Data-driven Wake Modelling for Reduced Uncertainties in short-term Possible Power Estimation

Paper

Göçmen, Tuhfe; Giebel, Gregor

Published in:
Journal of Physics: Conference Series

Link to article, DOI:
10.1088/1742-6596/1037/7/072002

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Data-driven Wake Modelling for Reduced Uncertainties in short-term Possible Power Estimation

To cite this article: Tuhfe Gökçmen and Gregor Giebel 2018 J. Phys.: Conf. Ser. 1037 072002

View the article online for updates and enhancements.

Related content
- Analysis of the reduced wake effect for available wind power calculation during curtailment
  M P C Bontekoning, S Sanchez Perez-Moreno, B C Ummels et al.
- Wind Speed Estimation and Parametrization of Wake Models for Downregulated Offshore Wind Farms within the scope of PossPOW Project
  Tuhfe Gökçmen Bozkurt, Gregor Giebel, Niels Kjølstad Poulsen et al.
- Wake behind an offshore wind farm observed with dual-Doppler radars
  Nicolai Gayle Nygaard and Alexander Christian Newcombe
Data-driven Wake Modelling for Reduced Uncertainties in short-term Possible Power Estimation

Tuhfe Göçmen, Gregor Giebel
Technical University of Denmark, DTU Wind Energy, Risø Campus, Roskilde, Denmark
E-mail: tuhf@dtu.dk

Abstract. One of the ancillary services the wind farms are required to provide to the system operators is reserve power, which is achieved by down-regulating the wind farm from its possible power. In order to estimate the reserves, the possible power needs to be calculated by correcting the reduced wake effects behind the down-regulated turbines. The most recent grid codes dictate the quality of the possible power at the wind farm level to be assessed within 1-min intervals for offshore wind power plants. Therefore, the necessity of a fast and reliable wake model is more prominent than ever.

Here we investigate the performance of two engineering wake models with 1-sec resolution SCADA data on three different offshore wind farms, given the quantified input uncertainty. The preliminary results show that, even wind farm specific training of the model parameters might fail to comply with the strict criteria stated in the grid codes, especially for the layouts with significant wake losses. In order to tackle the inadequacy of the engineering wake models to capture some of the dynamics in the wind farm flow due to the embedded assumptions, purely data-driven techniques are evaluated. The flexibility of such an on-line model enables 'site-turbine-time-specific' modelling, in which the parameters are defined per turbine and updated with each time-step in a specific wind farm.

1. Introduction & Motivation
For a long time, it has been known to the wind energy community that wakes within a wind farm induce significant losses in power production and a substantial increase in turbine loading at downstream locations. The latest developments in wind farm operation strategies and the requirements enforced by the system operators in Europe dictates fast data transfer, both at the turbine and the farm level. Accordingly, wind farms are expected to provide information regarding their power production in time scales much shorter than a year and conventional AEP (annual energy production) estimates. Even the usual 10 minutes averaging period is too long compared to the recent grid requirements in Germany, [1]. Pilot phase took effect in October 2017, where the possible (sometimes also called available) power is to be calculated for 60-seconds intervals for down-regulated wind farms. The standard deviation of the percentage error of the wind farm scale possible power signal for the pilot phase of the German grid code is required to be less than ±5%. This is the strictest criteria demanded by any system operators. The quality assessment of the possible power for offshore wind farms is well documented by the system operators around Europe. [25]. The longest time averaging interval for the required
possible power signal is 15 minutes only. The implied standards are difficult to achieve and are subject to penalty if not met. Since the key challenge in possible power estimation is to correct the reduced wake effects behind the curtailed turbines, there is a clear need for fast and accurate wake models with low levels of uncertainty embedded in the resulting power estimations.

Additionally, in order to mitigate the existing wake effects within the wind farm (either by induction, e.g. [19], or steering of the wake, e.g. [9]) various studies are performed within the area of aerodynamic wind farm control. Most of those model based optimum wind farm controllers are built upon low to medium fidelity wake models due to their computationally low-cost estimate of the flow within the wind farm. Those low-cost wake models are typically parameterised and calibrated using a high-fidelity (high-cost) flow model [4, 10, 23, 24] with quasi-steady inflow, where the uncertainty in the fitted parameters is left unclear. However, the accuracy of the short-term power estimations and the manifestation of their uncertainties within the control scenarios, especially for dynamic wind farm control, is one of the biggest research questions in the field. Here we aim to also provide an insight to the levels of ambiguity that is to be dealt with in the wind farm operations and the control scenarios.

In an attempt to address the need and the challenge to estimate short-term possible power with the desired accuracy, two of the existing and widely applied wake models are implemented to 1Hz SCADA signals (Supervisory Control And Data Acquisition systems) from Thanet, Horns Rev I and Liliggrund offshore wind farms. The first engineering, semi-empirical wake model investigated is the Larsen model [22, 21]. The Larsen model is then re-parameterised for high frequency data fit, and in total has six parameters to calibrate [16, 14]. The second low-cost model is the recently introduced analytical Gaussian deficit model [7, 3, 8] with a single parameter to calibrate. Lastly, the high frequency data from the upstream turbine(s) are fed into a large-scale machine learning platform, TensorFlow [2]. In order to represent the noisy pattern in the high frequency SCADA data and the time lag between the upstream and the downstream turbines, Long Short-Term Memory (LSTM) [17], which is a special building unit for Recurrent Neural Networks (RNNs), is implemented. A brief description of the investigated wind farms and the extracted signals is provided in Section 2. The model structures and the calibration procedures, together with the training and the validation of the dynamic neural networks, are elaborated in Section 3.

![Thanet Offshore Wind Farm Layout & Vestas V90 Power and Thrust Curve](image.png)
2. Offshore Wind Farm Data

Especially for large offshore wind farms, the turbine data, also referred as SCADA, are the most available and representative of the local flow and operational characteristics. Here we will describe a possible power estimation framework that re-calibrates itself using the most recent SCADA information at the turbine locations. Both for the training of the wake models and for the validation of their re-calibration, the 1-sec SCADA data from Thanet, Horns Rev I and Lillgrund wind farms are analysed. Due to the high frequency of the extracted dataset and the short-term focus of the study, the length of both the calibration and the validation data is limited to hours. In addition, the pre-processing has its main emphasis on continuous time series during perpendicular winds along the rows of turbines (i.e. (perpendicular upstream wind directions as indicated in Figures 1 & 2 & 3 ) ±15°).

![Horns Rev-I Offshore Wind Farm Layout & Vestas V80 Power and Thrust Curve](image1)

**Figure 2:** Horns Rev-I Offshore Wind Farm Layout & Vestas V80 Power and Thrust Curve

![Lillgrund Offshore Wind Farm Layout & Siemens 2.3MW Power and Thrust Curve](image2)

**Figure 3:** Lillgrund Offshore Wind Farm Layout & Siemens 2.3MW Power and Thrust Curve
The Thanet wind farm is located in the east of the UK and consists of 100 Vestas V90-3MW turbines [6]. The spacing between the turbines for a perpendicular wind is $5.3D$ (rotor diameters), Figure 1. The Horns Rev I wind farm is located in the west of Denmark and it originally consists of 80 Vestas V80-2MW turbines [5]. It is one of the most studied offshore wind farms with a regular layout aligned east-west, with a grid spacing of $7D$, Figure 2. The last investigated wind farm is Lillgrund, which is located east of Denmark, west of Sweden. It has 48 Siemens 2.3MW turbines with $3.3D$ and $4.3D$ spacing, which causes very significant wake losses and makes it a perfect case to test the performance of the wake models.

Since the main focus is to evaluate the performance of the investigated low-cost models for short-term power estimations, the additional uncertainties due to summation of the wakes and the meandering effects are avoided here in this study. Therefore, the single wake cases on turbines “Th-SW”, “HR-SW” and “Lill-SW” in Figures 1, 2, and 3 are investigated using 1-min averaged percentage error distributions over 1-hour full wake performances.

3. Methodology, Model Implementation & Results

The wake models within the framework are trained for individual turbines among the wind farm using the active (output) power, pitch angle $\theta$, rotational speed $\omega$, temperature and wind direction signals from single turbines’ SCADA systems. Since the uncertainty is the main concern of the short-term power estimation, the nacelle mounted anemometers are not considered as an input. Instead, the rotor effective wind speed approach introduced in [15] and validated using the same three wind farms [15, 12] is implemented as in Figure 4. At every second, the effective wind speed, $U_{eff}$ is calculated iteratively using the power output of the turbine(s) together with the power coefficient, $C_p(\omega, \theta, U_{eff})$, under the instantaneous pitch and rotational speed configuration. The temperature and pressure are also extracted from the SCADA at every turbine to correct the air density. At this point, it should be pointed out that the correction would be improved where/if the humidity information is also available. However for that study, it was not one of the extracted signals from the investigated SCADA systems, hence not included.

Figure 4: Effective wind speed estimation and (representative) input SCADA signals

For as the approach is seen to perform well under down-regulation also (where the output power is set to a certain value, see a representative signal in Figure 4), the effective wind speed at the downstream turbines is considered as the output of the training and validation. In other words, the models are using the effective wind speed at the upstream turbines as inputs (together with the individual wind direction signals) where the desired output is the downstream effective wind speed. As emphasised earlier, both the input and the propagated uncertainty are under the scope of this study. Therefore the quantified uncertainty of the rotor effective wind speed, $\sigma_{U_{eff}} = 0.3 \text{ m/s below rated}$ [13] (for the investigated turbines), is taken into account as input and output uncertainty for both the validation and re-calibration processes.
It is important to note that the trained parameters in the wake model are to be specific for a particular turbine placed in a particular wind farm at a particular time interval. That unprecedented approach which learns from local data (both spatially and temporally) enables to introduce the effects of the local flow characteristics such as atmospheric stability, wake meandering, turbulence intensity, etc. into the wake models properly, although no direct measurement is available. Hence a substantial increase in the accuracy of the wake model is anticipated with a considerable reduction in the uncertainty. Here, we present the improvement in the model performance when trained for the same wind farm historical data as the validation case. The uncertainty in the model results is analysed via 1-min averaged percentage error distribution (as required in the strictest grid code in Germany [1]), defined as:

$$\%\text{error}_{1\text{min}} = \frac{(U_{\text{model wake}}_{1\text{min}}) - (U_{\text{eff wake}}_{1\text{min}})}{(U_{\text{eff wake}}_{1\text{min}})} \times 100$$ (1)

To calculate the error, the second-wise effective wind speed and the wake model results are averaged over 1-minute, i.e. 60 samples per mean value. The model evaluation is performed using 1-sec dataset corrected for the time delay between the upstream and the downstream turbines, where the time delay is approximated using the correlation in the local effective wind speed. The percentage error distributions are presented in boxplot where the boxes go from the first quartile ($-0.6745\sigma$ around the mean) to the third ($0.6745\sigma$ around the mean) and the middle line is the median. Representation of the median instead of the mean is simply a better measure of the error, given that the distributions are not necessarily Gaussian. The whiskers in the boxplots follow Tukey’s descriptive statistics [26] where the boundaries are the lowest and the highest datum within the 1.5 inter-quartile range, IQR, corresponding to $\pm 2.7\sigma$ around the mean. Note that the strictest grid code in Germany requires the distribution to have $\sigma = \pm 5\%$ around the mean. Accordingly, the width of the IQR or the length of the boxes for the error distributions are required to be within $\pm 3.375\%$ throughout this study.

3.1. Larsen Model
The first engineering wake model to be re-calibrated at the turbine locations using 1-sec SCADA is the Larsen model [22, 21]. Previously, within the PossPOW project [16, 14], it has been re-calibrated using nonlinear LSE (Least Squares Estimation) for Thanet data. However, here we focus on further training the model using Bayesian calibration (BC) where the previously fitted parameter space is the prior distribution. Note that, the aim is to update the parameters of the wake model using the previous experience. In the process, the input uncertainty $\sigma_{U_{\text{eff}}}$ and the standard deviation in the prior parameter distribution is also taken into account.

Starting from the LSE fit in Thanet, the Larsen model is re-calibrated with the BC using the wind farms’ own data, which is carefully separated from the validation dataset. Both the calibration and the validation dataset is filtered for the perpendicular winds only and they cover 1 hour period of 1 Hz data. It can be clearly seen from Figure 5 that the wind farm specific training of the model decreases both the median and the standard deviation of the error under the German grid code limits; although the distribution is consistently skewed towards under-estimation of the wake. In order to test the hypothesis further, the third wind farm, Lillgrund with closely spaced turbines, is investigated. Having the original LSE fit updated with BC in Horns Rev-I, now the latter is to be updated further using 1-sec SCADA from Siemens 2.3MW turbines. Similar to Figure 5, the comparison of the 1-min percentage error between the two sets of BC parameters (trained in Horns Rev I and Lillgrund dataset) tested for a perpendicular wind in Lillgrund (for 4.3D spacing). The results in Figure 6 indicate a minor mitigation in the percentage error where the underestimation of the wake deficit (overestimation of the power production at the downstream turbine, positive error) is prominent.
3.2. Gaussian-deficit Model

The second engineering model in the re-calibration framework is the Gaussian deficit model introduced by Bastankhah et al. [7]. For a single wake, the model assumes the normalised wake deficit to follow a self-similar, axisymmetric Gaussian distribution where the width of the distribution expands linearly with downstream distance. The slope of this linear expansion, wake growth rate ($k^*$), is then specified to be a function of the local turbulence intensity (TI) at the turbine locations [24]. The parameters to define the linear relation between the local TI and the wake growth rate are calibrated using large eddy simulation (LES) of Vestas V80 turbine for below rated conditions and reported as

$$k^* = a \, TI + b$$

where $a = 0.3837$ and $b = 0.003678$. $TI$ at the turbine locations are defined using moving average with 10-min window on 1-sec effective wind speed, $U_{eff}$ as exemplified in [12], as it is the only measure to quantify turbulence using only SCADA in short time intervals. Consequently, the wake growth rate $k^*$ becomes time dependent and the expansion of the wake is linear only during 1-sec snapshots. The implementation of the model with the LES-fit parameters using 1Hz SCADA from the investigated wind farms are presented in Figure 7.

Compared to Figures 5 and 6, the error distribution is much less skewed and, similar to the modelled deficit itself, follows a Gaussian distribution. For Thanet, although the mean error is as low as 0.6%, the quartiles are -6% and 10% which is still broader than desired. In Horns Rev-I, for which the parameters are tuned for, the width of the error distribution is narrower but still higher than the requirements, and it is skewed more than 5% towards under-estimation of the wake. Lastly for Lillgrund, the model significantly over-estimates the wake deficit at 4.3D downstream, where both the mean and the variation of the error are quite high. It is very clear that there is a need for improvement in the model performance. With the same approach as in Larsen model, a new set of parameters can be defined per wind farm basis. Since the parameter distribution of the original LES-fit for Equation 2 is not defined, the prior distributions of $a$ and
b are assumed to be normal with the reported mean value and a standard deviation of 5% around the mean. The improvement in the resulting error distribution is presented in Figure 8. The resulting wind farm specific parameters for \( a \) and \( b \) are listed in Table 1. For all the wind farms the mean error is significantly reduced and the distribution is much narrower. The German grid code requirements are easily met in Horns Rev-I; whereas for closer turbine spacing in Thanet and Lillgrund, although the mitigation compared to Figure 7 is clear, the model performance is still far from desired.

\[
\begin{array}{ll}
\text{Thanet} & 0.421 \quad 0.0025 \\
\text{Horns Rev} & 0.223 \quad 0.0039 \\
\text{Lillgrund} & 2.42 \quad 0.0042
\end{array}
\]

**Table 1:** Mean values of the wind farm specific parameter fit for the Gaussian Deficit Model using Bayesian calibration

For both of the engineering/analytical low-cost models, especially the performance of the newly-trained parameters in Lillgrund indicate the need for more “accurate” sensors/data, more intelligent methods to analyse the data, and different perspectives on how to include the physical complexity of the wind farm flow into the wake models.

### 3.3. Deep-learning via LSTM

So far, it has been seen that the performance of computationally affordable physical models, although calibrated specifically for the wind farm, is less than targeted. Therefore, using the same pre-processed data and the included uncertainties as the re-calibration procedure, a purely data-driven approach is generated. Although the machine-learning techniques have been one of the most popular research topics and implemented to numerous fields, their application in wind farm flow and wake modelling has been rather limited. The major contribution to such implementations has been performed by the wind farm control community. One of the earliest works in the field with applied system identification principles is presented in [20]. Similar to the calibration of the simplified physical models, a data-driven parametric model specifically designed for wake steering control is introduced in [11]; where the parameters are trained using steady flow simulation results of LES. As far as the uncertainties and non-optimum conditions are concerned, [27] presents a surrogate model for the operational power curve to quantify and isolate the turbulence induced uncertainties. Comparable to the focus of this study, in [18], 3
individual wind speed and direction cases are considered in Horns Rev-I to train and validate the results of artificial neural network (ANN), support vector machine (SVM), and K-nearest neighbours (KNNs). In the quasi-steady pre-processing of the wind farm data, the turbulent effects are not included.

Here in this study, the state-of-the-art machine learning platform TensorFlow [2] is used to implement LSTM algorithm [17]. Given the sampling rate and the uncertainties included in our wind farm data, LSTM (thereby recurrent neural networks) is applied as our deep-learning algorithm. As stated earlier, LSTM is shown to perform much faster and better for highly fluctuating time series. The network is trained using the upstream turbine information for the historical data of previous 1-hour to predict the upcoming 1-min. Then the training data is shifted for 1-min to include the actual output of the first prediction, still consists of 1-hour data. Therefore, the model is constantly updated and by the end of the prediction period (the second hour) the training is stopped, providing a total of 1-hour estimation with 1Hz frequency. Consequently, a new model is created based on exactly 1-hour earlier at every minute per turbine. Having a fast machine learning platform enables such a network generation to be feasible.

![Figure 9: Performance of the LSTM network trained using 1-hour historical data from upstream turbines, updated at every 1-min](image)

For the studied single wake cases, the upstream turbines are defined at every minute depending on the incoming wind direction. Together with the wind direction, the effective wind speed, standard deviation of the effective wind speed (as turbulence indicator) and the time dependent thrust coefficient (especially for the down-regulated turbines) of the upstream turbines are fed as inputs to the network, together with their associated uncertainties. The output layer of the network is the effective wind speed at the downstream turbine. The percentage error of the 1-min averaged outputs of the LSTM network is presented in Figure 9.

It is clearly seen that, LSTM RNNs trained using 1-hour historical data from the upstream turbines (with 1-min forward moving sampling) easily comply with the strictest grid code requirements for all three of the wind farms. Generating essentially a new model per turbine at every minute enables to include all the local dynamics, in terms both of time and space. On the other hand, the detail required in training an adequate network also shows how high the demanded accuracy of the available power is.

4. Conclusion

Due to the the tightening demand from the system operators in terms of balancing services, the wind farm operators are required to provide short term (down to 1-min time averaging intervals) and highly accurate (maximum std of the percentage error up to 5%) estimation of possible (or available) power. In order to achieve that, the accuracy and the levels of uncertainty of the fast engineering models needs to be improved. Since the relevant parties are the operating wind farms, the turbine historical data can be exploited to enable such an improvement.

Here we describe first a re-calibration framework that updates the parameters in Larsen and Gaussian Deficit wake models using wind farm specific 1-sec historical data. The results of the recalibrated models from Thanet, Horns Rev-I and Lillgrund offshore wind farms are presented in
terms of 1-min percentage error distributions over a 1-hour time interval. Although a significant mitigation is observed, it is seen that the reduced uncertainty in the fitted parameters are not adequately manifested in the consecutive error spread. Especially for Lilgrund wind farm, where the turbines are very closely located, the engineering model assumptions seem to have failed and risk to comply with the accuracy requirements. It should also be noted that the input uncertainty to the wake models plays an equally important role in the error distribution. Accordingly, a more accurate sensor and data acquisition systems need to be installed to the operating wind farms if such strict criteria are to be fulfilled.

Lastly, a deep learning algorithm using LSTM RNN is implemented via the recently developed platform TensorFlow. The neural network is trained using the wind direction, effective wind speed, standard deviation of the wind speed and the thrust coefficient signals of the upstream turbines for 1-hour historical data (3600 samples in total). The trained model is set to predict the wind speed at the downstream turbine for the upcoming 1-min. The procedure is repeated for every minute where the training interval is continuously shifted to ensure the network to stay updated. By doing so essentially a new model is created at every minute per turbine, where the increasing flexibility and speed of the machine-learning algorithms make such configurations possible and computationally feasible. It is seen that the performance of the updated networks surpasses the physical models by far, even with wind farm specific parameters; and easily meets the requirements. Therefore, the results do indicate that the data-science approaches have the potential to be the new generation of wake modelling, at least for the short-term power estimation of the operating wind farms.

Acknowledgments
This work is a part of Concert and PossPOW projects funded by Energinet.dk under the Public Service Obligation (PSO) 2016-1-12396 and 2012-1-10763, respectively. The authors would like to thank to Dr. Juan Pablo Murcia Leon from Vestas Wind Systems A/S, Denmark for his contributions and Jesper Runge Kristoffersen from Vattenfall Renewables Wind DK A/S, Denmark for the provision of SCADA and his supervision.

References


