Improving wind farm flow models by learning from operational data

Johannes Schreiber, Carlo L. Bottasso, Bastian Salbert, and Filippo Campagnolo

Wind Energy Institute, Technische Universität München, 85748 Garching bei München, Germany

Correspondence: Carlo L. Bottasso (carlo.bottasso@tum.de)

Abstract. This paper describes a method to improve and correct an engineering wind farm flow model by using operational data. Wind farm models represent an approximation of reality and therefore often lack accuracy and suffer from unmodeled physical effects. It is shown here that, by surgically inserting error terms in the model equations and learning the associated parameters from operational data, the performance of a baseline model can be improved significantly. Compared to a purely data-driven approach, the resulting model encapsulates prior knowledge beyond the one contained in the training data set, which has a number of advantages. To assure a wide applicability of the method —including also to existing assets— learning is here purely driven by standard operational (SCADA) data. The proposed method is demonstrated first using a cluster of three scaled wind turbines operated in a boundary layer wind tunnel. Given that inflow, wakes and operational conditions can be precisely measured in the repeatable and controllable environment of the wind tunnel, this first application serves the purpose of showing that the correct error terms can indeed be identified. Next, the method is applied to a real wind farm situated in a complex terrain environment. Here again learning from operational data is shown to improve the prediction capabilities of the baseline model.

1 Introduction

Knowledge of the flow at the rotor disk of each wind turbine in a wind power plant enables several applications, including wind farm control, the provision of grid services, predictive maintenance, the estimation of life consumption, the feed-in to digital twins and power forecasting, among others.

This paper describes a new method to improve a wind farm flow model directly from standard operational data. The main idea pursued here is to use an existing wind farm flow model to provide a baseline predictive capability; however, as all models contain approximations and may lack the description of some physical phenomena, the baseline model is improved (or "augmented", which is the term used in this work) by adding parametric correction terms. In turn, these extra elements of the model are learnt by using operational data. The correction terms capture effects that are typically not present in standard flow models (as, for example, secondary steering (Fleming et al., 2018) or wind farm blockage (Bleeg et al., 2018)), or that are highly dependent on a specific site or difficult to model upfront (as, for example, non-uniform inflow caused by local orography and vegetation).

Various wind farm flow models have been developed and are described in the literature. Whereas Direct Numerical Simulation (DNS) is still out of reach for practical applications due to its overwhelming computational cost, Large Eddy Simulation

(LES) methods are now routinely used for the modeling of wind farm flows (Fleming et al., 2014; Breton et al., 2017). Although invaluable for the understanding of the behavior of the atmospheric boundary layer and of wakes, LES is however still very expensive, so that its use outside of some specialized applications is limited. To reduce cost, one can resort to lower fidelity CFD models (Boersma et al., 2017), or to the extraction of reduced order models (ROMs) from higher fidelity ones

5 (Bastine et al., 2014). Instead of deriving models from first principles, another widely adopted approach is to use engineering models, which are expressed in the form of parametric analytical formulas with a limited number of degrees of freedom and hence a much reduced numerical complexity (Frandsen et al., 2006; Gebraad et al., 2014; Bastankhah and Porté-Agel, 2016). The present paper uses this last family of methods, although ideas similar to the ones developed here could be applicable also to higher fidelity models.

10 Even though engineering models are constantly improved and refined (Fleming et al., 2018), they will most likely always exhibit only a limited accuracy in many practical applications, for example whenever an important role is played by effects such as orography, (seasonal) vegetation, spatial variability of the wind, sea state roughness, the erection of other neighbouring wind turbines, the presence of obstacles, and others. In addition, low fidelity models often lack some physics, e.g. the flow acceleration caused by wake and rotor blockage, secondary steering or others. The idea pursued in this paper is then to take a

15 rather pragmatic approach: based on the realization that it will always be difficult —if not altogether impossible— to include all effects and all physics in a model of limited numerical complexity, a given model is corrected by unknown parametric terms, which are then learnt by using operational data.

The idea of improving an existing model based on measurements is hardly new, and it is actually an important topic in the areas of controls and system identification. For example, in the field of wind farm flows, a Kalman filtering approach has

20 been proposed by Doekemeijer et al. (2017) to update model predictions based on Lidar measurements. Here again the present paper takes a more pragmatic approach, and model updating is based exclusively on data provided by the standard Supervisory Control And Data Acquisition (SCADA) systems that are typically available on contemporary wind turbines. On the one hand this has the advantage that the proposed method is applicable to existing assets, as it does not necessitate of extra sensors. On the other, given that stored SCADA data typically represents 10-minute averages, this also implies that the models obtained by

25 this technique are of a steady-state nature. Although unsteady effects in wind farms are clearly important, steady-state models are still very valuable and can support many of the applications listed above. In addition, nothing prevents the generalization of the proposed approach to unsteady flow models, assuming that the relevant higher frequency data sets are available, which is already the subject of ongoing work from these authors.

The contemporary literature —and not only in the field of wind energy— indicates an increasing interest in data-driven ap-

30 proaches. Just to give one single example related to wake modeling, a purely data-driven approach has been recently described by Göçmen and Giebel (2018). However, the current enthusiasm for data should not make one forget that physics-based and analytical models are also extremely valuable because they often encapsulate significant knowledge on a given problem, often corroborated by a long experience. In fact, purely data-driven approaches suffer from a number of limitations that descend directly from a very simple and inevitable fact: a model that is exclusively based on data can only know what is contained in

35 the data set that was used to build it. Typically, this means that a very significant amount of data is necessary to obtain a model

that is sufficiently general and accurate. Furthermore, the data has to cover the entire spectrum of operation of the system. This also means that the model might have very poor knowledge (and hence poor performance) for rare situations or conditions that take place at the boundaries of the operating envelope, where few if any data points might be available.

An alternative to the purely data-driven approach is presented in this work, where a reference baseline model is augmented with parametric error terms, which are then identified using data. The baseline model already includes prior knowledge based on physics, empirical observations and experience. Therefore, even prior to the use of data, a minimum performance can be guaranteed. The model is augmented with parametric error terms, whose choice is driven by physics and the knowledge of the limitations of the baseline model. Once the errors are identified using operational data, their inspection can clarify the causes of discrepancy between model and measurements. Eventually, this can be used to improve the underlying baseline model. Furthermore, by looking at the magnitude of the identified errors, significant deviations from the baseline model can be flagged to highlight issues with the model itself, the data or the training process.

Finally, it should be noted that the identification of the error terms can be combined with the tuning of the parameters of the baseline model. This addresses yet another problem: tuning the parameters of a model that lacks some physics may lead to unreasonable values for the parameters, as the model is "stretched" to represent phenomena that is does not contain. By the proposed hybrid approach, the simultaneous identification of the parameters of the baseline model together with the ones of the error terms eases this problem, as unmodeled phenomena can be captured by the model-augmenting terms, thereby reducing the chances of nonphysical tuning of the baseline parameters.

The baseline model parameters and the extra correction terms have a different functional form in the augmented governing equations. Hence, they should be distinguishable from each other, as they imply different effects on the model. However, as for many identification problems, it is in general not possible to guarantee that all unknown parameters are observable and non-collinear given a set of measurements and, hence, given a certain informational content. To address this problem, the method proposed by Bottasso et al. (2014a) is used here, where the original unknown parameters are recast into a new set of statistically uncorrelated variables by using the Singular Value Decomposition (SVD) of the inverse Fisher information matrix. Once the problem has been solved in the space of the orthogonal uncorrelated parameters, the solution is mapped back onto the original physical space. This approach not only avoids the ill-posedness of the original problem, but also allows one to clarify which physical parameters are visible given a certain data set.

The paper is organized as follows. First, the baseline model is introduced in §2.1, together with a detailed description of the proposed parametric corrections in §2.2. Next, the SVD-based parameter identification method is presented in §2.3. The approach is then applied in §3.1 to a cluster of scaled wind turbines operating in the atmospheric test section of the wind tunnel of the Politecnico di Milano (Bottasso et al., 2014b). Goal of this first application is to show that a correct identification of the error terms can be achieved. This is indeed possible in the controllable and repeatable conditions of a wind tunnel, where inflow and wake characteristics can be precisely measured, something that is hardly possible today in the field. Specifically, it is shown that the method can correctly learn the lack of uniformity of the wind tunnel inflow, which is akin to what happens in a real wind farm because of orographic effects. Similarly, it is shown that secondary steering, which is completely absent from the baseline model used here, can be learnt by using turbine power measurements only. A more extended view on the

wind tunnel results is reported in appendix A. After having demonstrated the method in the known and controlled wind tunnel environment, a second application is developed in §3.2 that targets a real 43-turbine wind farm. Here results indicate that the augmented model has a markedly improved prediction capability when compared to the baseline one, thanks primarily to the identification of orographic effects on the inflow and the tuning of other model parameters. Finally, conclusions are drawn in Sect. 4.

2 Methods

2.1 Baseline wind farm flow model

The proposed method is applied here to the baseline wake model of Bastankhah and Porté-Agel (2016), implemented within the FLORIS framework (Doekemeijer et al., 2018). Given ambient wind conditions, steady state velocities within a wind farm can be computed by this model, together with the corresponding operating states and power outputs of all its turbines. First, ambient conditions are estimated from un-waked machines operating in free stream, which are identified by the turbine yaw orientations and the wake model (Schreiber et al., 2018). Then, power and thrust of the upstream turbines are computed based on the turbine aerodynamic characteristics, regulation strategy, and alignment with the local wind direction. Next, the wakes shed by these turbines are calculated in terms of their trajectory and speed deficit. In turn, this yields the velocity at the rotor disks of the turbines immediately downstream. In case of multiple wake impingements on a rotor, a combination model is used to superimpose multiple wake deficits. Similarly, an added turbulence model is used to estimate the turbulence intensity at a downstream turbine rotor disk, as this local ambient parameter affects the expansion of the wake. This process is repeated marching downstream throughout the wind farm until the last downstream turbine is reached.

In this work, the implementation uses the *selfSimilar* FLORIS velocity deficit model, the *rans* deflection model, the *quadraticRotorVelocity* wake combination model, and the *crespoHernandez* added turbulence model. The interested reader is referred to Bastankhah and Porté-Agel (2016), Crespo and Hernández (1996) and Doekemeijer et al. (2018) and references therein for detailed descriptions and derivations of these models.

Engineering wake models depend on a number of parameters, which should be tuned in order to obtain accurate predictions. For the specific model used in this work, these tunable factors are the wake parameters α, β, k_a, k_b, a_d, and b_d, and the turbulence model parameters TI_a, TI_b, TI_c, TI_d (Bastankhah and Porté-Agel, 2016).

In this work, the parameters are first set to an initial value, either taken from the literature or identified with ad hoc measurements; these initial values are held fixed throughout the analysis and not changed further. Corrections to the initial values are then expressed as

$$k = k^* + p_k, \tag{1}$$

where k is a model parameter, k^* its initial value and p_k the correction. Although this is not strictly necessary, this redundant notation helps highlight the changes to the nominal model parameters obtained by the proposed procedure.

2.2 Model augmentation

The engineering model described earlier is a rather simple approximation of a flow through a wind power plant and it is therefore bound to have only a limited fidelity to reality, with a consequently only limited predictive accuracy. Even for more sophisticated future models, it is difficult to imagine that all relevant physics will ever be precisely accounted for. But even if such a model existed, in practice one might simply not have all necessary detailed information on the relevant boundary and operating conditions that would be required. For example, one might not know with precision the conditions of the vegetation around and within a wind farm, with its effects on roughness and, hence, on the flow characteristics. In other words, it is safe to assume that all models are in error to some extent, and will probably always be.

To address this problem, the model can be pragmatically augmented with correction terms. Here one could take two alternative approaches: either a generic all-encompassing error term is added to the model, or "surgical" errors are introduced at ad-hoc locations in the model to target specific presumed deficiencies. The first approach could be treated with a brute-force parametric modeling approach, as for example by using a neural network. Here, the second approach was used, as it allows for more insight into the nature of the identified corrections. The specific parametric corrections used in the present paper are reviewed next. It is clear that these are only some of the many corrections that could be applied to the present baseline model, so that the following does not pretend to be a comprehensive treatment of the topic. Nonetheless, results indicate that some of these corrections are indeed significant, and provide for a marked improvement of the baseline model.

Non-uniform inflow. The inflow to a wind farm can exhibit spatial variability, mostly because of orographic and local effects, especially in complex terrain conditions. For example, commercial wind resource assessment tools include topographic speed-up ratios customarily computed by CFD models (Jacobsen, 2019). In contrast to this established practice, no direct or equivalent modeling of orographic effects are at present available in engineering wake models. Another reason for inflow variability may be due to wind farm blockage effects (Bleeg et al., 2018). Indeed, current wake models as the one used here assume that upstream turbines affect downstream ones through their wakes, but do not model the effects of downstream machines on the upstream ones. In a wind farm, depending on the wind direction and cross-wind location considered, the number and operating state of downstream turbines varies, which may induce a cross-wind speed variability in the inflow.

To capture some of these effects, the model ambient flow speed V_∞ is expressed here as a function of height above ground Z, cross-wind lateral position Y and ambient wind direction Γ as

$$V_\infty(Y, Z, \Gamma) = \left(1 + f_{\mathrm{augm,speed}}(Y, \Gamma, \boldsymbol{c}_{\mathrm{speed}}, \boldsymbol{p}_{\mathrm{speed}})\right) V_{\infty,0} \left(\frac{Z}{z_{\mathrm{h}}}\right)^{\alpha_{\mathrm{vs}}}, \tag{2}$$

where $V_{\infty,0}$ is the reference (baseline uncorrected) ambient flow speed, and z_{h} the reference height of the vertically sheared flow with exponent α_{vs}. Function $f_{\mathrm{augm,speed}}(Y, \Gamma, \boldsymbol{c}_{\mathrm{speed}}, \boldsymbol{p}_{\mathrm{speed}})$ is the speed correction term. This function is defined in the 2D space $Y \in [Y_{\mathrm{min}}, Y_{\mathrm{max}}]$, $\Gamma \in [\Gamma_{\mathrm{min}}, \Gamma_{\mathrm{max}}]$. For each value of the ambient wind direction Γ, Y is a lateral coordinate orthogonal to it that spans the width of the farm; hence, by selecting Γ_{min} and Γ_{max} a lateral inflow non-uniformity can be modelled for a given sector or the whole wind rose of directions. The (Y, Γ) space is discretized

into rectangular cells with corner nodes $\boldsymbol{c}_{\mathrm{speed}} = [\ldots;(Y_i,\Gamma_i);\ldots]$ (for an example, see Fig. 16). The corresponding unknown error nodal values are stored in vector $\boldsymbol{p}_{\mathrm{speed}}$, and bilinear shape functions interpolate the error in each cell based on the nodal values at its corners. Equation (2) could be extended to include also a longitudinal wind-aligned coordinate, similarly to the localized speed-up ratios of Jacobsen (2019), to model wind farm blockage effects.

5   Local orographic effects and blockage may also induce variability in the wind direction Γ. Similarly, the vertical shear exponent α_{vs} and turbulence intensity I may vary, for example on account of non-uniform roughness induced by vegetation or other obstacles. To include these effects in the farm flow model, the baseline quantities are augmented as

$$\Gamma(Y) = \Gamma_{\mathrm{ref}} + Y f_{\mathrm{augm,dir}}(\Gamma_{\mathrm{ref}}, \boldsymbol{c}_{\mathrm{dir}}, \boldsymbol{p}_{\mathrm{dir}}), \tag{3a}$$

$$\alpha_{\mathrm{vs}}(\Gamma) = \alpha_{\mathrm{vs,ref}} + f_{\mathrm{augm,shear}}(\Gamma, \boldsymbol{c}_{\mathrm{shear}}, \boldsymbol{p}_{\mathrm{shear}}), \tag{3b}$$

$$I(\Gamma) = I_{\mathrm{ref}} + f_{\mathrm{augm,I}}(\Gamma, \boldsymbol{c}_{\mathrm{I}}, \boldsymbol{p}_{\mathrm{I}}). \tag{3c}$$

In all these expressions, $(\cdot)_{\mathrm{ref}}$ indicates a baseline reference quantity, while function $f_{\mathrm{augm},(\cdot)}$ is a correction term. This function is defined on the 1D space $\Gamma \in [\Gamma_{\mathrm{min}}, \Gamma_{\mathrm{max}}]$, discretized with nodes $\boldsymbol{c}_{(\cdot)} = [\ldots;\Gamma_i;\ldots]_{(\cdot)}$, using linear shape functions to interpolate the corresponding nodal values $\boldsymbol{p}_{(\cdot)}$. Here again, by selecting Γ_{min} and Γ_{max}, corrections can be 15   applied to the whole wind rose or just to a sector.

Secondary steering. By misaligning a wind turbine rotor with respect to the incoming flow direction, the rotor thrust force is tilted, thereby generating a cross-flow force that laterally deflects the wake. As shown with the help of numerical simulations by Fleming et al. (2018), this cross-flow force induces two counter rotating vortices that, combining with the wake swirl induced by the rotor torque, lead to a curled wake shape. As observed experimentally by Wang et al. (2018), 20   the effects of these vortices result in additional lateral flow speed components, which are not limited to the wake itself but extend also outside of it. By this phenomenon, the flow direction within and around a deflected wake is tilted with respect to the upstream undisturbed direction. Therefore, when a turbine is operating within or close to a deflected wake, its own wake undergoes a change of trajectory —termed secondary steering— induced by the locally modified wind direction. Although models of this phenomenon are being developed (Martínez-Tossas et al., 2019), they significantly 25   increase the computational cost and are not yet available in standard implementations of engineering wake models as the one used here.

The change of wind direction $\Delta\Gamma$ at a downstream turbine induced by secondary steering (indicated by the subscript ss) is modeled here as

$$\Delta\Gamma(y) = f_{\mathrm{augm,ss}}(\tilde{y}, \Gamma_{\mathrm{init}}, \boldsymbol{p}_{\mathrm{ss}}), \tag{4}$$

30   where $f_{\mathrm{augm,ss}}$ is the correction term and $\tilde{y} = Y - y_{\mathrm{wc}}$ is the lateral distance to the wake centerline (see Fig. 1), defined in the baseline wind farm model as the locus of the points of minimum flow speed. According to the notation used in

Eq. (6.12) of Bastankhah and Porté-Agel (2016), Γ_{init} indicates the initial wake direction of the closest upstream turbine. The correction term is expressed as the difference of two Gaussian functions, and more precisely

$$f_{\text{augm,ss}}(\tilde{y},\Gamma_{\text{init}},\boldsymbol{p}_{\text{ss}}) =$$

$$\Gamma_{\text{init}}\left(p_{\text{ss},1}\exp\left(-0.5\Big(\frac{\tilde{y}+\text{sgn}(\Gamma_{\text{init}})p_{\text{ss},3}}{p_{\text{ss},2}}\Big)^2\right) - p_{\text{ss},4}\exp\left(-0.5\Big(\frac{\tilde{y}+\text{sgn}(\Gamma_{\text{init}})p_{\text{ss},6}}{p_{\text{ss},5}}\Big)^2\right)\right), \quad (5)$$

where $\boldsymbol{p}_{\text{ss}} = (p_{\text{ss},1},p_{\text{ss},2},p_{\text{ss},3},p_{\text{ss},4},p_{\text{ss},5},p_{\text{ss},6})$ is the vector of free parameters, where parameters 1 and 4 are related to the amplitude, 3 and 6 to the standard deviation, and 2 and 5 to the location of the correction functions. Since the Gaussian functions are not centered at the wake centerline and the effect of secondary steering is assumed to be symmetric with respect to the misalignment angle, the correction term depends also on the direction of wake deflection $\text{sgn}(\Gamma_{\text{init}})$.

This particular choice of the shape functions is motivated by the results shown in Fig. 8b of Wang et al. (2018). Indeed, LES simulations and measurements reveal the presence of a stronger lateral velocity component directed towards the wake on the leeward side of the wake itself, and of an opposite and weaker lateral component on the windward side. Such a distribution can be approximated by two Gaussian functions using Eq. (5).

Note that the change in local wind direction also leads to a slight lateral deflection of the non-uniform wind farm inflow introduced previously. More precisely, for a turbine that is located ΔX behind an upstream turbine, the non-uniform inflow expressed by Eq. (2) is evaluated at $Y + \Delta X \sin(\Delta\Gamma)$ instead of Y.

The upper subplot of Fig. 1 shows the hub height flow speed for two wind turbines modeled in FLORIS, the turbine rotor disks being indicated with thick black lines. The wake centerlines and the undisturbed free stream wind direction are indicated by black dotted and dashed lines, respectively. The upstream turbine is misaligned with respect to the incoming flow and therefore its wake is deflected laterally. Using the baseline wake model, the downstream turbine wake develops along the free stream wind direction. The lower subplot of the same figure shows the effects of the secondary steering correction term given by Eq. (5). The plot clearly shows that the downstream turbine wake path is affected by the locally changed wind direction.

Non-Gaussian wake and flow acceleration. Engineering wake models are based, among other hypotheses, on assumed shapes of the speed deficit. For example, the present baseline model assumes a Gaussian distribution of the speed deficit within the wake. Another assumption is that the flow outside the wake is undisturbed, and equal to the free-stream. However, these assumptions can at times not be exactly satisfied, as already observed by Xie and Archer (2017) and Martínez-Tossas et al. (2019), among others. For example, aisle jets are local accelerations of the flow outside of the wake, produced by local blocking in the neighborhood of an operating turbine. It has been reported that aisle jets can induce local flow speedups in excess of 10% of the undisturbed inflow (Dörenkämper et al., 2015).

To account for such effects, the wake velocity V_{wake} of the baseline model is corrected as

$$V_{\text{wake}}(d_{\text{wc}}) = V_{\text{wake,FLORIS}}(d_{\text{wc}})\Big(1 + f_{\text{augm,acc}}(d_{\text{wc}},\boldsymbol{c}_{\text{acc}},\boldsymbol{p}_{\text{acc}})\Big), \tag{6}$$
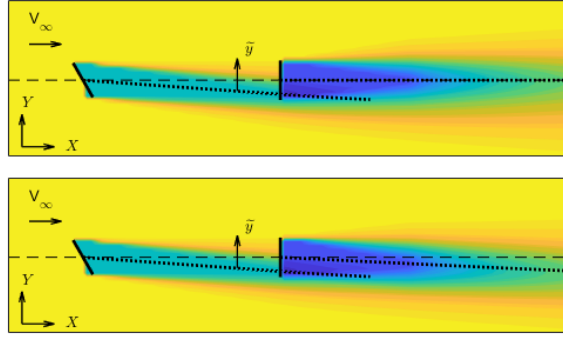
Figure 1. Effect of secondary steering on the trajectory of a downstream turbine. Top subplot: baseline wake model; lower subplot: baseline model augmented with the empirical correction term of Eq. (5).

where $V_{\text{wake,FLORIS}}$ is the baseline Gaussian wake speed profile, d_{wc} is the absolute distance to the wake center (which, at hub height, is equivalent to $|\tilde{y}|$), and $f_{\text{augm,acc}}$ represents the correction term, which —similarly to the previous corrections— is modeled with linear shape functions characterized by node locations $\boldsymbol{c}_{\text{acc}}$ (in terms of d_{wc}) and nodal values $\boldsymbol{p}_{\text{acc}}$.

5   **Reduced power extraction due to non-uniform wind turbine inflow.** Numerical simulations conducted in FAST (Jonkman and Jonkman, 2018) using its Blade Element Momentum (BEM) implementation yielded a slight reduction in the rotor power coefficient for horizontally sheared flow, when compared to unsheared conditions with the same hub wind speed. Even though BEM can only give a rough indication for such an effect, a correction of the power coefficient of the baseline model is introduced here in the form

$$C_{\text{P}} = C_{\text{P},\kappa=0}\Big(1 + p_\kappa \kappa^2\Big), \tag{7}$$

where $C_{\text{P},\kappa=0}$ is the nominal power coefficient, κ the equivalent horizontal linear shear coefficient on the rotor disk and p_κ the free correction parameter. The linear shear κ is either due to a lack of lateral uniformity of the inflow or to the impingement of a wake, and it is evaluated accordingly within the farm model.

Wind speed dependent power loss in yaw misalignment. The baseline formulation models the power extraction of a mis-
15   aligned wind turbine using the cosine-law $C_{\text{P}}(\gamma) = C_{\text{P}}\cos(\gamma)^{p_{\text{P}}}$, where C_{P} is the power coefficient of the wind-aligned turbine, γ the misalignment angle with respect to the local flow direction, and p_{P} the power loss exponent. Different power loss exponents have been reported in the literature, ranging from the value of 1.4 found by Fleming et al. (2017), to 1.8 according to Schreiber et al. (2017), 1.9 for Gebraad et al. (2015), all the way to the ideal value of 3 that is expected if only the rotor-orthogonal ambient flow component contributes to power extraction (Boersma et al., 2017). In addition,
20   p_{P} might also depend on the regulation strategy used by the turbine controller. Here, the power coefficient in misaligned

operation is augmented as

$$C_{\mathrm{P}} = C_{\mathrm{P}} \cos(\gamma + p_{\mathrm{P0}})^{p_{\mathrm{P}} + p_{\mathrm{P,a}}(V - V_{\mathrm{rated}}) + p_{\mathrm{P,b}}}, \tag{8}$$

where C_{P} is the power coefficient of the flow-aligned turbine (possibly reduced by shear effects, as argued above), p_{P0} is the misalignment angle at which the turbine produces maximum power, while V and V_{rated} are, respectively, the rotor effective and rated wind speeds. Finally, p_{P} is the baseline exponent, while $p_{\mathrm{P,a}}$ and $p_{\mathrm{P,b}}$ are free parameters that model a linear wind speed dependency of the cosine law.

2.3 Parameter identification method

The parameters of the baseline model and of its corrections terms are identified with the method developed by Bottasso et al. (2014a). The formulation of the parameter estimation problem is independent on whether the parameters belong to the baseline model or to its correction factors. In this sense, one can use the same method to just tune the baseline parameters without considering the correction terms, just identify the correction terms at frozen baseline model, or identify concurrently both sets.

The formulation is based on the classical likelihood function, which describes the probability that a given set of noisy observations can be explained by a specific set of model parameters. By numerically maximizing this function, a set of parameters is identified that most probably explains the measurements. Bound constraints are used to guide the process, and ensure convergence to meaningful results.

The accuracy with which the parameters can be estimated depends on how flat the likelihood function is with respect to changes in the parameters. For example, a flat maximum of the function implies that different nearby values of the model parameters are associated with similar values of the likelihood. These characteristics of the solution space are captured by the Fisher information matrix, which can be interpreted as a measure of the curvature of the likelihood function. Furthermore, it can be shown that the variance of the estimates is bound from below (Cramér-Rao bound) by the inverse of the Fisher matrix (Jategaonkar, 2015). Although the analysis of the Fisher information is useful for the understanding of the well-posedness of an estimation problem and of the quality of the identified model, it does not offer a constructive way of reformulating a given ill-posed problem. Indeed, a flat solution space and collinear parameters are to be expected in the present case, given the complex couplings and dependencies that may exist among the various parameters of a wind farm flow model and its correction terms.

To overcome this limitation of the classical maximum likelihood formulation, following Bottasso et al. (2014a), the original physical parameters of the model are transformed into an orthogonal parameter space, by diagonalizing the Fisher matrix using the SVD. This way, as the parameters are now statistically decoupled, one can set a lower observability threshold, and retain in the analysis only the ones that are in fact observable given the available set of measurements. Once the problem is solved, the uncorrelated parameters are mapped back onto the original physical space.

As shown later on, this approach achieves multiple goals: it allows one to successfully solve a maximization problem with many free parameters, some of which might be interdependent on one another or not observable in a given data set; it reduces the problem size, retaining only the orthogonal parameters that are indeed observable; it highlights, through the singular vectors,

the interdependencies that may exist among some parameters of the model, which provides for a useful interpretation tool that may guide the reformulation of parts of the model and its correction terms.

2.3.1 Maximum likelihood estimation of model parameters

A steady-state wind farm model can be mathematically expressed as

$$y = f(p, u),\tag{9}$$

where $f(\cdot, \cdot, \cdot)$ is the non-linear static function describing the wind farm model, which depends on free parameters $p \in \mathbb{R}^n$. These parameters can include both wake model parameters and/or model augmentation parameters. The model inputs $u \in \mathbb{R}^{n_u}$ include ambient wind conditions (i.e. ambient wind speed, direction, air density, turbulence intensity, etc.) and control inputs (i.e. yaw misalignment, partialization factor, blade pitch, rotor speed, etc. of each turbine). The model outputs $y \in \mathbb{R}^m$ represent quantities of interest for which measurements are available, in the present work these being the power outputs of each wind turbine in the farm. Experimental observations z of the simulated outputs y will in general result in a residual $r \in \mathbb{R}^m$, caused by measurement and process noise (e.g. plant-model mismatch), so that

$$z = y + r.\tag{10}$$

Given a set $S = \{z_1, z_2, ..., z_N\}$ of N independent observations, the likelihood function (Jategaonkar, 2015) can be defined as

$$\mathcal{L}(S\big|_p) = \prod_{i=1}^{N} p(z_i\big|_p),\tag{11}$$

where $p(\cdot)$ is the probability of S given p. Assuming the residuals r with covariance R to be statistically independent within the set of measurements (i.e. $E[r_i r_j{}^T] = R\delta_{i,j}$, where $\delta_{i,j}$ is the Kronecker delta), the likelihood function can be written, following Jategaonkar (2015), as

$$\mathcal{L}(S\big|_p) = \left((2\pi)^m \det(R^{-1})\right)^{-N/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{N} r_i{}^T R r_i\right).\tag{12}$$

Maximizing $\mathcal{L}$ (or minimizing its negative logarithm), a maximum likelihood estimate of the parameters can be obtained as

$$p_{\mathrm{MLE}} = \arg\min_{p} J(p),\tag{13}$$

where $J(p) = -\ln(\mathcal{L}(S\big|_p)$. The measurement noise covariance matrix R can be estimated under mild hypotheses as $R = \sum_{i=1}^{N} r_i{}^T r_i$, yielding $J(p) = \det(R)$, leading to an iteration between a solution at given covariance and a covariance update step (Jategaonkar, 2015). However, in this paper the measurement noise covariance matrix is estimated a priori and therefore assumed to be known. The cost function becomes therefore

$$J(p) = \frac{1}{2}\sum_{i=1}^{N} r_i{}^T R^{-1} r_i.\tag{14}$$

To ensure reasonable and physically viable solutions, parameters can be forced to stay within predefined upper (subscript ub) and lower (subscript lb) bounds, by adding the corresponding inequality constraints $\boldsymbol{p}_{\mathrm{lb}} \leq \boldsymbol{p} \leq \boldsymbol{p}_{\mathrm{ub}}$ to problem (13). As the parameter values and constraints can differ in magnitude, it is a good practice to scale all parameters such that a value of 1 corresponds to the upper bound p_{ub} and a value of -1 to the lower one p_{lb}. The optimization problem can finally be solved numerically by a suitable algorithm, such as Sequential Quadratic Programming (SQP) (Nocedal and Wright, 2006).

2.3.2 Identifiability of parameters

The Fisher information matrix $\boldsymbol{F} \in \mathbb{R}^{n \times n}$ is defined as

$$\boldsymbol{F} = \sum_{i=1}^{N} \left[\frac{\partial \boldsymbol{y}_i}{\partial \boldsymbol{p}} \right]^T \boldsymbol{R}^{-1} \left[\frac{\partial \boldsymbol{y}_i}{\partial \boldsymbol{p}} \right], \tag{15}$$

and describes the curvature of the likelihood function. It can be shown (Jategaonkar, 2015) that a lower bound (termed Cramér-Rao bound) of the covariance of the estimated parameter is given by

$$\boldsymbol{F}^{-1} = \boldsymbol{P} \leq \mathrm{Var}(\boldsymbol{p}_{\mathrm{MLE}} - \boldsymbol{p}_{\mathrm{true}}), \tag{16}$$

where $\boldsymbol{p}_{\mathrm{true}}$ are the true but unknown parameters. The k-th diagonal element of $\boldsymbol{P}$ is a lower bound on the variance of the k-th estimated parameter, while the correlation between different parameters is captured by the off-diagonal terms of that same matrix. The correlation coefficient between two parameters i and j is defined as

$$\Psi_{p_i, p_j} = \frac{P_{i,j}}{\sqrt{P_{i,i} P_{j,j}}}, \tag{17}$$

where $P_{i,j}$ denotes the i, j-th element (row, column) of $\boldsymbol{P}$. By analyzing the estimated parameter variance, as well as the correlation between parameters, valuable insight into the well-posedness of the parameter identification problem can be readily obtained.

2.3.3 Problem transformation and untangling using the SVD

When some parameters are highly correlated or have large variance, the problem is ill-posed: it might exhibit sluggish convergence, or not converge at all, and small changes in the inputs may lead to large changes in the estimates. Such situations are difficult to solve in the physical space, because parameters are typically coupled together to some degree through the model.

To untangle the parameters, one may resort to the SVD (Golub and van Loan, 2013). By this approach (Hansen, 1987; Waiboer, 2007; Bottasso et al., 2014a), the original parameters are mapped into a new set of uncorrelated (orthogonal) parameters. Since the new unknowns are uncorrelated, one can set a threshold to their variance by using the Cramér-Rao bound, and only retain in the optimization those that are observable within the given data set.

The Fisher matrix F is first factorized as $F = M^T M$, where $M \in \mathbb{R}^{Nm \times n}$ is defined as

$$M = \begin{bmatrix} R^{-1/2} \frac{\partial y_1}{\partial p} \\ R^{-1/2} \frac{\partial y_2}{\partial p} \\ \dots \\ R^{-1/2} \frac{\partial y_N}{\partial p} \end{bmatrix}. \tag{18}$$

Assuming a larger number of measurements than parameters ($Nm > n$), matrix M can be decomposed into

$$M = U \Sigma V^T, \tag{19}$$

where $U \in \mathbb{R}^{Nm \times Nm}$ and $V \in \mathbb{R}^{n \times n}$ are the matrices of left and right, respectively, singular vectors, while

$$\Sigma = \begin{bmatrix} S \\ 0 \end{bmatrix}, \tag{20}$$

where $S \in \mathbb{R}^{n \times n}$ is a diagonal matrix, whose entries s_i are the singular values sorted in descending order.

By using Eq. (19) and the factorization of F, the inverse of the Fisher information matrix can be written as

$$P = V S^{-2} V^T. \tag{21}$$

Note that the columns of the orthogonal matrix V are also the eigenvectors of P and s_i^{-2} the corresponding eigenvalues. Furthermore, P and F are symmetric and, based on the spectral theorem, diagonalizable.

The physical parameters p can now be transformed into a new set of orthogonal parameters Θ by a rotation performed with the right singular values:

$$\Theta = V^T p. \tag{22}$$

For the transformed set of parameters, the Cramér-Rao bound on the variance of the estimates is the diagonal matrix $S^{-2} \leq \mathrm{Var}(\Theta_{\mathrm{MLE}} - \Theta_{\mathrm{true}})$. Therefore, a small singular value s_i corresponds to a large uncertainty in the corresponding orthogonal parameter estimation.

To remove parameters that cannot be estimated with sufficient accuracy, matrix S can be partitioned as

$$S = \begin{bmatrix} S_{\mathrm{ID}} & 0 \\ 0 & S_{\mathrm{NID}} \end{bmatrix}, \tag{23}$$

where S_{ID} contains the identifiable singular values, i.e. those such that $s_i^{-2} < \sigma_t^2$, σ_t being a threshold on the highest acceptable standard deviation in the estimate. On the other hand, matrix S_{NID} contains singular values associated with parameters that cannot be identified with sufficient accuracy, and are therefore discarded. Accordingly, V is also partitioned as $V = [V_{\mathrm{ID}}, V_{\mathrm{NID}}]$, while the orthogonal parameters are partitioned as $\Theta = [\Theta_{\mathrm{ID}}^T, \Theta_{\mathrm{NID}}^T]^T$. Finally, the physical parameters are expressed in terms of the sole identifiable orthogonal parameters:

$$p \approx V_{\mathrm{ID}} \Theta_{\mathrm{ID}}. \tag{24}$$

Given that the Fisher matrix depends on the values of the parameters $\boldsymbol{p}$, an iterative procedure should be followed, where the diagonalization of the problem is repeated at each update of the parameter vector.

2.3.4 Identification method with variable measurement weights

In some cases, it may be useful to increase the importance of some measurements in the parameter estimation problem. This can be readily obtained by simply treating an observation with weight w as if it appeared w times in the observation data set (Karampatziakis and Langford, 2011). Cost function (14) then becomes

$$J(\boldsymbol{p}) = \frac{1}{2}\sum_{i=1}^{N} w_i \boldsymbol{r_i}^T \boldsymbol{R}^{-1} \boldsymbol{r_i}, \tag{25}$$

where w_i is the relative weight of observation i and $\sum_{i=1}^{N} w_i = N$. Similarly, the Fisher matrix becomes

$$\boldsymbol{F} = \sum_{i=1}^{N} w_i \left[\frac{\partial \boldsymbol{y_i}}{\partial \boldsymbol{p}}\right]^T \boldsymbol{R}^{-1} \left[\frac{\partial \boldsymbol{y_i}}{\partial \boldsymbol{p}}\right], \tag{26}$$

and its factorization is

$$\boldsymbol{M} = \begin{bmatrix} \sqrt{w_1}\boldsymbol{R}^{-1/2}\frac{\partial \boldsymbol{y_1}}{\partial \boldsymbol{p}} \\ \sqrt{w_2}\boldsymbol{R}^{-1/2}\frac{\partial \boldsymbol{y_2}}{\partial \boldsymbol{p}} \\ ... \\ \sqrt{w_N}\boldsymbol{R}^{-1/2}\frac{\partial \boldsymbol{y_N}}{\partial \boldsymbol{p}} \end{bmatrix}. \tag{27}$$

The remainder of the formulation is not affected by the introduction of weights.

3 Results

The proposed method is first applied in §3.1 to a wind tunnel experiment with a small cluster of three wind turbines, and then in §3.2 to a real wind farm consisting of 43 wind turbines. The former example aims at a verification of the correctness of the identified augmentations, given the known and controllable conditions of the scaled experiments, whereas the latter is meant to offer a first glimpse on the practical applicability of the new method in the field.

3.1 Wind tunnel verification

Whether identified model corrections are indeed physical or only an artefact of the model-measurement mismatch is difficult to prove in general. From this point of view, wind tunnel experiments provide for a unique opportunity to verify the concept proposed in this paper. Indeed, the overall flow within a cluster of turbines can be measured with good accuracy, and the experiments can be repeated in multiple desired operating conditions. The aim of this section is then to show that, even in the presence of multiple possibly overlapping model terms, the correct improvements to a baseline model can be learnt from operational data only.

3.1.1 Experimental setup

The experimental setup is composed by a scaled cluster of three G1 wind turbines, each of them equipped with active yaw, pitch and torque control. The turbines were operated in the boundary layer test section of the wind tunnel of the Politecnico di Milano. Details on the models and the wind tunnel are reported, among other publications, in Campagnolo et al. (2016a, b, c).

The turbines are labelled WT1, WT2 and WT3, starting from the most upstream one and moving downstream. The machines are mounted on a turntable, whose rotation is used to change the wind direction with respect to the wind farm layout. In the nominal configuration, i.e. for a turntable rotation $\gamma_{TT} = 0°$, the three turbines are aligned with the wind tunnel main axis —and hence with the flow velocity vector. The turbines are installed with a longitudinal spacing of 5 diameters (D), as shown in Fig. 2 with a view looking down towards the wind tunnel floor. As indicated in the figure, positive turntable rotations are clockwise. For $\gamma_{TT} \neq 0°$, the longitudinal distance between the turbines decreases slightly. However, considering that in this work the largest investigated turntable angle was $\pm 11.5°$, the longitudinal distance varied only between $4.9D$ and $5D$.

A pitot probe was placed at hub height, 3D upstream of the first G1 in the nominal configuration. The probe was therefore not placed on the turntable, and its position remained fixed with respect to the wind tunnel test section. A wind-tunnel-fixed reference frame, used in the following to discuss the results, is also depicted in Fig. 2. Its origin is placed at the turntable center, while the frame X axis is aligned with the wind direction, the Y axis points left looking downstream, and hence Z points vertically up from the floor to complete a right handed triad.

The yaw angle γ_{WTi} of the i-th wind turbine is positive for a counterclockwise rotation looking down onto the floor, as shown for WT1 in Fig. 2, and null when the rotor disk is orthogonal to X and, therefore, to the nominal wind direction.

Figure 3 shows a photo of the cluster of turbines, looking downstream with WT1 in the foreground. The wind tunnel floor is of a blue color, whereas the turntable is black.

The ambient wind speed $V_{\infty,0}$ measured by the pitot tube was, for all conducted experiments, between 5.20 and 5.75 m/s, which correspond to slightly below-rated conditions. The ambient turbulence intensity was equal to 6.12%, while the vertical shear was $\alpha_{vs} = 0.144$.

3.1.2 Model setup

The FLORIS model implementation used in this work is the one available online (Doekemeijer and Storm, 2018). All baseline model parameters are reported in Table 1 and taken from Campagnolo et al. (2019), where they were identified based on wake measurements of a single isolated G1 turbine.

Table 1. Initial FLORIS parameters for the G1 turbine.

α^*	β^*	k_a^*	k_b^*	a_d^*	b_d^*	TI_a^*	TI_b^*	TI_c^*	TI_d^*
0.9523	0.2617	0.0892	0.027	0	0	0.082	0.608	-0.551	-0.2773

Figure 4 shows the G1 power C_P and thrust C_T coefficients as functions of wind speed V. The curves were obtained from dynamic simulations conducted in turbulent inflow, using the same controllers implemented on the scaled models. The C_P and
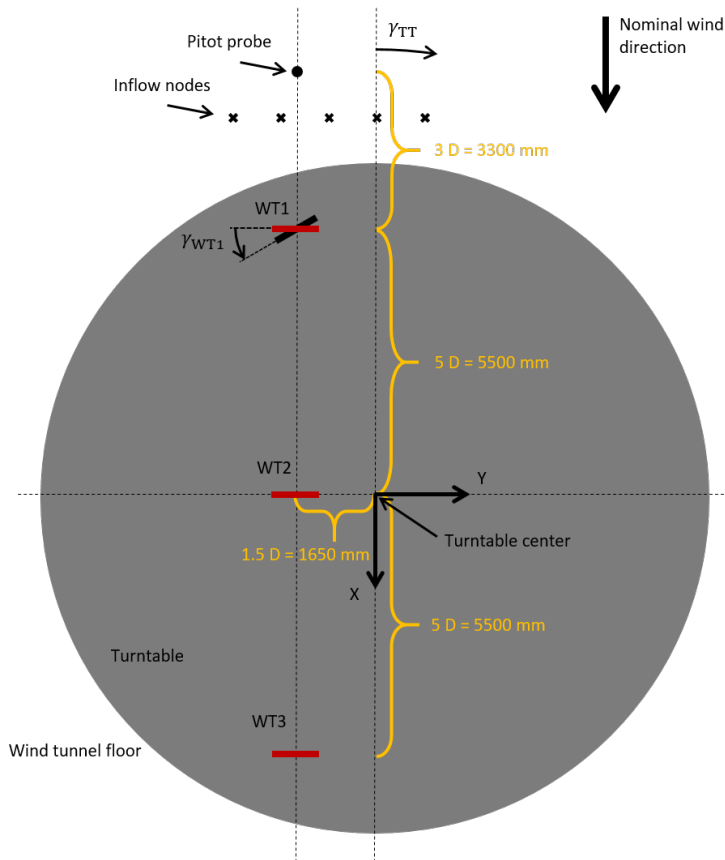
Figure 2. Wind farm layout for a null turntable rotation, looking down onto the wind tunnel floor.
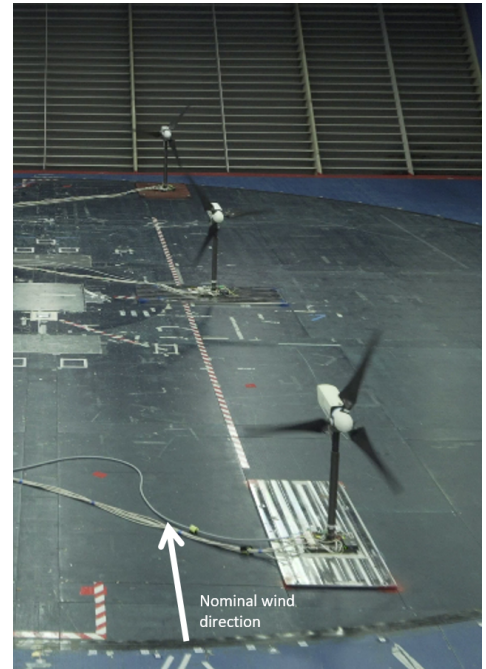
Figure 3. View looking downstream of the cluster of three G1 turbines.

C_T vs. tip-speed-ratio (TSR) and blade pitch setting curves were obtained with a BEM formulation using experimentally-tuned airfoil polars (Bottasso et al., 2014a). As the turbine controller does not consider variations of air density ρ, the coefficients shown in the figure exhibit a slight dependency on this ambient parameter. Within FLORIS, this effect is taken into account by interpolating within the coefficients based on the actual density measured in the wind tunnel during each experiment. For all reported test conditions, air density varied in the range $\rho \in [1.159, 1.185]$ kg/m^3. The power loss exponent in misaligned conditions was evaluated experimentally to be $p_P = 2.1741$, while for thrust the coefficient was found to be $p_T = 1.4248$.

The ambient wind speed was determined from the pitot tube. It was observed that, by using this value, the power of a free-stream turbine predicted by the FLORIS model was slightly underestimated, most probably due to the sheared flow. To correct for this effect, measurements provided by the pitot tube were scaled by the factor 1.0176, which was computed in order to match simulated and measured power. Furthermore, in the original FLORIS implementation the power of a turbine is computed as $P = 1/2\rho A V_{\text{avg}}^3 C_P$, where V_{avg} is the average wind speed at the rotor disk and A the rotor disk area. Here, power
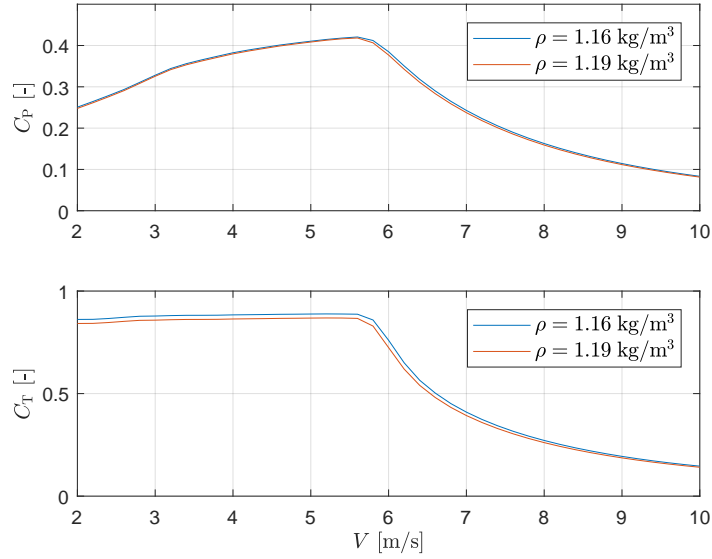
Figure 4. Power and thrust coefficients vs. wind speed for the G1 turbine.

was computed by integrating over the rotor disk area, i.e. $P = 1/2\rho \int_A V^3 C_{\text{P}} dA$, which is probably slightly more accurate even though it involves a minor increase in computational effort.

3.1.3 Ranking of correction terms

To initially assess the role of the various parameters, a ranking analysis was conducted. The parameters were clustered in sets, depending on their role in the model. A first identification was performed using all parameter sets, yielding the presumed best value, denoted as J_{ref}, of the cost function expressed by Eq. (14). The analysis was then repeated multiple times, each time removing one parameter set from the optimization. By looking at the resulting change in the value of the cost function, one may then rank the various parameter sets in order of importance. The analysis is based on a total of 190 experimental observations, as described in greater detail in the following.

All augmentation terms described in §2.2 were considered, except for the lateral variation in wind direction and the wind direction dependent vertical shear, as they are not applicable to the wind tunnel experiments. The non-uniform flow speed was modeled using five nodes located at $c_{\text{speed}}(Y) = [-3, -2, -1, 0, 1]$ m (which correspond to approximatively $[-2.7, -1.8, -0.9, 0, 0.9]$ D), and also indicated in Fig. 2 using $\times$ symbols. As only the turbine positions with respect to the flow are modified by rotating the turntable, a wind direction dependency was not included in this correction term. Table 2 reports the initial values and lower/upper bounds —chosen based on an educated guess— for the non-uniform inflow and secondary steering correction terms.

Table 2. Definition of the parameters, together with their initial values, lower and upper bounds, and identified values.

i	p_i	$p_{\text{lb},i}$	$p_{\text{ub},i}$	$p_{\text{init},i}$	$p_{\text{opt},i}$	Implementation
$1-5$	$\boldsymbol{p}_{\text{speed}}$	-0.1	0.1	0	$[0.079, 0.029, ...$ $-0.051, -0.006, 0]$	$f_{\text{augm,speed}}(Y, Z, 0, \boldsymbol{c}_{\text{speed}}, \boldsymbol{p}_{\text{speed}})$ $\boldsymbol{c}_{\text{speed}} = [-3, -2, -1, 0, 1]$ m
$6-11$	$\boldsymbol{p}_{\text{ss}}$	$[-3, 0, ...$ $-3, -3, ...$ $0, -3]$	$[3, 1.5, ...$ $3, 3, ...$ $1.5, 3]$	$[-0.5, 0.5, ...$ $0.2, -0.25, ...$ $0.5, -0.2]$	$[-0.94, 0.63, ...$ $0.20, -0.48, ...$ $0.73, -0.28]$	$f_{\text{augm,ss}}(\tilde{y}, \Gamma_{\text{init}}, \boldsymbol{p}_{\text{ss}})$

Figure 5 shows the relative increase of the cost function when eliminating one parameter set at a time. The figure clearly indicates that the most important parameters are the ones modeling laterally non-uniform speed and secondary steering. Indeed, this particular wind tunnel, due to its internal configuration and large width, does present a significant non-uniform flow speed, as already discussed by Campagnolo et al. (2019). Likewise, the effect of secondary steering is particularly important and should not be neglected for accurate predictions in misaligned conditions, as already reported in various publications. Based on these results, in the following only non-uniform inflow and secondary steering corrections are considered.
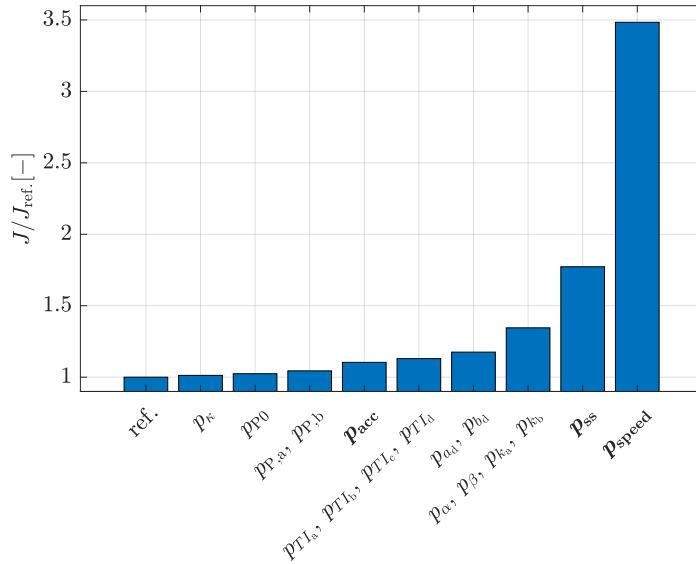
Figure 5. Relative increase of the optimization cost function when eliminating one parameter set at a time.

3.1.4 Results

A total of 451 observations were available, including 11 different turntable positions and thus wind farm layouts, with turbine yaw misalignments ranging from $-40°$ to $+40°$. A total of 190 observations were used to identify the 5 parameters associated

with non-uniform inflow speed and the 6 associated with secondary steering, whereas the remaining data points were used for model validation. The various tested configurations in terms of turbine misalignments and turntable positions are reported in the figures of Appendix A.

Among all the available measurements gathered at each operating condition, only the steady-state power of the wind turbines was utilized, mimicking what could be done at full scale in the field using SCADA data. The model outputs $\boldsymbol{y}$ (see Eq. (9)) are defined as

$$\boldsymbol{y} = \frac{1}{P_{\text{ref}}} \begin{bmatrix} P_{\text{WT1}} \\ P_{\text{WT2}} \\ P_{\text{WT3}} \end{bmatrix}, \tag{28}$$

where $P_{\text{WT}i}$ is the power of the i-th wind turbine and $P_{\text{ref}} = 37.6$ W is a reference value used as scaling factor. Based on experience, a diagonal measurement noise covariance matrix $\boldsymbol{R}$ with all three terms equal to $\sigma^2 = 0.025^2$ was specified.

The threshold of the highest acceptable standard variance σ_t^2 for the orthogonal parameters was set to 0.01. As the parameters are scaled within a range $[-1, 1]$, the threshold corresponds to a relative variance of 2%. Wind-aligned operating conditions (i.e. $\gamma_{\text{WT1}} = \gamma_{\text{WT2}} = \gamma_{\text{WT3}} = 0°$) were weighted with a factor of 2, to increase their importance in the parameter estimation process.

The constrained optimization problem (13) was solved in Matlab using the *fmincon* function with the *interior-point* algorithm (Mathworks, 2019). As the baseline model with its initial nominal values ($\boldsymbol{p} = \boldsymbol{p_{\text{init}}}$) is far away from the optimal solution, a first optimization was performed including only the inflow correction. Afterwards, three iterations were conducted including all 11 parameters. At each iteration, a total of 8 orthogonal parameters could be identified within the specified variance threshold. The method converged very quickly, as the identified parameters and the residual did not change significantly after the first iteration. Figure 6 shows on the left the initial variance of all 11 orthogonal parameters, and on the right the variance computed after the first iteration. The horizontal black line indicates the threshold σ_t^2.

Interestingly, the 11-th orthogonal parameter seems to have a very low observability. Table 3 shows the transformation matrix $\boldsymbol{V}^T$ that links the physical parameters to the orthogonal ones ($\boldsymbol{\Theta} = \boldsymbol{V}^T \boldsymbol{p}$, see Eq.(22)). The 11-th orthogonal parameter is almost entirely associated with $p_{\text{speed},5}$, which corresponds to the inflow speed augmentation node at position $Y = 1$ m. Indeed, the location of this node is such that it has only a very marginal effect on the turbine outputs and, hence, a very low observability, as shown later in Fig. 7. The transformation matrix reported in Table 3 also shows that the other two orthogonal parameters with low observability (9 and 10) represent secondary steering modes, mainly associated with the second Gaussian function of the correction term.

Table 4 presents the correlation matrix $\boldsymbol{\Psi}$ (cf. Eq. (17)), and shows a clear and to be expected dependency among neighbouring inflow parameters. Among the secondary steering parameters, strong but less obvious correlations are present, which suggest that a simplification of the assumed correction term might be possible.

Figure 7 shows the identified inflow augmentation function. In the picture, whiskers indicate the parameter uncertainty σ_i, computed based on the Cramér-Rao lower error bound as $\boldsymbol{\sigma} = \sqrt{\text{diag}(\boldsymbol{P})}$ (cf. Eq. (16)). The same figure reports also measurements obtained with hot wire probes in the empty wind tunnel at three different heights above the floor. These measurements,
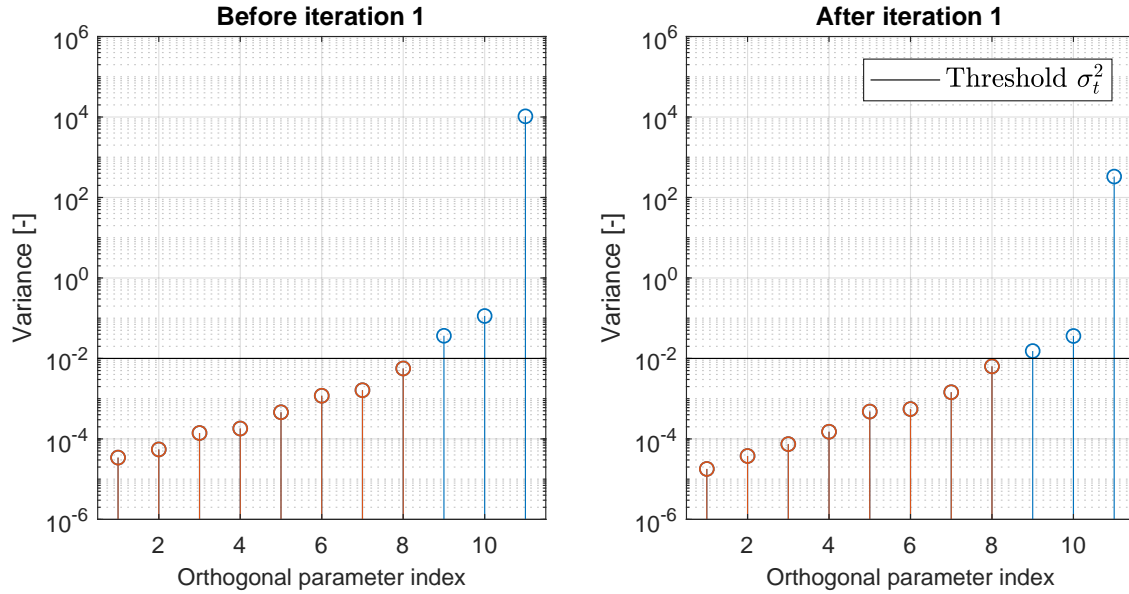
Figure 6. Variance of the orthogonal parameters before (left) and after (right) the first iteration. The identifiable orthogonal parameters are shown in red, whereas all others are shown in blue.

Table 3. Transformation matrix $\boldsymbol{V}^T$ after the first iteration. Each row corresponds to a different orthogonal parameter.

	$p_{\text{speed},1}$	$p_{\text{speed},2}$	$p_{\text{speed},3}$	$p_{\text{speed},4}$	$p_{\text{speed},5}$	$p_{\text{ss},1}$	$p_{\text{ss},2}$	$p_{\text{ss},3}$	$p_{\text{ss},4}$	$p_{\text{ss},5}$	$p_{\text{ss},6}$
1	-0.0	0.0	0.0	-0.0	-0.0	-0.7	0.2	-0.0	0.7	-0.1	-0.1
2	-0.2	-0.4	-0.3	-0.1	-0.0	0.2	-0.1	-0.7	0.3	0.1	0.3
3	0.0	-0.6	-0.6	-0.1	0.0	-0.1	0.0	0.4	-0.1	-0.0	-0.2
4	-0.4	-0.6	0.6	0.3	0.0	-0.0	0.0	0.1	-0.0	-0.0	-0.0
5	-0.7	0.2	-0.1	-0.2	-0.0	0.2	0.5	0.2	0.1	-0.1	0.1
6	-0.5	0.2	-0.1	-0.1	0.0	-0.4	-0.7	-0.0	-0.2	0.1	-0.2
7	0.1	-0.2	0.3	-0.9	-0.0	-0.0	-0.1	0.1	-0.0	-0.0	0.1
8	0.0	0.0	-0.0	0.1	-0.0	0.3	-0.5	0.5	0.5	0.1	0.4
9	-0.1	0.0	0.0	-0.1	0.0	0.2	0.1	0.0	0.2	0.8	-0.5
10	0.0	-0.0	0.0	-0.0	-0.0	0.4	-0.2	-0.1	0.3	-0.6	-0.6
11	0.0	-0.0	0.0	-0.0	1.0	-0.0	-0.0	0.0	-0.0	-0.0	0.0

Table 4. Correlation coefficients Ψ after the first iteration.

	$p_{\text{speed},1}$	$p_{\text{speed},2}$	$p_{\text{speed},3}$	$p_{\text{speed},4}$	$p_{\text{speed},5}$	$p_{\text{ss},1}$	$p_{\text{ss},2}$	$p_{\text{ss},3}$	$p_{\text{ss},4}$	$p_{\text{ss},5}$	$p_{\text{ss},6}$
$p_{\text{speed},1}$	1.0	-0.5	0.2	-0.1	0.2	-0.1	-0.1	-0.0	-0.1	-0.2	0.2
$p_{\text{speed},2}$	-0.5	1.0	-0.7	0.5	-0.2	-0.0	0.0	0.1	0.0	0.2	0.0
$p_{\text{speed},3}$	0.2	-0.7	1.0	-0.7	0.2	0.1	-0.0	-0.1	0.1	-0.1	-0.2
$p_{\text{speed},4}$	-0.1	0.5	-0.7	1.0	-0.4	-0.1	-0.0	0.1	-0.0	0.1	0.2
$p_{\text{speed},5}$	0.2	-0.2	0.2	-0.4	1.0	-0.1	-0.1	0.0	-0.1	-0.3	0.2
$p_{\text{ss},1}$	-0.1	-0.0	0.1	-0.1	-0.1	1.0	-0.6	-0.1	0.9	-0.4	-0.8
$p_{\text{ss},2}$	-0.1	0.0	-0.0	-0.0	-0.1	-0.6	1.0	-0.3	-0.7	0.6	0.3
$p_{\text{ss},3}$	-0.0	0.1	-0.1	0.1	0.0	-0.1	-0.3	1.0	0.2	0.4	0.6
$p_{\text{ss},4}$	-0.1	0.0	0.1	-0.0	-0.1	0.9	-0.7	0.2	1.0	-0.2	-0.6
$p_{\text{ss},5}$	-0.2	0.2	-0.1	0.1	-0.3	-0.4	0.6	0.4	-0.2	1.0	0.3
$p_{\text{ss},6}$	0.2	0.0	-0.2	0.2	0.2	-0.8	0.3	0.6	-0.6	0.3	1.0

and especially the ones at hub height, are in good agreement with the estimates provided by the proposed method. The figure also reports (with $\times$ symbols) the lateral position of the upstream turbine for the investigated turntable rotations. Noting that all points are shifted to the left, helps explain why the parameter associated with the inflow node at $Y = 1$ m has a very low —but still finite— observability.
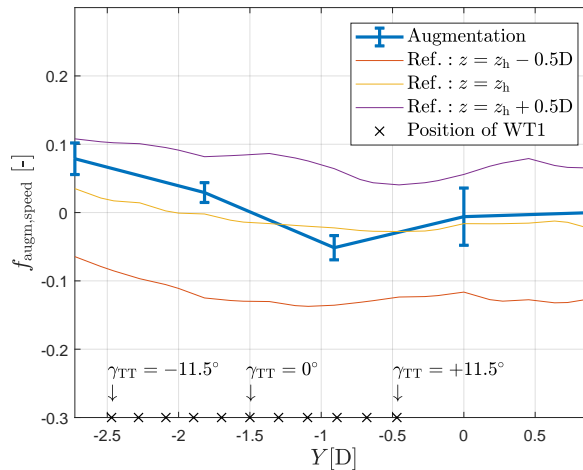
Figure 7. Identified non-uniform inflow speed augmentation term (solid line) and associated standard deviation (whiskers). Hot wire measurements at different heights above the floor are shown in thin solid lines. The upstream turbine (WT1) y position for all investigated turntable rotations is shown by $\times$ markers placed along the lower edge of the figure.

The identified secondary steering augmentation term is visualized in Fig. 8. The plot shows the wind direction change $\Delta\Gamma$ as a function of the distance $\tilde{y}$ to the wake centerline for a turbine misalignment of $20°$. The gray shaded area shows the uncertainty band $p_{\text{opt},i} \pm \sigma_i$. Consistently with the findings of Wang et al. (2018), the maximum change in wind direction is found at approximatively $0.3\,D$ on the leeward side of a deflected wake. The maximum magnitude of secondary steering in this operating condition is $1.9°$, which is again comparable to the results of Wang et al. (2018).
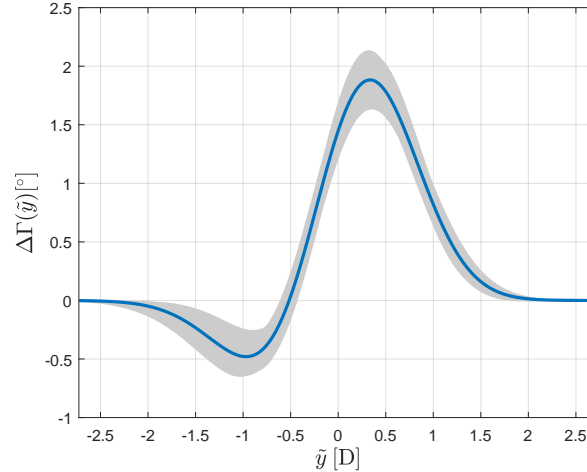
Figure 8. Identified wind direction change $\Delta\Gamma$ due to secondary steering as a function of distance $\tilde{y}$ to the wake centerline for a turbine misalignment of $20°$. The grey shaded area shows the uncertainty band.

The validity of the augmentation terms, identified as explained, was assessed by comparing the results of the simulation model with experimental wake measurements from a different test campaign. The setup was identical to the one considered here, except for the fact that only the first two upstream wind turbines were installed in the wind tunnel. At the downstream distance where the third wind turbine should have been installed, flow velocity measurements were obtained at turbine hub height using hot wire probes. Figure 9 shows wake profiles for the turntable position $\gamma_{\text{TT}} = 0°$ for various combinations of turbine yaw misalignments, as indicated by the subplot titles. Each subplot is accompanied by two flow visualizations, one based on the baseline FLORIS model and the other on its augmented version. The figures also include the points at which the flow was measured with the probes.

In the left subplots, the improvements of the augmented model with respect to the baseline FLORIS are exclusively due to the inflow correction, as the upstream turbine is aligned with the flow and therefore there are no secondary steering effects. In the right subplots, the upstream turbine is misaligned ($\gamma_{\text{WT1}} = 30°$) and secondary steering effects are present. Taking into account that model augmentation was obtained exclusively by turbine power measurements, the improved matching of the wake profiles is remarkable. Still, even with the extra correction terms some small model mismatches are present; these might be caused by the wake combination model, which was not augmented in this study.
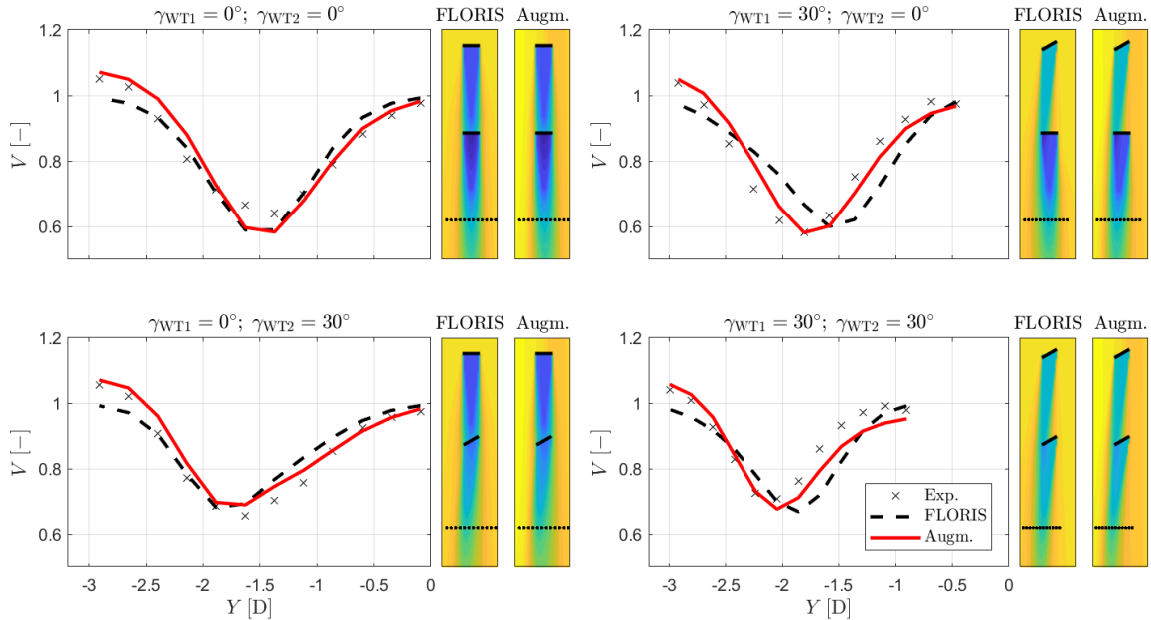
Figure 9. Wake profiles 5D behind WT2 for various combinations of turbine yaw misalignment. Experimental values are indicated by the × symbols. Each subplot is accompanied by two flow visualizations based on the FLORIS model and its augmented version.

The turbine power coefficients are computed as

$$C_{P,i} = \frac{P_{WTi}}{0.5 \rho A V_\infty (Y_{WTi}, z_h, 0)^3}, \tag{29}$$

where V_∞ is the augmented inflow function given by Eq. (2), evaluated at the respective turbine position Y_{WTi} and hub height z_h. A detailed overview of the results is offered by the figures of Appendix A, which report the power outputs and the model errors for all wind farm configurations. For readability, here a more synthetic overview of the results is presented, by condensing the information contained in Figs. A1, A2 and A3 in the probability density plots of Fig. 10. This figure shows the results for the baseline FLORIS model using a black dashed line, for the 11-parameter augmented model (i.e. including only non-uniform inflow speed and secondary steering corrections) using a red solid line, and for the 27-parameter augmented model (i.e. including all additional augmentation terms presented earlier) using a red dotted line. The root mean squared errors ϵ_{RMS} are shown in the respective legends.

Note that the FLORIS error distribution shows two peaks for WT1 and WT3, indicating the presence of two uncorrelated errors. The 11-parameter model removes these peaks, even though a smaller pair of peaks remains for WT2 and WT3, indicating additional errors that only the 27-parameter augmented model is able to capture.

Here again the trend is clear: the addition of non-uniform speed and secondary steering increases substantially the accuracy of the baseline model, with additional small —but not insignificant— gains offered by the additional correction terms. Finally, there is still room for improvement, possibly through extra correction terms not yet explored.
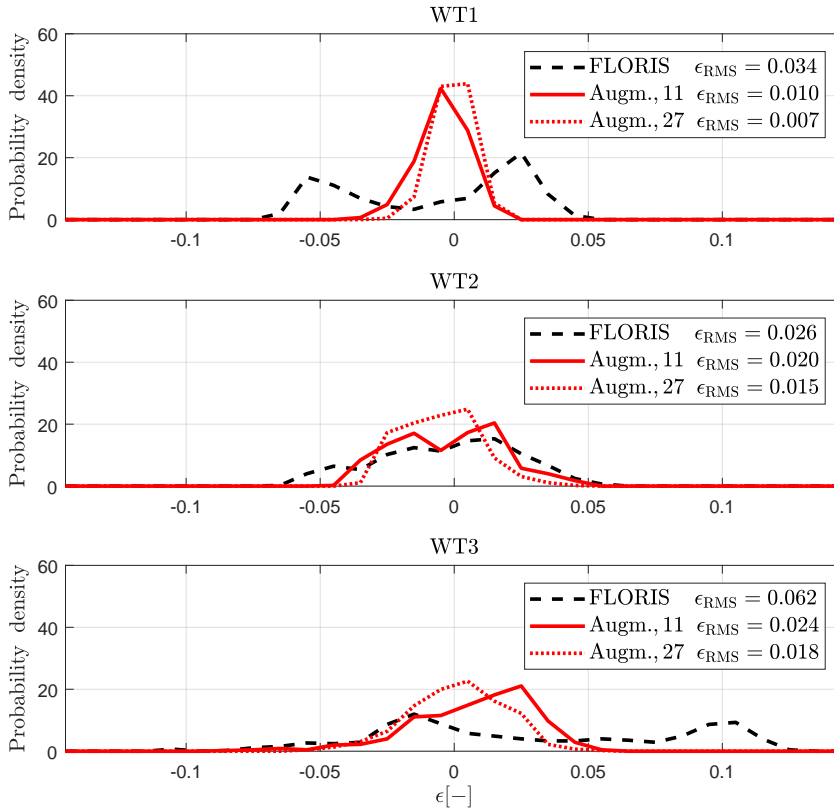
Figure 10. Error distributions for each turbine for all tested configurations, for the baseline FLORIS model (black dashed line), the 11-parameter augmented model (red solid line) and the 27-parameter augmented model (red dotted line).

3.2 Field application

In this section the model augmentation and identification method is applied to a full scale wind farm, to test its applicability and usability in a realistic scenario. In such conditions, it is often difficult to assess whether the identified model corrections are indeed physical or not, due to a lack of knowledge of the actual ground truth. To deal with this problem, the classical approach
5 of splitting the data set was used here: first, a relatively small subset of measurements is used for model and error identification; then, the rest of the data set is used for a verification of the generality of the identified model, and of its improved performance with respect to the baseline one.

3.2.1 Wind farm and data pre-processing

The onshore wind farm is situated close to Sedini, on the Italian island of Sardinia, and it consists of 43 GE1.5s and GE1.5sle wind turbines, as specified in Table 5.

Table 5. Turbine specifications

Type	Rated power [MW]	Cut-in wind speed [m/s]	Rated wind speed [m/s]	Rotor diameter [m]	Hub height [m]	Installed units [-]
GE1.5s	1.5	4	13	70.5	65	36
GE1.5sle	1.5	3.5	12	77	80	7

5 The wind farm is located at a rather complex terrain site, as shown in Fig. 11. Blue turbines are of type GE1.5sle, black and red turbines are of type GE1.5s, the latter being used as sensing turbines as explained later. Figure 12 shows a top view of the wind farm, including the turbine identifiers.
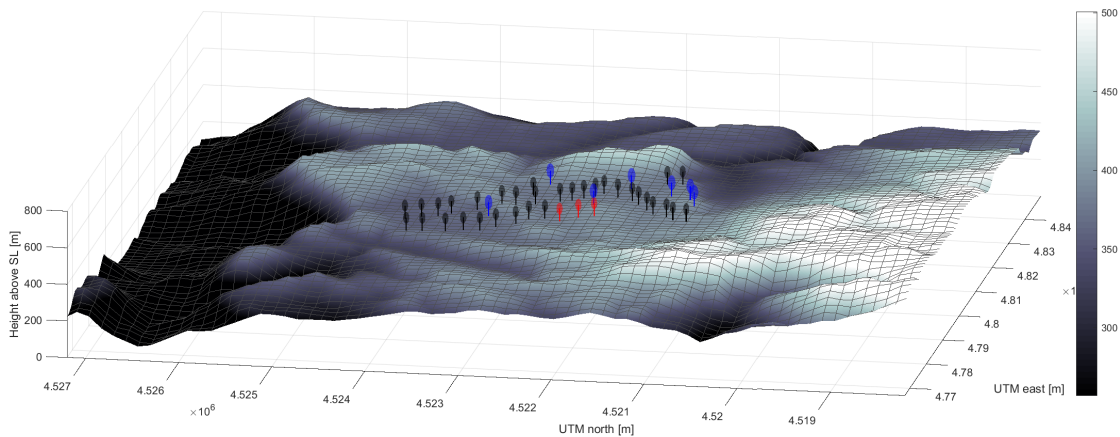
Figure 11. 3D view of the Sedini wind farm with terrain elevation, as seen from $\Gamma = 260°$.

Historical 10 min SCADA data was made available for this research for a period of 24 months, throughout the years 2015 and 2016. The recorded turbine yaw orientations exhibit sudden jumps and long term drifts. An ad-hoc algorithm was developed for detecting and correcting these data issues. On average, for each turbine 45% of the data points were missing, and 23% were 10 discarded because of low power output (< 5 kW) or rotor speed (< 1 rpm). As a result, about $33,700$ data points were available for each turbine. Regarding the missing data points, it is unknown whether the turbines were operating or just not reporting. To avoid eliminating a large fraction of the data set, it was assumed that the turbines were indeed operational and thus shedding wakes. This way, even if recordings of one or more turbines were missing at a specific time instance, the data points of the other turbines could still be used.
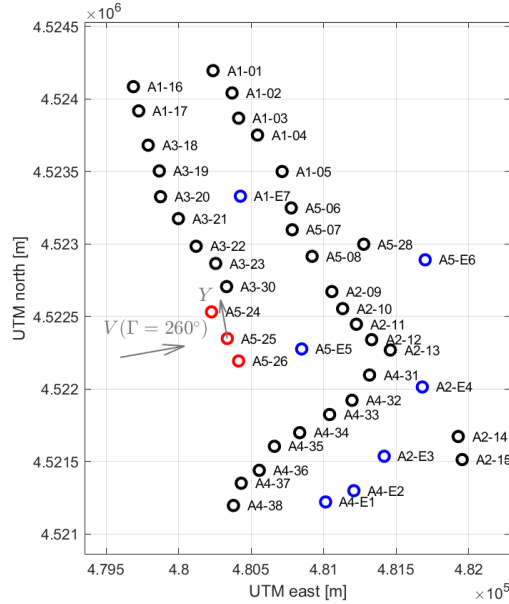
Figure 12. Top view of the Sedini wind farm with turbine identifiers. The gray arrows indicate the X and Y axes for an ambient wind direction $\Gamma = 260°$.

As no direct measurements of ambient conditions were available, the method described by Schreiber et al. (2018) was used to identify ambient wind speed and direction. The procedure works as follows. First, the ambient wind direction is estimated from turbine yaw orientations. Second, the ambient wind speed is estimated from the rotor effective wind speed of the free-stream turbines, computed from the turbine power curve below rated wind speed. To this purpose, the three sensing turbines A5-24, A5-25 and A5-26 indicated in red in Fig. 12 were used, checking that they were unwaked by using the flow model; the average of these speeds was attributed to the location of turbine A5-25. This way, $5,667$ ambient wind conditions could be processed for a range of wind directions $\Gamma \in [184°, 320°]$. Based on the ambient wind conditions, the data of all turbines was aggregated in two-dimensional bins: ambient wind speed (bin width of 2 m/s) and ambient wind direction (bin width of $5°$). Figure 13 shows the scaled number of measurements in each bin between 6 and 12 m/s.

3.2.2 Model setup

Here again the FLORIS implementation was based on the version available online (Doekemeijer and Storm, 2019). The initial values of both the wake and turbulence model parameters were set according to Bastankhah and Porté-Agel (2016) for (α^*, β^*), Crespo and Hernández (1996) for $(TI_a^*, TI_b^*, TI_c^*, TI_d^*)$, Niayifar and Porté-Agel (2015) for (k_a^*, k_b^*), and Gebraad et al. (2014) for (a_d^*, b_d^*), as reported in Table 6.
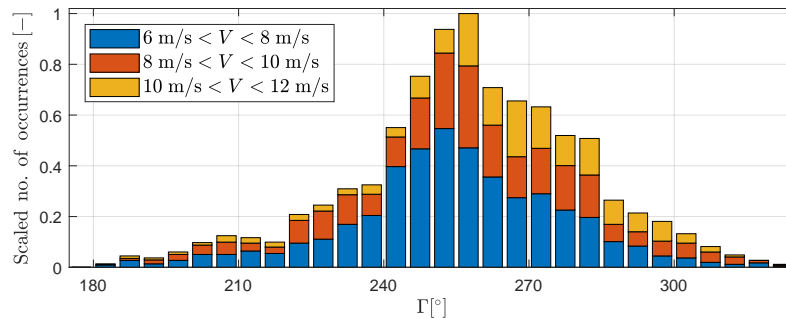
Figure 13. Scaled number of measurement data points (10 min mean) within each speed and direction bin.

Table 6. Initial FLORIS parameters for the Sedini wind farm.

α^*	β^*	k_a^*	k_b^*	a_d^*	b_d^*	TI_a^*	TI_b^*	TI_c^*	TI_d^*
2.32	0.154	0.3837	0.0037	-0.0356	-0.01	0.73	0.8325	0.0325	-0.32

The required turbine power and thrust versus wind speed curves were provided by the turbine manufacturer. The vertical shear exponent of the inflow was set to $\alpha_\mathrm{vs} = 0.143$ and the turbulence intensity to 14%, which represent annual average values measured at 65 m of height by an on-site met-mast. Air density was set to the constant value $\rho = 1.177$ kg/m^3.

The different turbine foundation heights were accounted for by accordingly increasing the tower heights, using the lowest

5   foundation height as reference (turbine A1-02). Indeed, power measurements of the upstream turbines show a correlation with the actual turbine hub height with respect to sea level (SL), as shown in Fig. 14. As indicated by the only approximate correlation shown by the figure, it is clear that such simple correction might not provide satisfactory results for all wind directions and all turbines, because complex orthographic flow effects might also play a role. Nonetheless, this approximate correction seems to be a step in the right direction. In addition, some of these effects may be corrected by the lateral non-

10   uniformity terms added to the augmented model. The reference height of the sheared inflow z_h (see Eq. (2)) was set to the hub height of the sensing turbine A5-25.

3.2.3   Ranking of correction terms

As for the wind tunnel experiments, here again a first analysis was aimed at ranking the various correction terms. However, since the turbines were operated with a conventional wind-aligned strategy, secondary steering corrections were neglected. The

15   ranking is based on data points in the range $V \in [8, 10]$ m/s, as described in greater detail in the following.

Figure 15 shows the relative increase in the cost function after optimization eliminating one set of parameters at a time. The results clearly indicate that the non-uniform wind farm inflow speed $\boldsymbol{p}_\mathrm{speed}$ is the most important correction. In fact, this was to be expected, given that the Sedini wind farm is located at a rather complex terrain site. Results indicate also a non-negligible effect of the wake deflection parameters for non-misaligned operation $(a_\mathrm{d}, b_\mathrm{d})$.
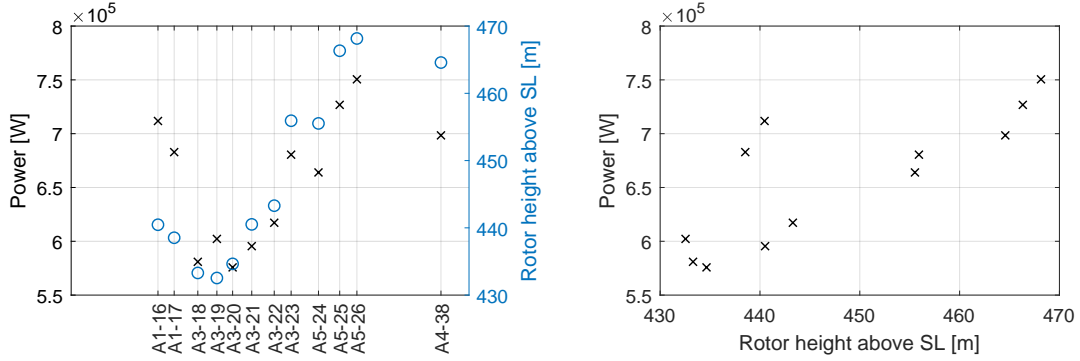
Figure 14. Correlation between power output and hub height with respect to SL. Left subplot: power ($\times$ symbols and left y axis) and rotor height above SL ($\circ$ symbols and right y axis) vs. lateral turbine position for a wind direction $\Gamma = 240°$. Right subplot: power vs. rotor height above SL for $\Gamma \in [220°, 275°]$ and $V_\infty \in [8, 10]$ m/s. All conditions are free-stream and all turbines of type GE1.5s.

On the other hand, the additional model augmentation parameters $(\boldsymbol{p}_{\mathrm{TI}}, \boldsymbol{p}_{\mathrm{winddir}}, \boldsymbol{p}_{\mathrm{acc}}, \boldsymbol{p}_{\mathrm{shear}})$ do not seem to contribute to a significant extent. Note also the slight retuning of parameters $(\alpha, \beta, k_{\mathrm{a}}, k_{\mathrm{b}})$ and $(TI_{\mathrm{a}}, TI_{\mathrm{b}}, TI_{\mathrm{c}}, TI_{\mathrm{d}})$, which can be explained with the fact that their initial values were taken from the literature, and therefore apply to different turbine types and sites.
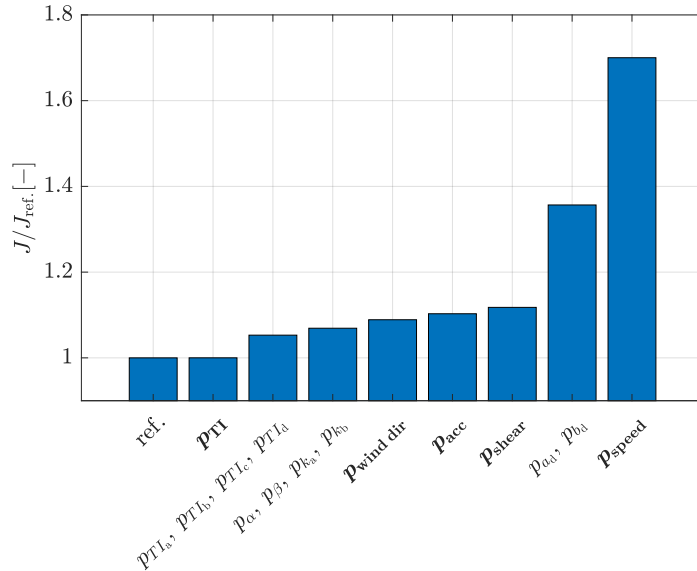
Figure 15. Relative increase of the optimization cost function for the Sedini wind farm when eliminating one parameter set at a time.

Given these results, the rest of the analysis is based only on the sub-set of parameters $\boldsymbol{p}_{\mathrm{inflow}}$, $(p_{a_{\mathrm{d}}}, p_{b_{\mathrm{d}}})$, (p_α, p_β), $(p_{k_{\mathrm{a}}}, p_{k_{\mathrm{b}}})$, $(p_{TI_{\mathrm{a}}}, p_{TI_{\mathrm{b}}}, p_{TI_{\mathrm{c}}}, p_{TI_{\mathrm{d}}})$. The augmentation term for non-uniform inflow speed is modeled using five nodes along the lateral position Y located at $[-2000; -1000; 0; 1000; 2000]$ m (which is approximatively $[-28; -14; 0; 14; 28]$ $\mathrm{D}_{\mathrm{GE1.5s}}$) and six nodes

in wind direction Γ at $[180; 210; 140; 270; 300; 330]°$, resulting in 30 nodes. The Y-coordinate axis is orthogonal to the wind direction and its origin $Y = 0$ m is located at the position of wind turbine A5-25, as shown in Fig. 12.

The definitions of the correction parameter, together with their bounds and converged values, are reported in Table 7. Note that all parameters were set to zero at the beginning of the identification process.

Table 7. Definition of the parameters, together with their lower and upper bounds, and initial and identified values.

i	p_i	$p_{\mathrm{lb},i}$	$p_{\mathrm{ub},i}$	$p_{\mathrm{init},i}$	$p_{\mathrm{opt},i}$	Implementation
$1-30$	$\boldsymbol{p}_{\mathrm{inflow}}$	-0.1	0.1	0	see Fig. 16	$f_{\mathrm{augm,speed}}(Y, Z, \Gamma, \boldsymbol{c}_{\mathrm{speed}}, \boldsymbol{p}_{\mathrm{speed}})$
31	p_α	$-\alpha^*$	4	0	0.7837	$\alpha = \alpha^* + p_\alpha$
32	p_β	$-\beta^*$	2	0	1.063	$\beta = \beta^* + p_\beta$
33	$p_{k_{\mathrm{a}}}$	$-k_{\mathrm{a}}^*$	1	0	-0.2440	$k_{\mathrm{a}} = k_{\mathrm{a}}^* + p_{k_{\mathrm{a}}}$
34	$p_{k_{\mathrm{b}}}$	$-k_{\mathrm{b}}^*$	0.1	0	0.01862	$k_{\mathrm{b}} = k_{\mathrm{b}}^* + p_{k_{\mathrm{b}}}$
35	$p_{a_{\mathrm{d}}}$	-0.5	0.5	0	-0.3169	$a_{\mathrm{d}} = a_{\mathrm{d}}^* + p_{a_{\mathrm{d}}}$
36	$p_{b_{\mathrm{d}}}$	-0.1	0.1	0	-0.02246	$b_{\mathrm{d}} = b_{\mathrm{d}}^* + p_{b_{\mathrm{d}}}$
37	$p_{TI_{\mathrm{a}}}$	$-TI_{\mathrm{a}}^*$	1	0	-0.09577	$TI_{\mathrm{a}} = TI_{\mathrm{a}}^* + p_{TI_{\mathrm{a}}}$
38	$p_{TI_{\mathrm{b}}}$	-1	1	0	0.3403	$TI_{\mathrm{b}} = TI_{\mathrm{b}}^* + p_{TI_{\mathrm{b}}}$
39	$p_{TI_{\mathrm{c}}}$	-1	1	0	0.4452	$TI_{\mathrm{c}} = TI_{\mathrm{c}}^* + p_{TI_{\mathrm{c}}}$
40	$p_{TI_{\mathrm{d}}}$	-1	1	0	-0.3337	$TI_{\mathrm{d}} = TI_{\mathrm{d}}^* + p_{TI_{\mathrm{d}}}$

3.2.4 Results

To identify the 40 parameters of Table 7, only aggregated mean power measurements for wind speeds $V \in [8, 10]$ m/s were used. In addition, only one third of all wind direction bins were employed, with bin centers at $[192.5 : 15 : 312.5]°$, resulting in a total of 9 measurement corridors. The remaining wind direction and speed bins were reserved for validation.

The model outputs $\boldsymbol{y}$ (cf. Eq. (9)) were defined as

$$\boldsymbol{y} = \frac{1}{P_{\mathrm{ref}}} \begin{bmatrix} P_{\mathrm{WT1}} \\ ... \\ P_{\mathrm{WT43}} \end{bmatrix}, \tag{30}$$

where $P_{\mathrm{WT}i}$ is the power of wind turbine i and $P_{\mathrm{ref}} = 1.11$ MW a reference wind turbine value used as scaling factor. A diagonal measurement noise covariance matrix $\boldsymbol{R}$ was used, with all diagonal terms equal to $\sigma^2 = 0.01^2$. The threshold of the highest acceptable variance in the orthogonal parameter estimate was set to $\sigma_t^2 = 0.01$, which corresponds to a relative variance of 2%. The relative weight of each observation was set proportional to the number of measurement points within the respective bin. In a first iteration, 29 orthogonal parameters could be identified. In the second and third iteration only 23 and 25 orthogonal parameters fell below the threshold, although results changed only marginally after the first iteration.

The identified optimal parameter values $p_{\mathrm{opt},i}$ are included in Table 7 and, for the inflow augmentation, are also reported in Fig. 16. The plot shows, according to the colormap, the inflow augmentation function values $f_{\mathrm{augm,speed}}(Y, \Gamma, \boldsymbol{c}_{\mathrm{speed}}, \boldsymbol{p}_{\mathrm{speed}})$

in the left part of the figure. Each nodal point is indicated by a circle marker. The figure shows that significant variations in the inflow speed have been detected: for example, considering $\Gamma = 270°$, the inflow speed at $Y = +1000$ m (approximately at the location of wind turbines A3-19/20/21) is 3.5% smaller than the one measured at the reference turbines A5-24/25/26. For the same wind direction, the speed at $Y = -1000$ m (approximately located at the wind turbines A4-36/37/38) is 4.8% larger. These variations are expected to be mainly caused by terrain effects. The right subplot of Fig. 16 shows the parameter uncertainty (Cramér-Rao bounds). The parameter at the nodal point ($Y = -2000$ m; $\Gamma = 330°$) is completely unobservable, because it lies far outside of the wind farm perimeter (see Fig. 12). Some of the outer nodal points at $Y = \pm 2000$ m do show significantly increased uncertainties. However, the corresponding augmentation parameters (left subplot) are approximatively zero.
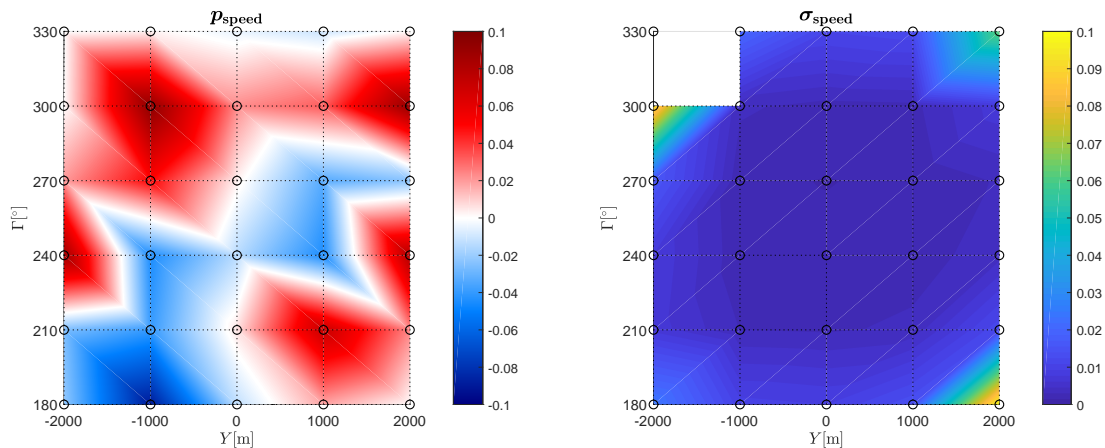
Figure 16. Identified inflow augmentation parameters (left subplot) and their uncertainties (right subplot). Nodal points are indicated by the circle markers.

Figure 17 shows the power coefficient of each individual wind turbine, as indicated by the subplot title, as function of wind direction. The power coefficient is computed as $C_P = P/(0.5\rho A V^3)$, where $\rho = 1.177$ kg/m^3 is the constant air density, $A = \pi(70.5/2)^2$ m^2 a reference rotor area, and V the corresponding estimated ambient wind speed. Blue crosses indicate SCADA data points, the ones used for identification having been encircled. The gray shaded area indicates the standard deviation within the binned measurements. The FLORIS (non-augmented) power estimates are shown by black dashed lines, whereas the augmented model results are shown using red solid lines.

Even though the baseline FLORIS power estimates already exhibit a reasonable correlation with the measurements for many turbines and wind directions, a significant improvement is achieved by the augmented model. Note that for $\Gamma < 210°$ and $\Gamma > 300°$ the number of measurement points within each bin is reduced (see Fig. 13), limiting the measurement quality/trustworthiness. More specifically, the augmented model shows improvements in the modeling of the free stream turbine power, due to the effects of the wind farm inflow augmentation terms. Furthermore, the predictions of the wake-induced power deficits are corrected, improving in many cases the deficit depth as well as the deficit location in terms of wind direction.
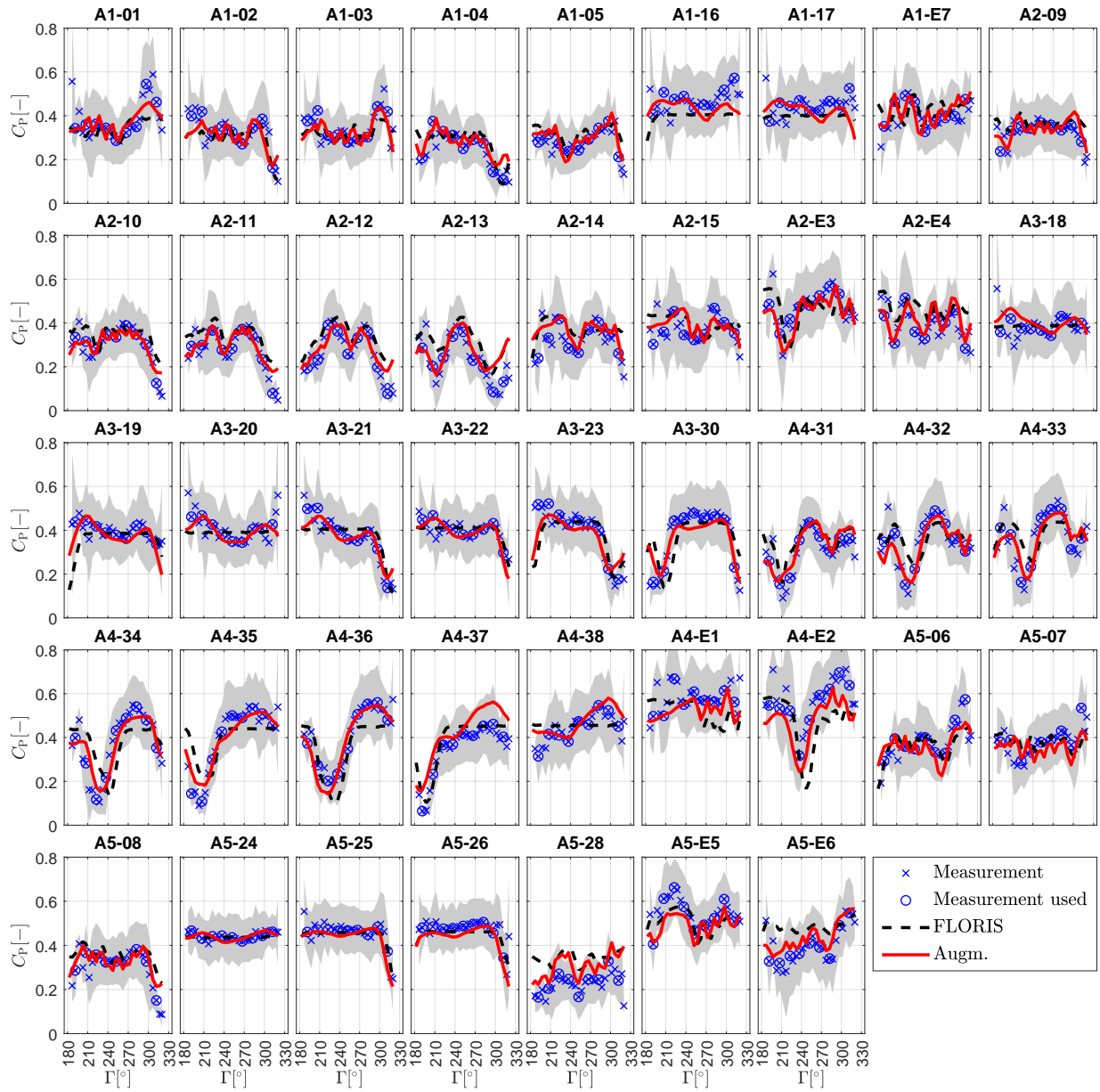
Figure 17. Power coefficient of each individual wind turbine, as indicated by the subplot title, as function of wind direction Γ for wind speeds $V \in [8, 10]$ m/s. The gray shaded area indicates the standard deviation within the binned measurements. The number of measurements within each bin is reported in Fig. 13.

The same results of Fig. 17 are also presented in a more synthetic form in terms of error probability densities in Fig. 18, where the error is defined as $\epsilon = C_{\mathrm{P,Meas.}} - C_{\mathrm{P,FLORIS/Augm.}}$. Each subplot shows the results for a different wind speed range. Note that the modeling error is reduced also for wind speed ranges that have not been used for model identification (i.e. $V \in [6,8]$ m/s and $V \in [10,12]$ m/s). The overall root mean squared error is reported within the legend, showing error reductions of 14%, 22% and 19%, respectively, highlighting the generality of the identified model and augmentation parameters.
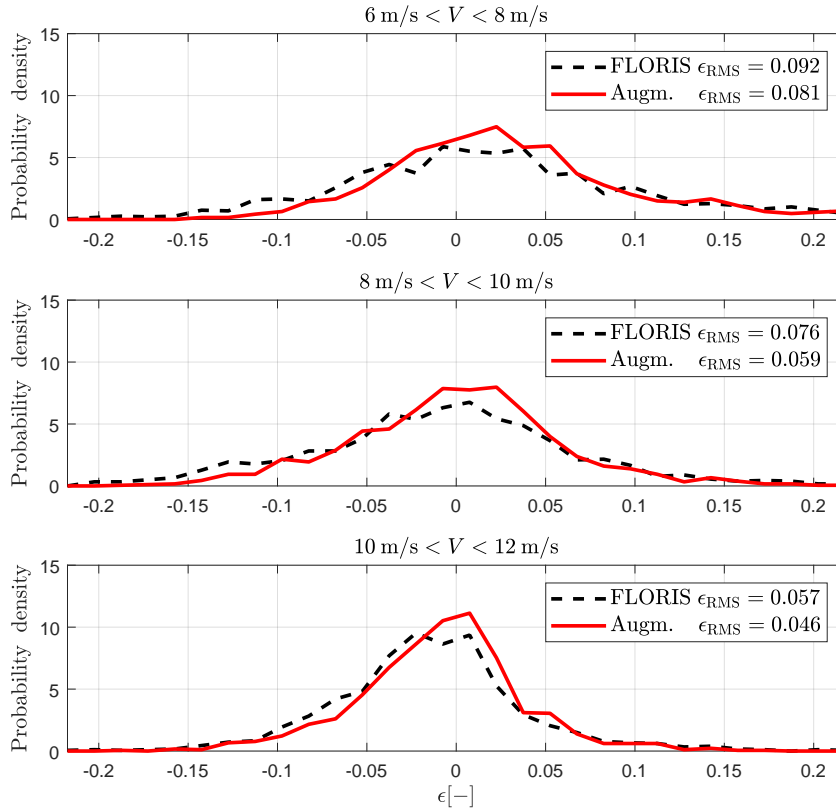
Figure 18. Error probability density functions for different wind speed ranges.

4   Conclusions

This paper has presented a new method to calibrate and augment parametric wind farm models. The proposed approach builds on the vast body of knowledge and experience embedded in available reduced wind farm flow models. However, recognizing that any such model will always have only a limited prediction accuracy, the present approach augments a baseline model with ad-hoc extra terms designed to correct some of its presumed specific deficiencies. These additional elements of the model

are then learnt from operational data. Optionally, the baseline model parameters can also be tuned within a single integrated process. By design, the method has been exclusively based here on SCADA power measurements; therefore, it is readily applicable to most operational wind farms, whenever such data is available. However, the concept of model augmentation is very general and could clearly be used also with other measurements.

5   To limit the number of free parameters and to overcome the fact that the identification problem can be over-parameterized and hence ill-posed, a parameter transformation into an orthogonal space has been used. Thereby, only parameters that are sufficiently visible within a given data set enter into the identification process.

The method was first applied to a large data set obtained with scaled wind turbines operating in a boundary layer wind tunnel. Thereby, it was shown that a correct learning of the extra modeling terms is achieved. These conclusions are made possible by

10   the fact that, in this case, the flow and wake characteristics are known with good accuracy. Next, the method was tested on a real wind farm, in a realistic and highly complex situation.

Based on the results shown here, the following conclusions can be drawn:

- Within the wind tunnel environment, a correct learning of non-uniform wind farm inflow speed and of secondary steering effects has been achieved. In particular, the latter shows a good match with detailed wake measurements in wind

15   misaligned conditions. It is remarkable, and very promising, that such detailed features of the solution could be inferred purely from operational power data, even when starting from a baseline model that does not consider at all secondary steering.

- The application to field data has shown that, as expected for the complex terrain site analyzed here, orographic effects play a driving role. A marked model improvement could be observed, even in conditions where the model was used for

20   extrapolating outside of the training conditions. It is worth noting that, in many practical onshore applications, orographic effects will be present, and the fact that one can learn them from simple and readily available operational data is very encouraging. Again, it should be explicitly pointed out that the baseline model did not include any orographic corrections.

- It has been shown that model tuning and the learning of extra correction terms can be achieved simultaneously. This reduces the risk of adapting the baseline parameters beyond their reasonable limits, driven by unmodeled physics.

25   - Although the augmented models show a much improved accuracy with respect to the baseline, some model mismatch still remains. Although these remaining errors may often be caused by issues in the data rather than in the model, additional improvements are thought to be possible.

Future work will apply the proposed method to other wind farms, to increase confidence in the obtained results. From longer and richer data sets, possibly in conjunction with meteorological reanalyses, it is presumed that yearly and seasonal variations

30   could be observed. The integration of CFD analyses can be used to support and confirm the identification of orographic effects. Attention should also be paid to improved and additional forms of model corrections, including wake overlap models. Finally, it is worth pointing out again that an improved knowledge of the flow within a wind farm finds applicability in a potentially large range of digitally-driven applications, including wind farm control, lifetime estimation, power forecasting, predictive

maintenance and others. Therefore, it is expected that methods for high-accuracy flow predictions in wind farms will be the subject of significant future research efforts.

Code and data availability. A MATLAB implementation of the wind farm model can be obtained by contacting the authors.

Appendix A: Extended wind tunnel results

5   Figures A1, A2 and A3 report the power outputs of WT1, WT2 and WT3, respectively, for all tested configurations. In each figure, clusters of three subplots represent a unique turntable position, as indicated by the title and the wind farm layout sketch therein. The left part of each subplot shows the turbine power coefficient $C_{\mathrm{P,WT}i}$ as a function of γ_{WT1} (x-axis) and γ_{WT2} (y-axis). All measured configurations are indicated by a small cross symbol, whereas the measurements used for parameter identification are circled. The central part of each subplot shows the FLORIS model error $\epsilon_{\mathrm{FLORIS}} = C_{\mathrm{P,Meas.}} - C_{\mathrm{P,FLORIS}}$,

10   including an annotation of the root mean squared error ϵ_{RMS}. Similarly, the right part of each subplot shows the augmented model error $\epsilon_{\mathrm{Augm.}}$.

For the first upstream wind turbine, WT1, the baseline FLORIS shows significant errors depending on the turntable position. For $\gamma_{\mathrm{TT}} < 0°$ the model under-predicts turbine power because of the lack of uniformity of the flow, as also shown in Fig. 7. The opposite behaviour can be seen for $\gamma_{\mathrm{TT}} > 0°$. The augmented model however shows significant improvements, which are due

15   to the inflow correction. Still some under-prediction for $\gamma_{\mathrm{TT}} = -11.5°$ is present, which is probably caused by an excessively small number of parameters in the inflow augmentation function and/or by the third wind turbine power measurements, which are also strongly affected by lateral inflow variations.

The power of WT2, shown in Fig. A2, is only weakly affected or improved by the model corrections. In fact, in all investigated conditions, the second turbine lateral position remains almost constant, such that the inflow correction does not have a

20   significant direct effect. However, secondary steering only slightly changes the inflow direction at WT2; for example, as shown in Fig. 8, a $20°$ misalignment of WT1 changes the wind direction by about $1.9°$. This leads to small misalignments and thus only very small changes in power output considering the cosine-law. In addition, secondary steering leads also to a slight lateral deflection of the non-uniform inflow.

The power of WT3, reported in Fig. A3, shows significant improvements when using the augmentation terms. For example,

25   for $\gamma_{\mathrm{TT}} > 0°$ the baseline model under-predicts the real flow velocities —and hence the power output— at WT3, an error that is corrected by the augmented model. In addition, for $|\gamma_{\mathrm{WT1}}| > 0$, secondary steering augmentation affects the deflection of the second turbine wake (as shown in Fig. 8), leading to further improvements.

Author contributions. JS conducted the main research work. CLB developed the core idea of model augmentation, its formulation and the overall solution methodology, and supervised the whole research. JS and CLB wrote the manuscript. BS pre-processed the field measure-
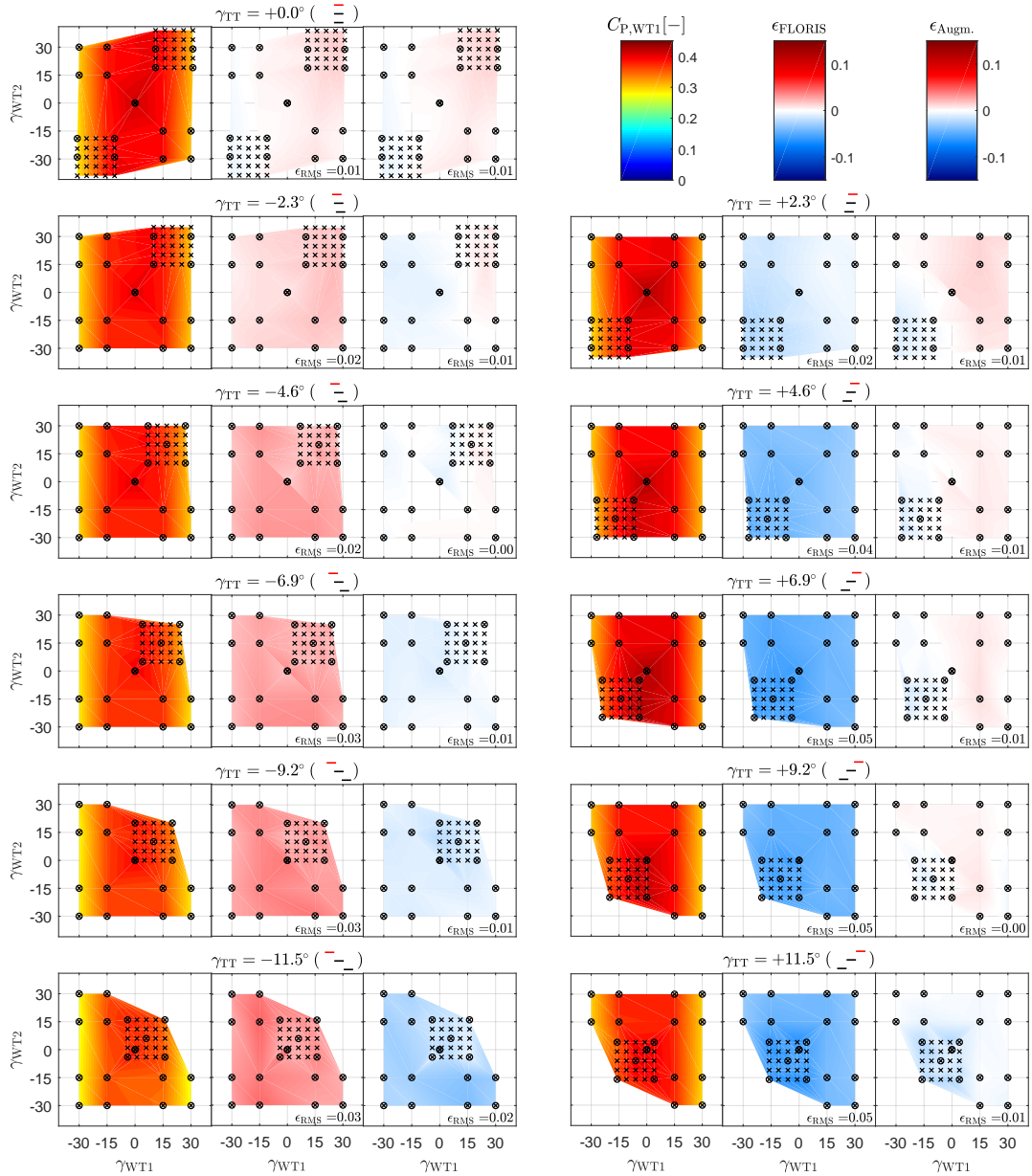
Figure A1. Wind turbine WT1. Each cluster of three subplots represents a unique turntable position, as indicated by the title and the wind farm layout sketch. Left subplot: turbine power coefficient $C_{P,WT1}$ as a function of γ_{WT1} (x-axis) and γ_{WT2} (y-axis). Middle subplot: FLORIS model error. Right subplot: augmented model error. Cross symbols: all measured configurations. Circles: conditions used for parameter identification.
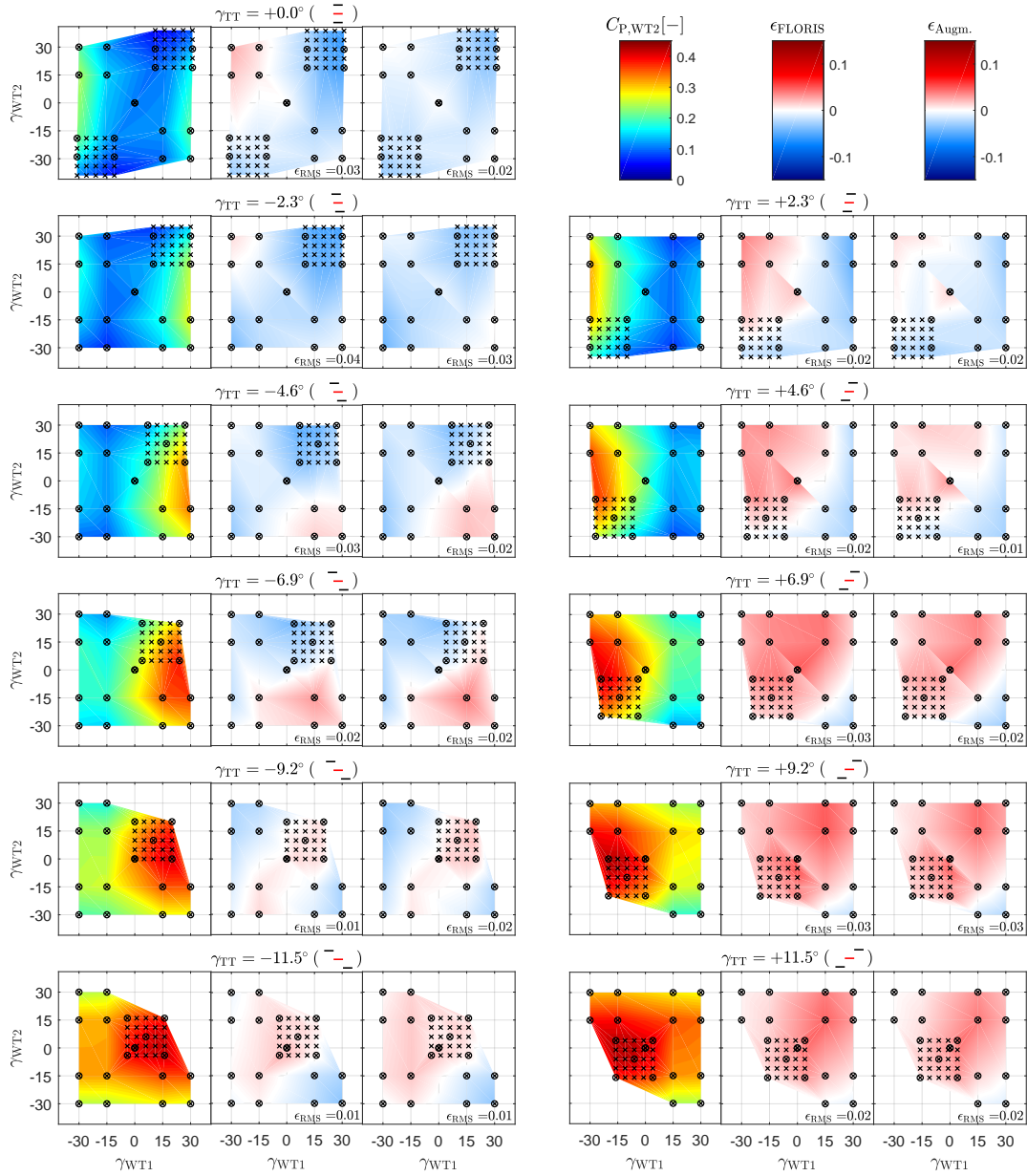
Figure A2. Wind turbine WT2. Each cluster of three subplots represents a unique turntable position, as indicated by the title and the wind farm layout sketch. Left subplot: turbine power coefficient $C_{P,WT2}$ as a function of γ_{WT1} (x-axis) and γ_{WT2} (y-axis). Middle subplot: FLORIS model error. Right subplot: augmented model error. Cross symbols: all measured configurations. Circles: conditions used for parameter identification.
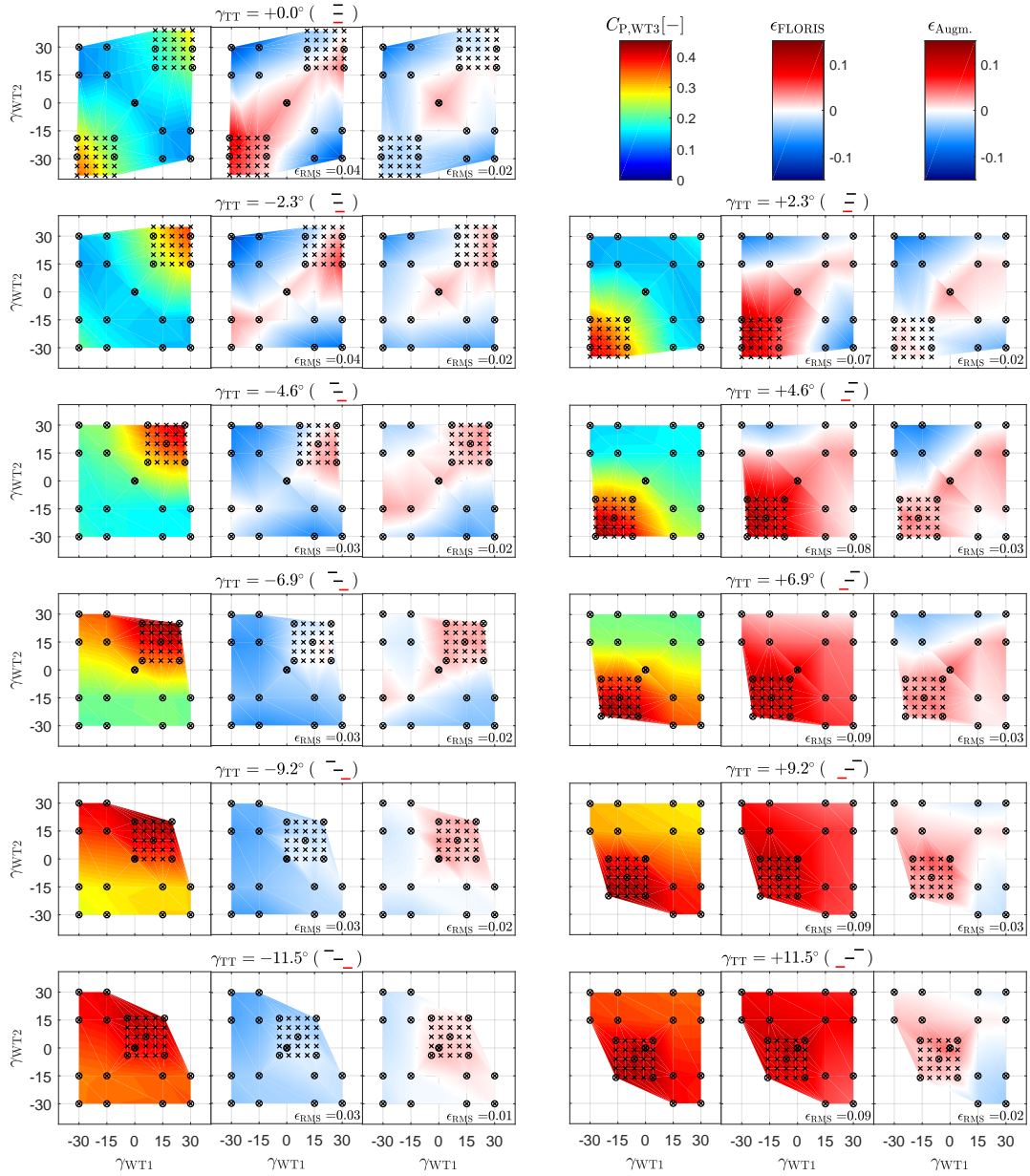
Figure A3. Wind turbine WT3. Each cluster of three subplots represents a unique turntable position, as indicated by the title and the wind farm layout sketch. Left subplot: turbine power coefficient $C_{P,WT3}$ as a function of γ_{WT1} (x-axis) and γ_{WT2} (y-axis). Middle subplot: FLORIS model error. Right subplot: augmented model error. Cross symbols: all measured configurations. Circles: conditions used for parameter identification.

ments. FC was responsible for the execution of the wind tunnel tests and the elaboration of the experimental results. All authors provided important input to this research work through discussions, feedback and by improving the manuscript.

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

Acknowledgements. The authors express their gratitude to Enel Green Power S.p.A., which granted access to the field data, and to Stefan
5   Kern of GE Renewable Energy for helping with the data post-processing. Special thanks go to Robin Weber of Technische Universität München and Stefano Cacciola, Alessandro Croce, Paolo Schito, and Alberto Zasso of Politecnico di Milano for their help in the conduction of the wind tunnel experiments. This work has been supported in part by the CL-WINDCON project, which receives funding from the European Union Horizon 2020 research and innovation program under grant agreement No. 727477. Further support has been provided by the CompactWind II project (FKZ: 0325492G), which receives funding from the German Federal Ministry for Economic Affairs and Energy
10   (BMWi).

References

Bastankhah, M. and Porté-Agel, F.: Experimental and theoretical study of wind turbine wakes in yawed conditions, Journal of Fluid Mechanics, 806, 506–541, 2016.

Bastine, D., Witha, B., Wächter, M., and Peinke, J.: POD Analysis of a Wind Turbine Wake in a Turbulent Atmospheric Boundary Layer, Journal of Physics: Conference Series, 524, 012 153, https://doi.org/10.1088/1742-6596/524/1/012153, 2014.

Bleeg, J., Purcell, M., Ruisi, R., and Traiger, E.: Wind Farm Blockage and the Consequences of Neglecting Its Impact on Energy Production, Energies, 11, 1609, 2018.

Boersma, S., Doekemeijer, B., Vali, M., Meyers, J., and van Wingerden, J.-W.: A control-oriented dynamic wind farm model: WFSim, Wind Energy Science Discussions, pp. 1–34, https://doi.org/10.5194/wes-2017-44, 2017.

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, https://doi.org/10.1016/j.jweia.2013.11.003, 2014a.

Bottasso, C. L., Campagnolo, F., and Petrović, V.: Wind tunnel testing of scaled wind turbine models: Beyond aerodynamics, Journal of Wind Engineering and Industrial Aerodynamics, 127, 11–28, https://doi.org/10.1016/j.jweia.2014.01.009, 2014b.

Breton, S.-P., Sumner, J., Sørensen, J. N., Hansen, K. S., Sarmast, S., and Ivanell, S.: A survey of modelling methods for high-fidelity wind farm simulations using large eddy simulation, Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 375, https://doi.org/10.1098/rsta.2016.0097, 2017.

Campagnolo, F., Petrović, V., Bottasso, C. L., and Croce, A.: Wind tunnel testing of wake control strategies, Proceedings of the American Control Conference (ACC), pp. 513–518, 2016a.

Campagnolo, F., Petrovic, V., Nanos, E. M., Tan, C. W., Bottasso, C. L., Paek, I., Kim, H., and Kim, K.: Wind tunnel testing of power maximization control strategies applied to a multi-turbine floating wind power platform, in: Proceedings of the The 26th International Ocean and Polar Engineering Conference, Rhodes, Greece, 2016b.

Campagnolo, F., Petrović, V., Schreiber, J., Nanos, E. M., Croce, A., and Bottasso, C. L.: Wind tunnel testing of a closed-loop wake deflection controller for wind farm power maximization, Journal of Physics: Conference Series, 753, 7, 2016c.

Campagnolo, F., Molder, A., Schreiber, J., and Bottasso, C. L.: Comparison of Analytical Wake Models with Wind Tunnel Data, Journal of Physics: Conference Series, 1256, 012 006, https://doi.org/10.1088/1742-6596/1256/1/012006, 2019.

Crespo, A. and Hernández, J.: Turbulence characteristics in wind-turbine wakes, Journal of Wind Engineering and Industrial Aerodynamics, 61, 71–85, https://doi.org/10.1016/0167-6105(95)00033-X, 1996.

Doekemeijer, B. M. and Storm, R.: FLORISSE M, https://github.com/TUDelft-DataDrivenControl/FLORISSE_M/tree/f0653ef3e8e56e284cdfa0a188aba0a7e42fce5c, accessed: 2019-07-30, 2018.

Doekemeijer, B. M. and Storm, R.: FLORISSE M, https://github.com/TUDelft-DataDrivenControl/FLORISSE_M/tree/ed2885107ca2600d0616f86cab3b3d59c5f0cb38, accessed: 2019-07-30, 2019.

Doekemeijer, B. M., Boersma, S., Pao, L. Y., and Van Wingerden, J. W.: Ensemble Kalman filtering for wind field estimation in wind farms, in: 2017 American Control Conference (ACC), pp. 19–24, IEEE, Piscataway, NJ, https://doi.org/10.23919/ACC.2017.7962924, 2017.

Doekemeijer, B. M., Fleming, P. A., and van Wingerden, J.-W.: A tutorial on the synthesis and validation of a closed-loop wind farm controller using a steady-state surrogate model, ArXiv, abs/1810.12066, 2018.

Dörenkämper, M., Witha, B., Steinfeld, G., Heinemann, D., and Kühn, M.: The impact of stable atmospheric boundary layers on wind-turbine wakes within offshore wind farms, Journal of Wind Engineering and Industrial Aerodynamics, 144, 146 – 153,

https://doi.org/10.1016/j.jweia.2014.12.011, selected papers from the 6th International Symposium on Computational Wind Engineering CWE 2014, 2015.

Fleming, P., Annoni, J., Shah, J. J., Wang, L., Ananthan, S., Zhang, Z., Hutchings, K., Wang, P., Chen, W., and Chen, L.: Field test of wake steering at an offshore wind farm, Wind Energy Science, 2, 229–239, https://doi.org/10.5194/wes-2-229-2017, 2017.

5 Fleming, P., Annoni, J., Churchfield, M., Martinez-Tossas, L. A., Gruchalla, K., Lawson, M., and Moriarty, P.: A simulation study demonstrating the importance of large-scale trailing vortices in wake steering, Wind Energy Science, 3, 243–255, https://doi.org/10.5194/wes-3-243-2018, 2018.

Fleming, P. A., Gebraad, P. M., Lee, S., van Wingerden, J.-W., Johnson, K., Churchfield, M., Michalakes, J., Spalart, P., and Moriarty, P.: Evaluating techniques for redirecting turbine wakes using SOWFA, Renewable Energy, 70, 211 – 218, https://doi.org/10.1016/j.renene.2014.02.015, special issue on aerodynamics of offshore wind energy systems and wakes, 2014.

Frandsen, S., Barthelmie, R., Pryor, S., Rathmann, O., Larsen, S., Højstrup, J., and Thøgersen, M.: Analytical modelling of wind speed deficit in large offshore wind farms, Wind energy, 9, 39–53, 2006.

Gebraad, P., Fleming, P. A., and Van Wingerden, J. W.: Comparison of actuation methods for wake control in wind plants, in: American Control Conference (ACC), 2015, pp. 1695–1701, IEEE, Piscataway, NJ, https://doi.org/10.1109/ACC.2015.7170977, 2015.

15 Gebraad, P. M. O., Teeuwisse, F. W., Van Wingerden, J. W., Fleming, P. A., Ruben, S. D., Marden, J. R., and Pao, L. Y.: A data-driven model for wind plant power optimization by yaw control, in: American Control Conference (ACC), 2014, pp. 3128–3134, IEEE, Piscataway, NJ, https://doi.org/10.1109/ACC.2014.6859118, 2014.

Göçmen, T. and Giebel, G.: Data-driven Wake Modelling for Reduced Uncertainties in short-term Possible Power Estimation, Journal of Physics: Conference Series, 1037, 072 002, https://doi.org/10.1088/1742-6596/1037/7/072002, 2018.

20 Golub, G. H. and van Loan, C. F.: Matrix computations, Johns Hopkins studies in mathematical sciences, Johns Hopkins Univ. Press, Baltimore, Md., 4. ed. edn., 2013.

Hansen, P. C.: The truncatedSVD as a method for regularization, BIT, 27, 534–553, https://doi.org/10.1007/BF01937276, 1987.

Jacobsen, H. S.: WAsP CFD: Wind model for complex terrain, https://www.wasp.dk/waspcfd#flow-model, accessed: 2019-08-01, 2019.

Jategaonkar, R. V.: Flight Vehicle System Identification, American Institute of Aeronautics and Astronautics, Reston, https://ebookcentral. proquest.com/lib/gbv/detail.action?docID=5056852, 2015.

25 Jonkman, J. and Jonkman, B.: FAST v8, https://nwtc.nrel.gov/FAST8, 2018.

Karampatziakis, N. and Langford, J.: Online Importance Weight Aware Updates, http://arxiv.org/pdf/1011.1576v4, 2011.

Martínez-Tossas, L. A., Annoni, J., Fleming, P. A., and Churchfield, M. J.: The aerodynamics of the curled wake: A simplified model in view of flow control, Wind Energy Science, 4, 127–138, https://doi.org/10.5194/wes-4-127-2019, 2019.

30 Mathworks: Matlab Documentation, fmincon, https://www.mathworks.com/help/releases/R2017b/optim/ug/fmincon.html, accessed: 2019-07-30, 2019.

Niayifar, A. and Porté-Agel, F.: A new analytical model for wind farm power prediction, Journal of Physics: Conference Series, 625, 012 039, https://doi.org/10.1088/1742-6596/625/1/012039, 2015.

Nocedal, J. and Wright, S. J.: Sequential Quadratic Programming, pp. 529–562, Springer New York, New York, NY, https://doi.org/10.1007/978-0-387-40065-5_18, 2006.

35 Schreiber, J., Nanos, E. M., Campagnolo, F., and Bottasso, C. L.: Verification and Calibration of a Reduced Order Wind Farm Model by Wind Tunnel Experiments, Journal of Physics: Conference Series, 854, 012 041, https://doi.org/10.1088/1742-6596/854/1/012041, 2017.

Schreiber, J., Salbert, B., and Bottasso, C. L.: Study of wind farm control potential based on SCADA data, Journal of Physics: Conference Series, 1037, 032 012, https://doi.org/10.1088/1742-6596/1037/3/032012, 2018.

Waiboer, R.: Dynamic modelling, identification and simulation of industrial robots: – for off-line programming of robotised laser welding –, Ph.D. thesis, University of Twente, Netherlands, 2007.

Wang, C., Wang, J., Campagnolo, F., Carraón, D. B., and Bottasso, C. L.: Validation of large-eddy simulation of scaled waked wind turbines in different yaw misalignment conditions, Journal of Physics: Conference Series, 1037, 062 007, https://doi.org/10.1088/1742-6596/1037/6/062007, 2018.

Xie, S. and Archer, C. L.: A Numerical Study of Wind-Turbine Wakes for Three Atmospheric Stability Conditions, Boundary-Layer Meteorology, 165, 87–112, https://doi.org/10.1007/s10546-017-0259-9, 2017.