will require careful consideration. Firstly, using a single operating condition for calibrating the lift polars,

as is currently the case, makes their validity to other operating conditions limited. Secondly, the discrepancy between model and measurement data (consisting of both model and measurement errors), E in equation (18), has been fully accounted for by calibrating the lift coefficient. It is highly likely that E depends also on other factors, such as the simplifications (missing physics) present in BEM theory, the unsteadiness of the atmospheric conditions, the uncertainty in the measurements, etc. This is perhaps the cause that relatively large changes (around 20%) in the sectional lift coefficients are needed to achieve a match between the Aero-Module and the experimental data. Lastly, the values obtained for the hyperparameters $\theta_{E,1}$ - $\theta_{E,4}$ are very much dependent on the aerodynamic model and the data used. These values currently include both the measurement noise and the model inadequacy, which are not expected to be the same for a different set of measurements, a different operating condition, or a different aerodynamic model.

5.2 Yaw model calibration with New MEXICO data

In the previous section we showed as a proof of concept how the combination of surrogate modelling, sensitivity analysis and Bayesian inference can be used to calibrate parameters of the Aero-Module. The test case was relatively simple in the sense that only a single operating condition was used, and because the relation between lift coefficient and normal force (equation (25)) is linear, the sensitivity analysis and calibration results were quite straightforward. In this section we move to a more advanced test case, in which the parameters of a yaw model are calibrated based on normal force measurements.

5.2.1 Sensitivity analysis

As mentioned in section 2.2, experiments and corresponding simulations were carried out for three different operating conditions; the values are shown in Table 1. The parameters to be calibrated are the yaw model parameters given by equation (7) and repeated here for convenience:

$$\theta_M = (\mathrm{AM}_{11}, \mathrm{AM}_{12}, \mathrm{AM}_{13}, \mathrm{AM}_{14}, \mathrm{AM}_{15}). \tag{27}$$

The nominal (uncalibrated) values for these parameters are listed in table C1 in Appendix C. A normal distribution is assumed for each amplitude coefficient, based on consultation with the developer of the yaw model (Schepers, 2012):

$$\mathrm{AM} \sim \mathcal{N}(\mu, \sigma), \tag{28}$$

where μ equals the nominal value provided in table C1, and σ is taken equal to 0.1. This value is such that the spread in the experimental data can be captured by the AeroModule (as will be needed for calibration) and also makes sure that the induced velocity is likely to remain positive.

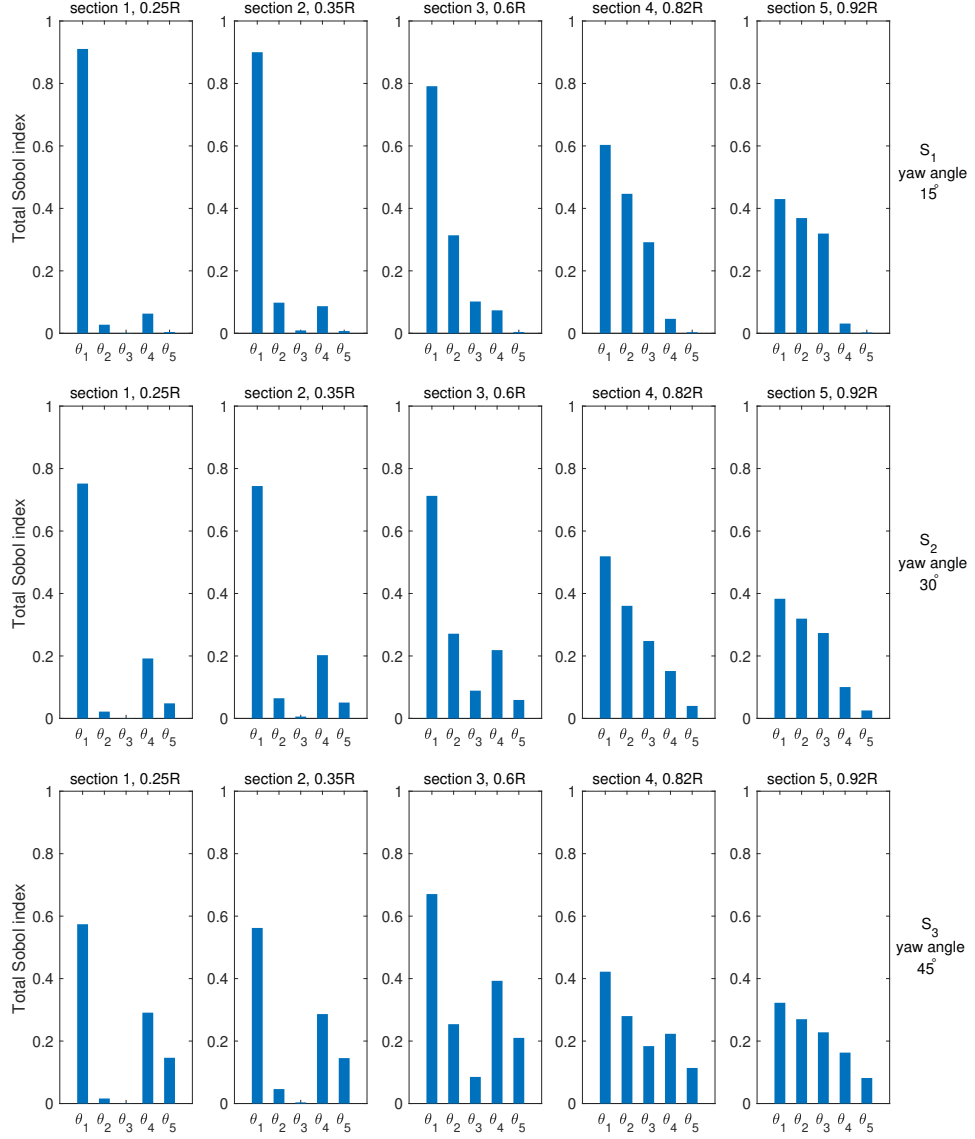



Figure 10. Sensitivity analysis for three different operating conditions: $\mathcal{S}_1$ (top), $\mathcal{S}_2$ (middle), $\mathcal{S}_3$ (bottom). Here $\theta_M = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5) = (\mathrm{AM}_{11}, \mathrm{AM}_{12}, \mathrm{AM}_{13}, \mathrm{AM}_{14}, \mathrm{AM}_{15})$.

Similar to the previous case, a PCE is set-up by drawing random samples from the parameter vector. In appendix C the convergence of the LOO error is assessed, and it is shown that by taking $N = 256$ samples (for each scenario) the LOO error is at most on the order of a few percent, which is sufficient to obtain accurate Sobol' indices (and to perform calibration). The total order Sobol' indices S^T following equation (17) are computed for the five dimensional parameter vector θ_M. In contrast to the DanAero case, (where the normal force at section i depended exclusively on the lift coefficient at section i), in this case the normal force at a certain section depends on the value of all model parameters, so that many more simulations are required to obtain an accurate surrogate model.

The resulting plots for the total Sobol' indices are shown in Figure 10 for all three operating conditions. It is clear that parameter AM_{11} is especially important at the inner part of the blade, whereas parameters AM_{12} and AM_{13} become increasingly important for the outboard sections. AM_{14} and AM_{15} have little dependency on r, but instead increase in importance when the yaw angle is increased. This behavior is consistent with expression (C1). Overall, it can be concluded that all parameters are significantly influencing the normal force behavior (under the assumed distributions).

5.2.2 Calibration

The sensitivity analysis did not identify clear non-influential parameters, so all five yaw model parameters will be included in the calibration process. The experimental data for the calibration consists of the normal force measurements obtained from the New MEXICO experiment. However, the normal force measurements at section 3 were not included in the calibration process, for the following reasons. Firstly, in Schepers et al. (2018) it was shown that the normal force amplitude measured at section 3 (for case 2.1) appeared to be much lower than predicted by both BEM and CFD codes. Secondly, it turned out that under the assumed range for the AM-parameters, it was not possible to get amplitudes as small as reported in the measurement data (and this could not be fixed by increasing the range). Perhaps this could be fixed by including the other parameters of the yaw model, but it is also likely that the measurement data is off at this point. Therefore, the normal force at section 3 has been removed from the surrogate model constructed in section 5.2.1 for the purpose of calibration. The resulting PCE-based surrogate model for the quantity of interest, $\mathcal{Q}^{PC}(\theta_M)$, is based on 768 samples (256 for each scenario) and used in place of the Aero-Module throughout.

The prior on the model parameters is taken the same as in equation (28). Given the limited amount of measurement data, the discrepancy parameters are not calibrated in this test case, but are chosen to be fixed (with the same value for each scenario and each radial section) at $\theta_E = \sigma^2$, where σ is taken equal to 3 as a rough estimate based on the uncertainty bands reported in Schepers et al. (2018). Like in the previous test case, the AIES MCMC algorithm with 10^3 steps and 10^2 parallel chains is deployed. The posterior samples and the marginal distributions are shown in Figure 11, and corresponding statistics are given in table 3 (for MCMC trace plot examples, we refer to Figure C2 in appendix C). One can observe that the posterior distributions are still Gaussian-like, but with a shifted mean and a smaller standard deviation than the prior distribution. The largest shift (in absolute sense) is incurred for parameters AM_{14} and AM_{15}; the smallest shift happens for AM_{13}, which is hardly changed when compared to the prior. In contrast to the DanAero test case, where the posterior was very much dominated by the data, the posterior for the New MEXICO case is still close to the prior, because fewer data points are used. The posterior samples

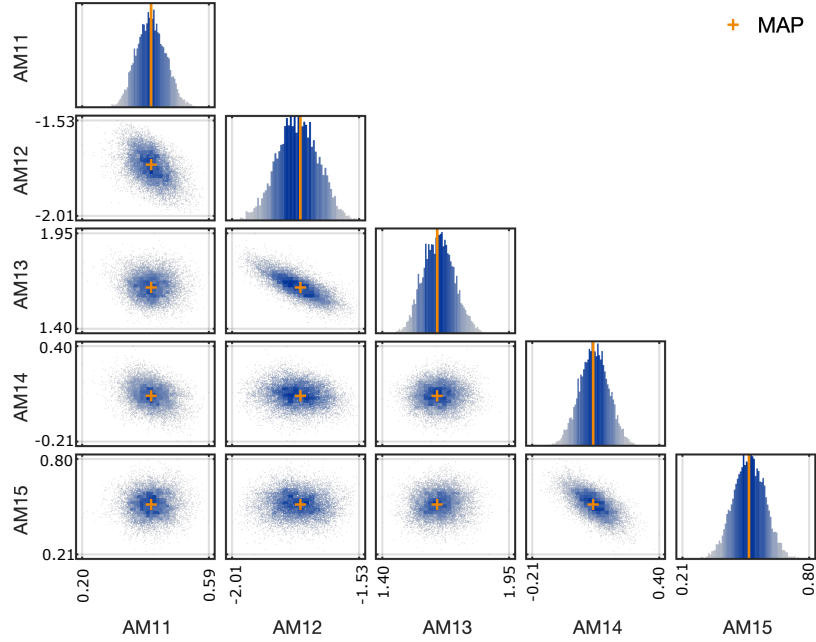



Figure 11. Samples of the posterior distribution of yaw model parameters for the New MEXICO test case. The orange crosses correspond to the MAP estimates.

indicate a clear correlation between parameters AM_{12} and AM_{13}, and between AM_{14} and AM_{15}. This result is consistent with the yaw model expression (equation (C1)), since AM_{12} and AM_{13} both relate to the relative radius, whereas AM_{14} and AM_{15} both relate to the yaw angle.

The normal force amplitudes that are obtained with the calibrated model parameters (based on the MAP) are shown for all operating conditions in Figure 12, together with the measurement data and the uncalibrated model output. Overall, we observe that the calibration of the parameters has led to much improved model predictions. This is especially evident for the outboard sections for operating conditions $\mathcal{S}_2$ and $\mathcal{S}_3$. For other scenarios and/or radial sections also an improvement is generally observed, except for a few points, where the match is slightly worse (e.g. $\mathcal{S}_2$, section 2). Since the likelihood function in Bayes (equation (21) with covariance matrix chosen as $\Sigma = \sigma^2 I$) weighs the discrepancy in the different model outputs equally, it is not surprising that at some points the discrepancy can increase slightly, while at other points it is significantly reduced: on *average* the model fits the data much better.

Table 3. Summary of prior and posterior distribution for New MEXICO yaw model calibration.

Parameter	Prior	Posterior - MAP	Posterior - Standard deviation
AM_{11}	$\mathcal{N}(0.445, 0.1)$	0.41	0.046
AM_{12}	$\mathcal{N}(-1.78, 0.1)$	-1.75	0.078
AM_{13}	$\mathcal{N}(1.63, 0.1)$	1.64	0.068
AM_{14}	$\mathcal{N}(-0.0543, 0.1)$	0.085	0.071
AM_{15}	$\mathcal{N}(0.367, 0.1)$	0.52	0.075

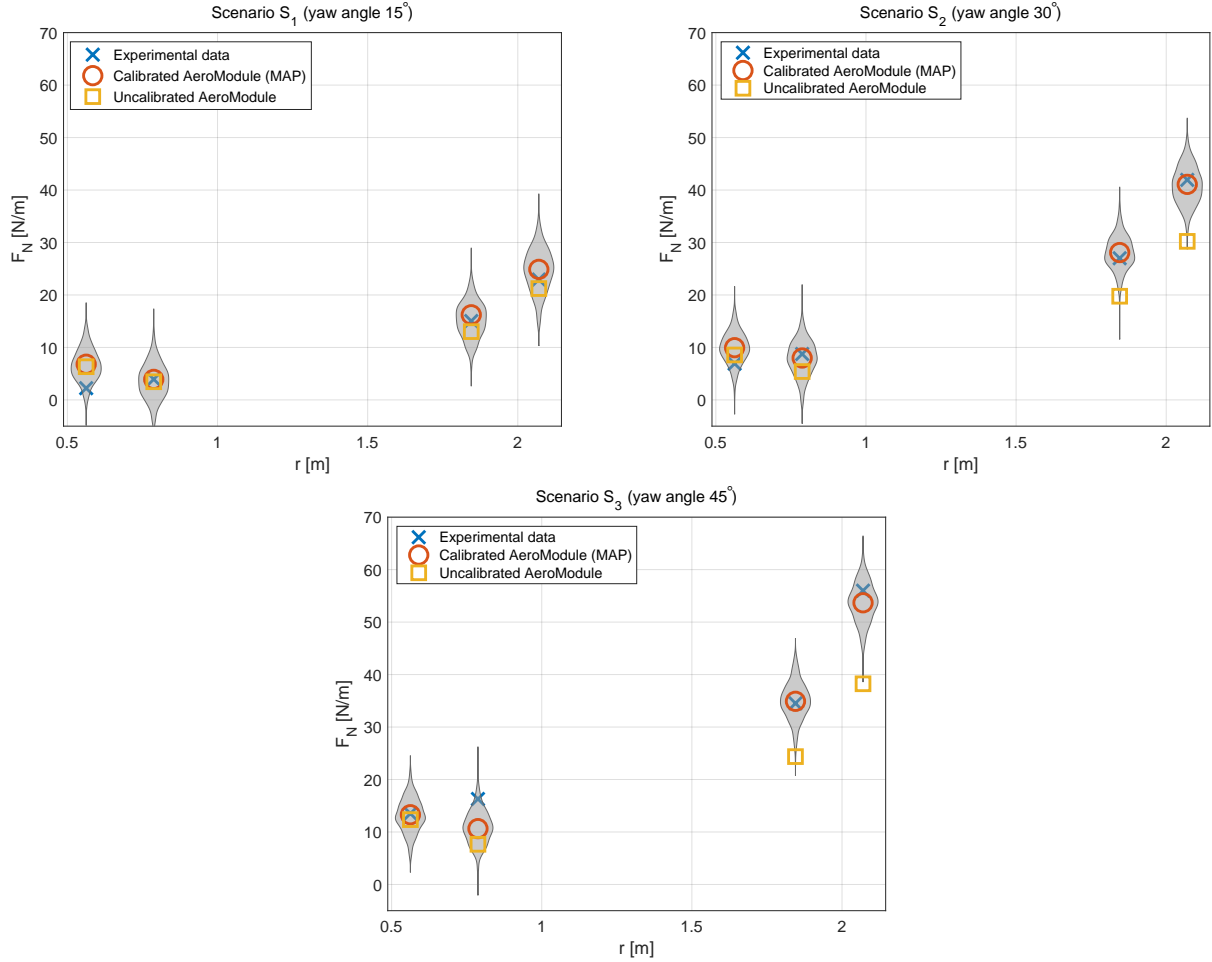



Figure 12. Comparison of uncalibrated and calibrated AeroModule predictions with the New MEXICO measurements in terms of the normal force amplitude. The grey areas (violin plots) are constructed from 500 draws of the posterior predictive distribution. The Aero-Module predictions are obtained from the surrogate model and are essentially indistinguishable from the full Aero-Module results, which have therefore been omitted for the sake of clarity.

6 Conclusions

In this article we have proposed a computationally efficient framework to calibrate model parameters in aerodynamic wind turbine models. The three main ingredients that we use are: (i) a (polynomial) surrogate model that approximates the wind turbine model, (ii) a sensitivity analysis to determine the most influential parameters, and (iii) Bayesian inference to calibrate parameters in a probabilistic setting. The Bayesian inference step, which is typically computationally very expensive to solve, is made computationally affordable through the use of the surrogate model. Evaluating 10^5 MCMC iterations takes less than an hour with the surrogate model, whereas it would take weeks with the full Aero-Module model (when running on a desktop computer). The polynomial nature of the surrogate model furthermore allows quick evaluation of the Sobol' indices in the sensitivity analysis. The entire framework, known as UQ4Wind, is built around the UQLab software and tested on TNO's aerodynamic code Aero-Module.

Two realistic calibration studies have been performed with our proposed UQ4Wind framework in this paper. In the first test case, we have used part of the DanAero experimental dataset to show how airfoil polars can be calibrated using normal force measurements. The sensitivity analysis clearly indicated that out of the lift, drag and moment coefficients, the lift coefficient is most influential. After calibrating the lift coefficient values at the four radial sections, an excellent match with the experimental data was observed.

In the second test case, we have used part of the New MEXICO experimental dataset to calibrate five parameters of the yaw model that is used in the Aero-Module to estimate the induced velocity in yawed conditions. In order to handle the time dependence of both measurements and code output, we used a Fourier transform and considered the amplitude of the most dominant Fourier mode as the quantity of interest for the calibration. The calibrated model leads to much improved model predictions, especially regarding the normal force amplitude at the outboard sections of the blade under significant yaw misalignment.

In both cases, the result of the Bayesian approach consists of *distributions* of the calibrated model parameters (the posterior distribution). The posterior distribution allows us to make predictions under uncertainty, for example by computing the posterior predictive distribution, from which probabilistic statements can be deduced. At the same time, existing knowledge on the model parameters (e.g. expert knowledge) can be included via the prior distribution, and any relation (not necessarily Gaussian) between model and measurement data can be specified by choosing a likelihood function. These aspects form the true strength of the Bayesian approach. However, it should be realized that in the practical setting of calibrating an aerodynamic wind turbine model, it is not always clear how representative choices for the prior distribution or the likelihood are to be made. For the cases investigated in this paper, we have relied on expert knowledge and inspection of the measurement data. We acknowledge that this process should be carefully performed when considering different experimental datasets and/or model parameters. Similarly, the choice of distribution used in the sensitivity analysis (and in particular the corresponding variance) can determine to a large extent the Sobol' indices and has to be performed with care.

Another aspect that requires careful attention is the selection of proper datasets. Initially, the plan was to use more datasets from both DanAero and New MEXICO in the calibration runs, but it turned out that many were not directly useful, for example

because of non-constant operating conditions (DanAero) or because the normal forces obtained from pressure distributions were not considered accurate enough in yawed conditions (New MEXICO). As alternative, it is also possible to use models with a higher physical fidelity (e.g. free vortex wake models, which perform well in yawed conditions) to generate data for the calibration.

555   A last aspect for future consideration is that of the steady nature of the DanAero test case, for which the airfoil polars could be calibrated without taking into account the effect of the dynamic stall model. In more realistic settings (e.g. turbulent inflow), one would have to calibrate both the airfoil polars and the dynamic stall model simultaneously. In the current case, time-dependent data was not available for further cross-validation, and the accuracy of the calibrated polars in dynamic flow conditions remains uncertain. We note that if such data would be available, one should realize that the simultaneous calibration of both dynamic

560   stall model parameters and airfoil polars would constitute a high dimensional problem that might be computationally very expensive. Our approach, in which such effects are separated by using a time-dependent and a time-independent case, is effectively a way to reduce the high dimensionality of such a calibration problem.

Overall, we believe that the combination of surrogate modelling, sensitivity analysis and Bayesian inference provides a powerful approach towards model calibration. Calibrated models with a quantified level of uncertainty have many applications

565   in the wind energy industry beyond the aerodynamic models considered in this study, such as calibrating a dynamic wind farm control model (this is part of our ongoing work). Another topic within wind energy that could benefit from the UQ4WIND framework could be the calibration of low-order acoustic models using empirical correction factors for wind turbine noise estimation. Furthermore, calibration of engineering wake models, which typically contain several uncertain model parameters (such as wake expansion coefficients), would benefit from calibration using high-fidelity models such as CFD results.

570   *Code availability.* The UQ4Wind software is available from https://github.com/bsanderse/uq4wind. UQ4Wind is built around the UQLab software package (Marelli and Sudret, 2014), version 1.4, available from https://www.uqlab.com/. UQ4Wind does not include a license for the Aero-Module, which has to be obtained separately from TNO).

Appendix A: Surrogate model details and UQLab settings

The surrogate model is built using LARS. The sampling scheme is Latin Hypercube Sampling (LHS) in all cases. Adaptive,

575   sparse LARS is used, with possible polynomial degrees from 1 to 4 and truncation parameter 0.75 for DanAero, and polynomial degrees from 1 to 10 and truncation parameter range from 0.5 to 1.5 for New MEXICO.

For the sensitivity analysis, the main UQLab commands are:

- `uq_createInput` (definition of model parameters),

- `uq_createModel` (once for definition of model and quantities of interest, and once for setting up the PCE surrogate

580   model),

– `uq_createAnalysis` (determining Sobol indices based on surrogate model).

For the Bayesian calibration, the same sequence of commands is used, with the difference that the `uq_createAnalysis` command then takes as input the options for the Bayesian calibration (prior, likelihood, experimental data, MCMC settings).

Appendix B: Details lift polar calibration

 ### B1   LOO convergence

The surrogate model should be accurately approximating the full Aero-Module in order to use it for sensitivity analysis and Bayesian inference. Figure B1a shows that the LOO error of the normal force at each airfoil section rapidly converges upon increasing the number of samples. The convergence of the LOO also becomes more regular when adapting the thickness distribution as will be described in the next section B2, which can be observed in Figure B1b. This is because the surrogate model at a certain section becomes almost independent of the parameters (lift coefficients) at other sections, making it easier to train. Note that the reported datapoints are obtained by averaging over 5 simulation runs (so for $N = 16$ we perform 5×16 simulations) in order to smooth out the randomness introduced by the LHS sampling method. For the results in sections 5.1 we use the surrogate model with $N = 32$.

(a) Original thickness distribution.   (b) Adapted thickness distribution.

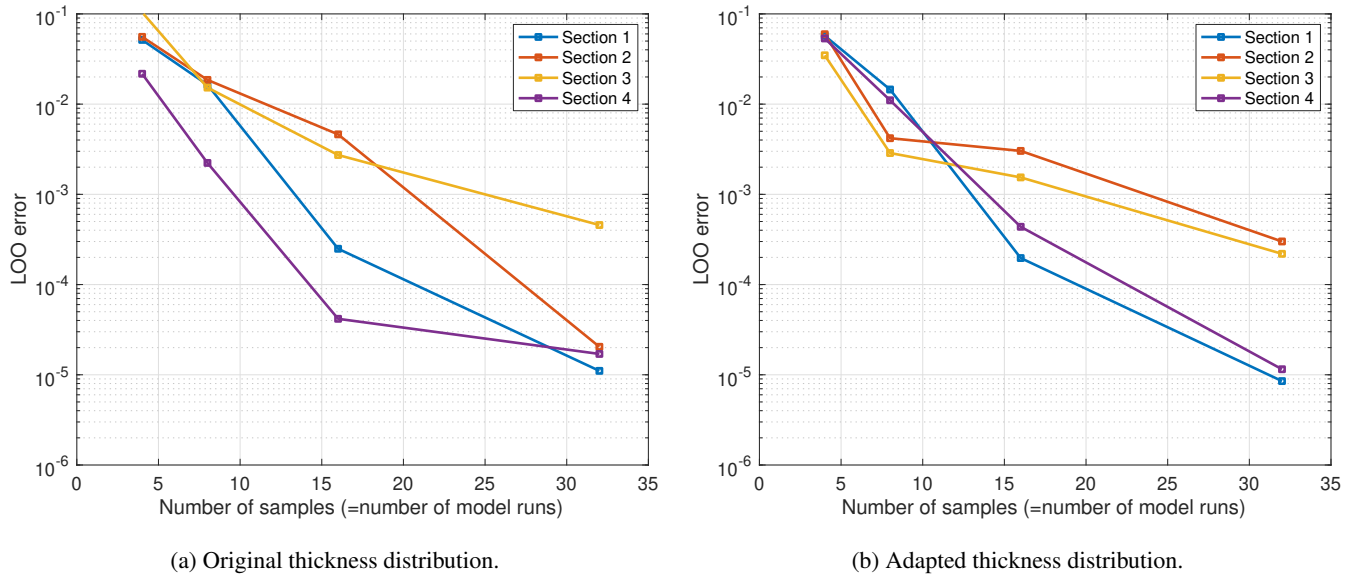



Figure B1. Convergence of LOO error as function of number of samples for DanAero test case with 12 parameters. Each datapoint corresponds to the average over 5 runs.

B2 Thickness adaptation

595   When using the original input files to perform the DanAero sensitivity study, it turned out that the normal force at section 3 depended on the lift coefficient at both sections 3 and 4 (see Figure B2). This peculiarity is caused by an inconsistency between the provided ('planform') thickness distribution of the blade and the provided thickness of the four airfoil sections, as shown in Figure B3. The lift coefficient at any radial position along the blade is determined by checking the local thickness in the planform graph, and then interpolating the lift coefficient from nearby airfoil sections, based on the relative thickness. For

600   example, at section 3 ($r = 29$m), the planform thickness is around 0.189. This value lies in between the values of section 3 ($t/c = 0.197$) and section 4 ($t/c = 0.187$), but is much closer to section 4 than section 3. This explains the large effect of $\Delta C_{l,4}$ on the force at section 3.

     After consultation with the DanAero experts[1], the thickness of the airfoil sections was changed to match the planform data (see Figure B3). Figure 6 shows the results of the sensitivity analysis, indicating that with the adapted sectional thicknesses,

605   we correctly obtain the expected dependency of the sectional normal force on the corresponding sectional lift coefficient. Thus, apart from identifying influential parameters, the sensitivity analysis step in our framework can also be used to correct inconsistencies in the model formulation.

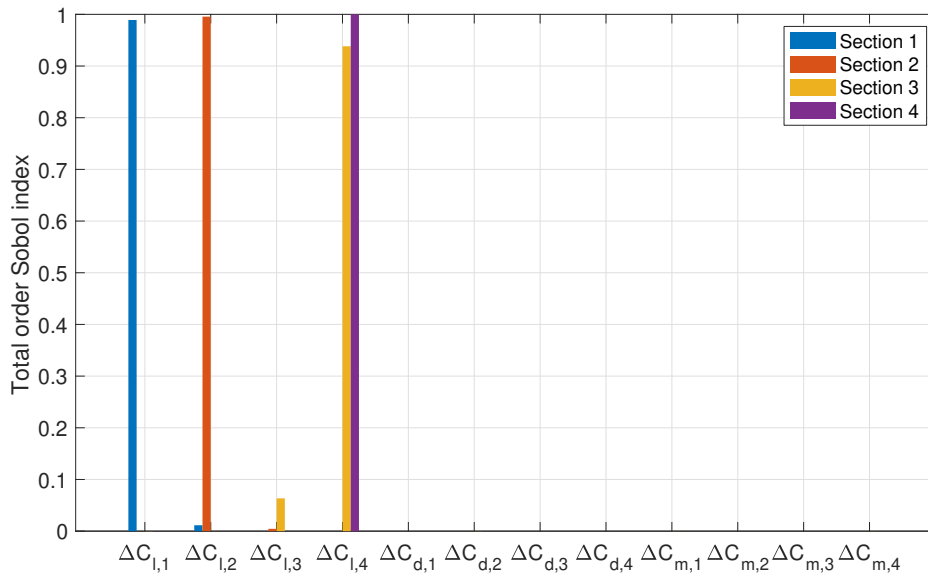



Figure B2. Sensitivity of sectional normal force with respect to perturbations in airfoil polars with original thickness.

---

[1]We thank Georg Pirrung from DTU for providing us the adapted thickness values for the airfoil sections.

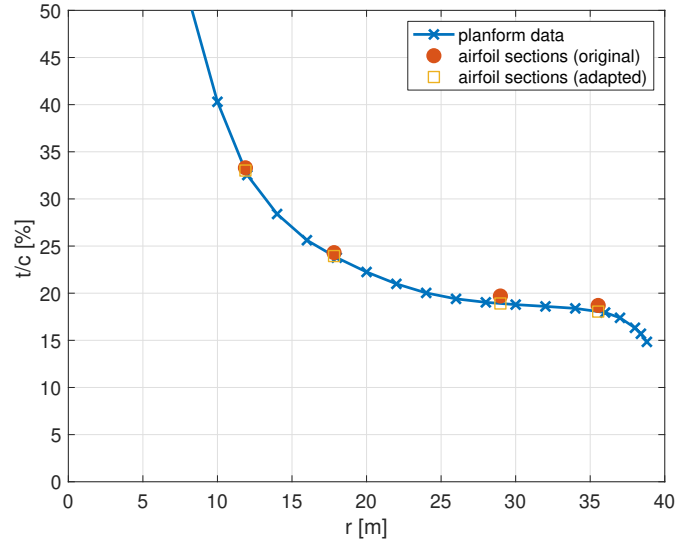



Figure B3. Thickness distribution of DanAero blade according to the planform data and the original and adapted airfoil sectional data. The discrepancy explains the observed sensitivity of the force at section 3 towards the lift coefficient at section 4.

B3 Calibration

Figure B4 shows two examples of the convergence of the MCMC chains for the parameter $\Delta C_{l,1}$ and hyperparameter $\theta_{E,1}$, where 200 datapoints were used for the calibration. The plot shows 100 chains that have been ran for 1000 steps using the AIES algorithm. The trace plots for the other parameters show very similar behavior.

(a) $\Delta C_{l,1}$.   (b) $\theta_{E,1}$.

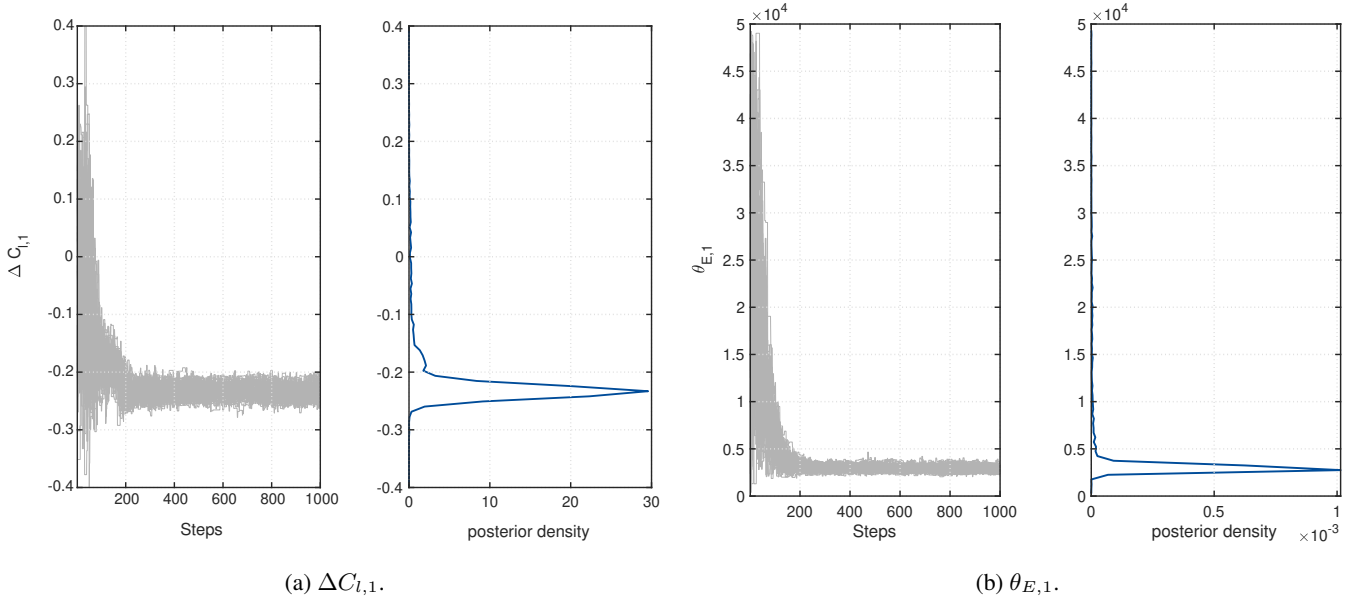



Figure B4. Examples of MCMC chains and resulting posterior for model parameter $\Delta C_{l,1}$ and hyperparameter $\theta_{E,1}$, using 200 measurement points. For the results reported in section 5.1, the first 50% of the chains is discarded (so-called burn-in).

Appendix C: Details yaw model calibration

The yaw model that is calibrated concerns an expression for the induced velocity (Schepers, 2012). The general form is

$$u_i(\bar{r}, \beta, \phi_r, \theta_M) = u_{i,0}\left(1 - A_1(\bar{r}, \beta, \theta_M)\cos(\phi_r - \psi_1(\bar{r}, \beta, \theta_M)) - A_2(\bar{r}, \beta, \theta_M)\cos(2\phi_r - \psi_2(\bar{r}, \beta, \theta_M))\right). \tag{C1}$$

615   Here $\bar{r}$ is the relative radius ($\bar{r} = r/R$), ϕ_r the azimuth angle, and β the yaw angle. In this study, we focus on $\theta_M = (\mathrm{AM}_{11}, \ldots, \mathrm{AM}_{15})$, which appears in the expression for A_1:

$$A_1(\bar{r}, \beta, \theta_M) = \mathrm{AM}_{11} + \mathrm{AM}_{12}\bar{r} + \mathrm{AM}_{13}\bar{r}^2 + \mathrm{AM}_{14}\sin|\beta| + \mathrm{AM}_{15}\sin^2\beta. \tag{C2}$$

Note that each parameter is multiplied by a factor that is between 0 and 1, since $0 \le \bar{r} \le 1$ and $0 \le \beta \le 90°$.

The nominal values for the yaw model parameters are given in table C1 and taken from (Schepers, 2012).

Table C1. Summary of nominal value for yaw model parameters: amplitude coefficients (dimensionless).

AM_{11}	AM_{12}	AM_{13}	AM_{14}	AM_{15}
0.445	-1.78	1.63	-0.0543	0.367

620   The convergence of the LOO error for the New MEXICO case, for all three operating conditions, is shown in Figure C1. Examples of MCMC trace plots for two selected parameters (AM_{11} and AM_{14}) are shown in Figure C2. The trace plots for the other parameters are similar.

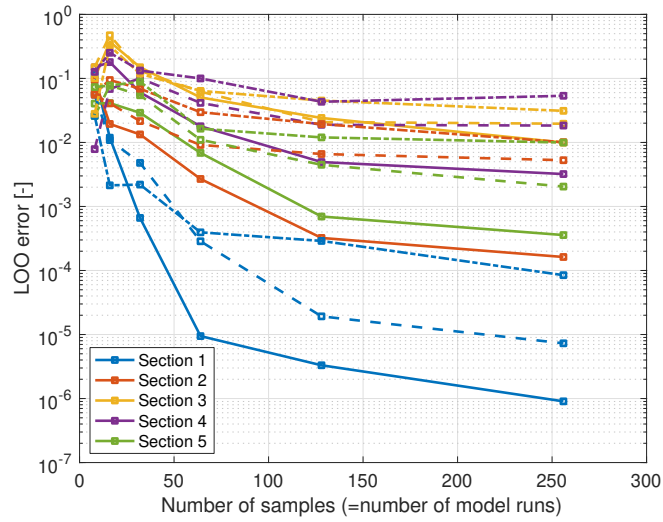



Figure C1. Surrogate model convergence for $\mathcal{S}_1$ (dash-dot), $\mathcal{S}_2$ (dashed) and $\mathcal{S}_3$ (solid). Each datapoint corresponds to the average over 5 runs.

(a) AM_{11}.

(b) AM_{14}.

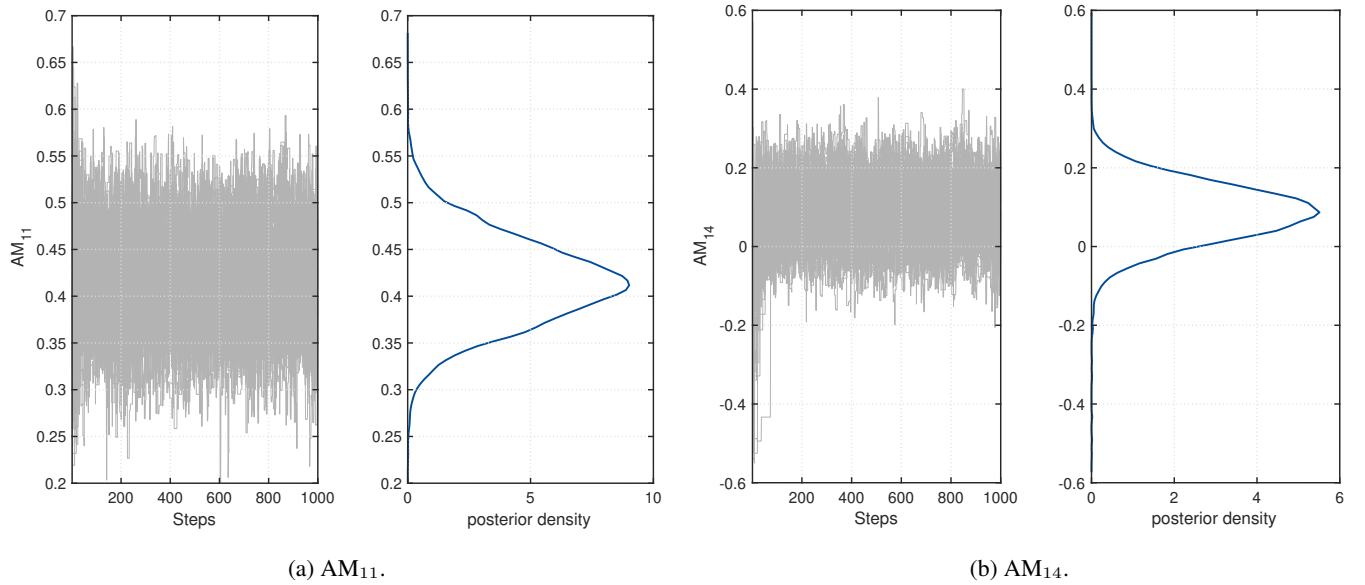



Figure C2. Examples of MCMC chains and resulting posterior for parameters AM_{11} and AM_{14}. For the results reported in section 5.2, the first 50% of the chains is discarded (so-called burn-in).

Acknowledgements. We would like to thank our colleagues from the WindTrue project for all the stimulating discussions. Furthermore, we want to thank the DanAero and New MEXICO consortia for making their datasets available for this study. Lastly, we would also like to thank

625 Laurent van den Bos, whose PhD thesis initiated this work and whose comments have been highly appreciated. The research was sponsored by the Topsector Energy Subsidy from the Dutch Ministry of Economic Affairs and Climate.

Author contributions. BS (Conceptualization, Methodology, Software, Writing - Original Draft, Funding Acquisition); VD (Software, Writing - Original Draft); KB (Resources, Validation, Review & Editing); GS (Validation, Review & Editing, Funding Acquisition , Project Administration).

630 *Competing interests.* The authors declare that they have no conflict of interest.

References

Abdallah, I., Natarajan, A., and Sørensen, J. D.: Impact of uncertainty in airfoil characteristics on wind turbine extreme loads, Renewable Energy, 75, 283–300, 2015.

Andrieu, C., De Freitas, N., Doucet, A., and Jordan, M. I.: An introduction to MCMC for machine learning, Machine learning, 50, 5–43, 2003.

Bak, C., Madsen, H. A., Gaunaa, M., Paulsen, U. S., Fuglsang, P., Romblad, J., Olesen, N. A., Enevoldsen, P., Laursen, J., and Jensen, L.: DAN-AERO MW: Comparisons of airfoil characteristics for two airfoils tested in three different wind tunnels, Torque 2010: The science of making torque from wind. EWEA, 2010, 59–70, 2010.

Bayes, T.: LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S, Philosophical transactions of the Royal Society of London, pp. 370–418, 1763.

Blatman, G.: Adaptive sparse polynomial chaos expansions foruncertainty propagation and sensitivity analysis, Ph.D. thesis, Universite Blaise Pascal, Clermont-Ferrand, France, 2009.

Boorsma, K. and Grasso, F.: ECN Aero-Module: User's Manual, Tech. rep., Energy research Centre of the Netherlands, 2015.

Boorsma, K. and Schepers, J.: New MEXICO experiment - Preliminary Overview with Initial Validation, Tech. Rep. ECN-E–14-048, Energy research Centre of the Netherlands, 2014.

Boorsma, K. and Schepers, J.: Rotor experiments in controlled conditions continued: New Mexico, Journal of Physics: Conference Series, 753, 022 004, https://doi.org/10.1088/1742-6596/753/2/022004, 2016.

Boorsma, K., Grasso, F., and Holierhoek, J.: Enhanced approach for simulation of rotor aerodynamic loads, Tech. Rep. ECN-M–12-003, Energy research Centre of the Netherlands, 2012.

Bottasso, C. L., Cacciola, S., and Iriarte, X.: Calibration of wind turbine lifting line models from rotor loads, Journal of Wind Engineering and Industrial Aerodynamics, 124, 29–45, 2014.

Buhl, M. and Manjock, A.: A Comparison of Wind Turbine Aeroelastic Codes Used for Certification, in: 44th AIAA Aerospace Sciences Meeting and Exhibit, American Institute of Aeronautics and Astronautics, https://doi.org/10.2514/6.2006-786, 2006.

Dutta, S., Ghosh, S., and Inamdar, M. M.: Optimisation of tensile membrane structures under uncertain wind loads using PCE and kriging based metamodels, Structural and Multidisciplinary Optimization, 57, 1149–1161, 2018.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., et al.: Least angle regression, The Annals of statistics, 32, 407–499, 2004.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B.: Bayesian data analysis, CRC press, 2013.

Goodman, J. and Weare, J.: Ensemble samplers with affine invariance, Communications in applied mathematics and computational science, 5, 65–80, 2010.

Guo, X., Dias, D., Carvajal, C., Peyras, L., and Breul, P.: Reliability analysis of embankment dam sliding stability using the sparse polynomial chaos expansion, Engineering Structures, 174, 295–307, 2018.

Kennedy, M. C. and O'Hagan, A.: Bayesian calibration of computer models, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63, 425–464, 2001.

Kumar, P., Sanderse, B., Boorsma, K., and Caboni, M.: Global sensitivity analysis of model uncertainty in aeroelastic wind turbine models, Journal of Physics: Conference Series, 1618, 042 034, 2020.

Laloy, E., Rogiers, B., Vrugt, J. A., Mallants, D., and Jacques, D.: Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov chain Monte Carlo simulation and polynomial chaos expansion, Water Resources Research, 49, 2664–2682, 2013.

Leishman, J. G.: Challenges in modelling the unsteady aerodynamics of wind turbines, Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology, 5, 85–132, 2002.

670 Madsen, H., Bak, C., Schmidt Paulsen, U., Gaunaa, M., Fuglsang, P., Romblad, J., Olesen, N., Enevoldsen, P., Laursen, J., and Jensen, L.: The DanAero MW experiments: final report, Tech. Rep. Risø-R-1726(EN), Danmarks Tekniske Universitet & Risø National laboratory, 2010.

Madsen, H. A., Sørensen, N. N., Bak, C., Troldborg, N., and Pirrung, G.: Measured aerodynamic forces on a full scale 2MW turbine in comparison with EllipSys3D and HAWC2 simulations, Journal of Physics: Conference Series, 1037, 022 011, https://doi.org/10.1088/1742-675 6596/1037/2/022011, 2018.

Marelli, S. and Sudret, B.: UQLab: A framework for uncertainty quantification in Matlab, in: Vulnerability, uncertainty, and risk: quantification, mitigation, and management, pp. 2554–2563, American Society of Civil Engineers, 2014.

Marelli, S. and Sudret, B.: UQLab user manual – Polynomial chaos expansions, Tech. rep., Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, report # UQLab-V1.3-104, 2019.

680 Matthäus, D., Bortolotti, P., Loganathan, J., and Bottasso, C. L.: Propagation of Uncertainties Through Wind Turbine Models for Robust Design Optimization, in: 35th Wind Energy Symposium, AIAA SciTech Forum, 9-13 January 2017, Grapevine, Texas, American Institute of Aeronautics and Astronautics, https://doi.org/10.2514/6.2017-1849, 2017.

Murcia, J. P.: Uncertainty Quantification in Wind Farm Flow Models, Ph.D. thesis, Technical Unversity of Denmark, 2016.

Murcia, J. P., Réthoré, P.-E., Dimitrov, N., Natarajan, A., Sørensen, J. D., Graf, P., and Kim, T.: Uncertainty propagation through an aeroelastic 685 wind turbine model using polynomial surrogates, Renewable Energy, 119, 910–922, 2018.

Oakley, J. E. and O'Hagan, A.: Probabilistic sensitivity analysis of complex models: a Bayesian approach, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 66, 751–769, 2004.

Özçakmak, Ö. S., Madsen, H. A., Sørensen, N. N., Sørensen, J. N., Fischer, A., and Bak, C.: Inflow Turbulence and Leading Edge Roughness Effects on Laminar-Turbulent Transition on NACA 63-418 Airfoil, Journal of Physics: Conference Series, 1037, 022 005, 690 https://doi.org/10.1088/1742-6596/1037/2/022005, 2018.

Papageorgiou, A. and Traub, J.: Beating Monte Carlo, Risk, 9, 63–65, 1996.

Schepers, J.: Engineering models in wind energy aerodynamics, Ph.D. thesis, Delft University of Technology, 2012.

Schepers, J., Lutz, T., Boorsma, K., Gomez-Iradi, S., Herraez, I., Oggiano, L., Rahimi, H., Schaffarczyk, P., Pirrung, G., Madsen, H., Shen, W., and Weihing, P.: Final report of IEA Wind Task 29 Mexnext (Phase 3), Tech. rep., Energy research Centre of the Netherlands, eCN-695 E-18-003, 2018.

Schepers, J., Boorsma, K., Madsen, H., Pirrung, G., Bangga, G., et al.: IEA Wind TCP Task 29, Phase IV: Detailed Aerodynamics of Wind Turbines, https://zenodo.org/record/4813068#.YMB2HpMzY61, https://doi.org/10.5281/zenodo.4813068, 2021.

Schöbi, R.: Surrogate models for uncertainty quantification in the context of imprecise probability modelling, IBK Bericht, 505, 2019.

Severini, T. A.: Likelihood methods in statistics, Oxford University Press, 2000.

700 Simms, D., Schreck, S., Hand, M., and Fingersh, L. J.: NREL unsteady aerodynamics experiment in the NASA-Ames wind tunnel: a comparison of predictions to measurements, Tech. rep., National Renewable Energy Lab., Golden, CO (US), 2001.

Smith, R. C.: Uncertainty quantification: theory, implementation, and applications, SIAM, 2013.

Sobol, I. M.: Sensitivity analysis for non-linear mathematical models, Mathematical modelling and computational experiment, 1, 407–414, 1993.

705 Soize, C. and Ghanem, R.: Physical systems with random uncertainties: chaos representations with arbitrary probability measure, SIAM Journal on Scientific Computing, 26, 395–410, 2004.

Sørensen, J. D. and Toft, H. S.: Probabilistic design of wind turbines, Energies, 3, 241–257, 2010.

Sudret, B.: Global sensitivity analysis using polynomial chaos expansions, Reliability engineering & system safety, 93, 964–979, 2008.

van Beek, M. T., Viré, A., and Andersen, S. J.: Sensitivity and Uncertainty of the FLORIS Model Applied on the Lillgrund Wind Farm,
710 Energies, 14, https://www.mdpi.com/1996-1073/14/5/1293, 2021.

van Kuik, G., Peinke, J., Nijssen, R., Lekou, D., Mann, J., Sørensen, J. N., Ferreira, C., van Wingerden, J., Schlipf, D., Gebraad, P., et al.: Long-term research challenges in wind energy–a research agenda by the European Academy of Wind Energy, Wind energy science, 1, 1–39, 2016.

Wagner, P.-R., Nagel, J., Marelli, S., and Sudret, B.: UQLab user manual – Bayesian inversion for model calibration and validation, Tech.
715 rep., Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, report # UQLab-V1.3-113, 2019.

Wagner, P.-R., Fahrni, R., Klippel, M., Frangi, A., and Sudret, B.: Bayesian calibration and sensitivity analysis of heat transfer models for fire insulation panels, Engineering Structures, 205, 110 063, 2020.

Wang, L., Liu, X., and Kolios, A.: State of the art in the aeroelasticity of wind turbine blades: Aeroelastic modelling, Renewable and Sustainable Energy Reviews, 64, 195–210, 2016.