Wind farm flow reconstruction and prediction from high frequency SCADA Data

Andreas Rott\textsuperscript{1}, Vlaho Petrović\textsuperscript{1}, Martin Kühn\textsuperscript{1}

\textsuperscript{1}ForWind – University of Oldenburg, Institute of Physics, Küpkersweg 70, 26129, Oldenburg, Germany
E-mail: andreas.rott@uni-oldenburg.de

Abstract. The flow inside a typical large wind farm propagates through the array from turbine to turbine. We present an algorithm that intuitively processes measured values from as much as possible turbines in real-time and uses them to determine a flow reconstruction and minute-scale forecast for the downstream turbines. For the validation we used full-field measurements from the offshore wind farm Global Tech I. The flow reconstruction is compared to the measurements of one turbine, which was excluded from the algorithm and achieved a root mean square error of 0.55 ms\textsuperscript{-1} for the wind speed estimation. The flow forecasting is tested for three prediction horizons 30 s, 60 s and 120 s. Together with automated error correction to account for calibration errors and wake effects, the flow prediction achieves a root mean square error of 0.52 ms\textsuperscript{-1} for the 120 s-forecast of the wind speed, which beats the persistence forecasting method. The reconstruction allows to analyse the flow in the wind farm, to detect abnormal turbine behaviour and to estimate fatigue loads, and the minute-scale forecasting is a useful tool for predictive wind farm control and estimating the available power of a wind farm, which becomes more and more necessary for grid stability.

1. Introduction
Wind energy is one of the central drivers of the energy transition towards renewable energy. In the last 10 years (2009 to 2019), the share of net energy generation by wind turbines in Germany has more than tripled (from 7.7\% to 24.6\%) \cite{1}. Especially in the offshore sector, the installed capacity has increased strongly. As a consequence, both wind turbines and wind parks are becoming larger in scale and more frequent. Each turbine is equipped with meteorological measuring instruments and constantly monitors the wind. Conventionally, the meteorological measurements are filtered and used to control the respective turbine and only simple statistics are recorded in a database, since the individual measurements are subject to strong disturbances caused by the rotor-induced turbulence. On the other hand, the high temporal resolution and the typically uniform spatial distribution over a large area in the wind farm are good prerequisites for flow reconstruction and very short-term forecasts. Usually statistical time series models are used for very short term predictions (see \cite{2}). More recently, the use of remote sensing systems such as radar and lidar for predictions have been researched (e.g. \cite{3}). This study will investigate to what extent a reliable flow reconstruction and flow forecast can be implemented in a real-time environment based solely on turbine measurements without further sensors or external information. This is similar to the investigation of Bossanyi \cite{4}, who proposed to use the measurements of the adjacent turbines to optimize the yaw control. However, in the present...
study the measurements from the Global Tech I wind farm are used instead of simulations and the difficulty of using real measurement data and the influence of different turbine behaviour, operating conditions and measurement inaccuracies is discussed.

A flow reconstruction allows determining the conditions the turbines are exposed to, e.g. to better estimate loads and to detect abnormal behaviour [5].

Wind forecasting becomes more and more important for maintaining grid stability as the share of wind energy in the grid increases [6], but it can also be used for predictive turbine and wind farm control [7, 8].

It is to be considered that the measurements are recorded on the nacelle and are therefore exposed to the influence of the rotor. Furthermore, the common nacelle anemometry is typically not calibrated individually. More sophisticated devices, like e.g. spinner anemometer [9], are hardly used, yet. Nevertheless, the collective measurements of several turbines can be used to reduce uncertainties for robust predictions.

This contribution aims to present an intuitive and easy-to-implement method that uses wind measurements from turbines and creates a flow field in real-time, which can be further propagated to make wind speed and wind direction predictions on downstream turbines. The general approach of this investigation is similar to the one in [10], but here a simple Lagrangian advection scheme is used and no complex Kalman filters and the algorithm is designed to be fast and omnidirectional to take wind direction changes into account. The method is demonstrated and validated by using real data from the offshore wind farm Global Tech I. Finally we are introducing an automated error correction to reduce deterministic deviations due to calibration differences and wake effects.

2. Methodology
In this manuscript, we introduce a Lagrangian advection scheme (Section 2.1) in a spatio-temporal interpolation framework (Section 2.2) to reconstruct and predict the flow inside a wind farm from turbine measurements. The method is validated using full-field wind farm measurements (Section 2.3).

2.1. Lagrangian Advection Scheme
The flow reconstruction and prediction are based on a Lagrangian advection scheme, which is a simplified solver of the advection term of the Navier-Stokes equation for incompressible flow when the external forces and viscosity are neglected as given in eq. (1):

$$\frac{\partial \mathbf{u}}{\partial t} = - (\nabla \cdot \mathbf{u}) \mathbf{u} = - \left( u_x \frac{\partial u}{\partial x} + u_y \frac{\partial u}{\partial y} + u_z \frac{\partial u}{\partial z} \right)$$

where \( \mathbf{u} = (u_x, u_y, u_z)^T \) is the flow velocity vector. The Lagrangian advection scheme treats every wind speed measurement as an individual parcel, which moves freely at its own velocity, without consideration of parcel collision. Since the available wind measurements consist only of horizontal wind speed and wind direction and no information about the elevation angle, the parcels move in the 2D-plane at hub height and we assume \( u_z = \theta \). The location \((x_i, y_i)\) of the \(i\)-th parcel is calculated according to eq. (2).

$$ \begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + t_i \cdot \begin{pmatrix} u_{x,i} \\ u_{y,i} \end{pmatrix} $$

Where \( t_i \) is the ‘age’ of the \(i\)-th parcel, \( u_{x,i} \) and \( u_{y,i} \) are its velocity components and \((x_0, y_0)^T\) is the origin of the parcel i.e. the location of the respective turbine. Figure 1 illustrates the advection scheme for the measurements of one turbine. The turbine is depicted by the
Figure 1. Schematic illustration of the measurements as parcels.

red marker, while the parcels are the green and blue dots. The blue dots depict parcels of measurements which were recorded at the turbine, they are propagated with their own wind velocity originating at the turbine. The green dots represent measurements, which will be measured by the same turbine in the future, i.e. they have a negative 'age'. These measurements are only available for the reconstruction in post-processing, where we can use the propagation also backwards in time to improve the accuracy.

2.2. Spatio-Temporal interpolation

To utilize the parcels for reconstructing or forecasting the flow, they are used in a spatio-temporal interpolation framework. First, this interpolation should take into account the spatial distance of the parcel to a target point for which we want to estimate the flow situation. Second, the age of the parcel should also be considered to reflect the increasing uncertainty of the assumptions of the Lagrangian advection in the weighting. All this is expressed in the following equation, which corresponds to the inverse distance weighting, also called Shepard weighting, where in our case the distance is a spatio-temporal distance. Let $i$ be the index of the $i$-th parcel, then the spatio-temporal weighting for the velocity information of this parcel for one evaluation point in space-time $(x, y, t)$ is defined as follows:

$$
\hat{w}_{i,\nu,p}(x, y, t) = \frac{w_{i,\nu,p}(x, y, t)}{\sum_i w_{i,\nu,p}(x, y, t)} \quad (3)
$$

where $\nu$ is a constant to scale the weighting of the temporal distance, i.e. age of the parcel, in accordance to the spatial distance, $p$ is a tuning parameter for the Shepard interpolation, $x_i, y_i$ are the spatial coordinates of the parcel and $t_i$ is the time of origin of the $i$-th parcel. For the theoretical case, that the spatio-temporal distance approaches 0 the weighting-factor $w_{i,\nu,p}(x, y, t)$ would approach $\infty$ for this parcel. The total weight of the respective parcel $\hat{w}_{i,\nu,p}(x, y, t)$ would be 1 and the weights for all other parcels 0, which means that the Shepard interpolation is a true interpolation at the supporting points. In our case, this does not occur as we will explain in Section 3.

The estimation of the wind speed components for the evaluation point $(x, y, t)$ is defined by:
\[ u_x := u_{x,\nu,p}(x, y, t) = \sum_{\forall t_i \leq t_{\text{max}}} \hat{w}_{i,\nu,p}(x, y, t) \cdot u_{x,i} \]  
\[ u_y := u_{y,\nu,p}(x, y, t) = \sum_{\forall t_i \leq t_{\text{max}}} \hat{w}_{i,\nu,p}(x, y, t) \cdot u_{y,i} \]

where \( u_{x,i} \) and \( u_{y,i} \) are the measured velocity components. \( t_{\text{max}} \) is a parameter that restricts the maximum absolute age of the parcels. Commonly at each turbine, wind direction \( \varphi \) and wind speed \( U \) are measured. To get the velocity components, we transform these quantities from the polar representation to the Cartesian with 
\[ u_x = -\sin(\varphi_i) \cdot U_i \]
\[ u_y = -\cos(\varphi_i) \cdot U_i \]
In total, we now can calculate the estimated wind speed and wind direction in the polar representation:
\[ U_{\text{est}} = \|(u_x, u_y)\| = \sqrt{u_x^2 + u_y^2}, \quad \varphi_{\text{est}} = \text{atan2}(u_x, u_y) \]

In fact, this interpolation can be used as an extrapolation, because the evaluation points are not restricted to be in between the input values. The Shepard weighting has the general property that the results converges to the mean value of all input values the further away the evaluation point is from the inputs (measured values).

To make the temporal weighting easier to understand, we can look at Figure 1 again and imagine that the parcels move away from the 2D surface in the direction of the \( z \)-axis with increasing age, where the slope is parameterized by \( \nu \). The distance to the evaluation point on the surface determines the weighting factor of the respective parcel.

### 2.3. Full-field Data

For the validation of the presented method, we are using full-field data from the offshore wind farm Global Tech I. Global Tech I is located in the German North Sea with more than 100 km distance to the coast. The wind farm consists of 80 turbines of the type AD 5-116 from Adwen and spans over an area of 41 square kilometres. Neighbouring turbines have a distance of roughly 700 m to 900 m to each other. The layout of the wind farm is illustrated in Figure 2. The turbine T60 was chosen for the evaluation of our method and is marked in red on the figure. From wind turbines producing power, only the rotor-equivalent wind speed was available instead of the measurements by the turbine anemometer. Such information was reconstructed based on the pitch angles of the blades, the generator speed and torque, and it has significantly lower fluctuations compared to the point measurements by the anemometer. Only when a
turbine was not producing power, the measurements from the anemometer were provided. The wind direction measurement, on the other hand, is the sum of the turbine alignment and the relative wind direction measurement from the 2D Sonic Anemometer. The Sonic Anemometer is mounted on the nacelle directly behind the rotor, which significantly influences the accuracy of the measurement and introduces strong fluctuations. The turbine alignment is not standardized for all turbines so that each turbine has its own north. Therefore, in a first step, an attempt is made to estimate the alignment deviation. Fortunately, the alignment of some turbines could be checked more precisely by means of additional sensors. These turbines were equipped with iSpin devices from the company ROMO Wind AG that used a GPS to estimate the alignment and additionally with a Windfit box of the company Sereema which uses a magnetic compass. Together with the assumption that the average wind direction over a day should be the same for all turbines, the other turbine alignments were corrected.

3. Results

In this part, we are evaluating the quality of the reconstruction (Section 3.1) and prediction (Section 3.2) of the flow from the turbine measurements. For the evaluation wind speed and wind direction measurements from all turbines over one month (April 2019) in a sampling rate of 1 Hz were available. We are discussing error sources and propose an automated error correction (Section 3.3) to further improve the results for the prediction and compare the prediction to the persistence forecast (Section 3.4). We have chosen the turbine T60 (see Figure 2) as our target for the reconstruction and prediction evaluation.

3.1. Flow Reconstruction

The sampling rate of the flow reconstruction was set to 0.2 Hz, which was interpolated to the full sampling rate of 1 Hz. To evaluate the flow reconstruction at the location of the turbine T60, we are excluding the measurements of T60 from the algorithm and reconstruct the flow at the T60 from the measurements of the remaining turbines. Consequently, the spatio-temporal distance can not be 0 at the location of T60. Finally, we compare the estimated wind speed and wind direction to the measurements of T60.

For the flow reconstruction, we consider measurements with an absolute age of $|t_i| \leq t_{\text{max}} := 180 \text{s}$. We set the spatio-temporal scaling constant (see eq. (4)) to $\nu = 1 \frac{\text{m}}{\text{s}}$ and the Shepard interpolation parameter to $p = 2$. The Shepard parameter controls how strong the weighting factors of the interpolation are localized. A smaller parameter distributes the weights more evenly to the parcels resulting in a smoother estimate, while a larger value focuses the weights on the closest parcels.

Figure 3 shows two small excerpts of the reconstructed wind speed for $p = 2$ and $p = 8$ compared to the target wind speed measured at turbine T60. The left graphs shows four-hour time series of a rather common situation with high fluctuations. The flow reconstructions fits well with the original data and it is clearly visible, that the lower Shepard parameter leads to a more averaged flow estimation, where a lot of the fluctuations are smoothed out. The right graph shows one-hour time series for a situation we did not encounter often in the measurements but serves as a good demonstration of the reconstruction. In the first 30 minutes, strong ramps appear and the turbines change from idling to production. In the second half-hour, the wind speed follows sinusoidal variations. The origin of these rather uncommon speed variation is still unknown, but it could be observed at all turbines and also by a lidar device that scanned the flow in front of the wind farm. The reconstruction follows the target values, but the sinusoidal fluctuations are damped significantly for the Shepard parameter of $p = 2$. We can see the effect of this parameter in Figure 3 for $p = 2$ and $p = 8$. The smaller value achieved more robust results in general, hence we chose $p = 2$ for the quantitative evaluation that follows.
Figure 3. (left) Four-hour time series with high turbulence intensity. (right) One-hour time series with a strong wind ramp and sinusoidal fluctuations. In both figures is the wind speed measurement at T60 represented by the blue graph and flow reconstruction is represented by the orange graph for $p = 8$ and the green graph for $p = 2$.

Figure 4. (left) Histogram of the wind speed error of the flow reconstruction. (right) Histogram of the wind direction error of the flow reconstruction.

We investigated the statistical reconstruction error of wind speed and wind direction for the whole data set (April 2019). Figure 4 is a histogram of the deviations and the statistics are summarized in Table 1. In the table, the mean error, the standard deviation of the error, the 25% and 75% quartiles and the median (50%) are listed. Finally, the root mean square error (RMSE) is given at the bottom of the table.

The results indicate an overall good agreement of the wind speed reconstruction. 50% of the errors fall in between an interval of $[-0.27 \text{ m/s}, 0.43 \text{ m/s}]$ and RMSE is $0.55 \text{ m/s}$. The error of the wind direction, however, is relatively high with a RMSE of $7.02^{\circ}$. This is due to the large fluctuations of the wind vane measurements.
### Table 1. Error statistics of the reconstruction.

| Error statistic | Wind speed [ms$^{-1}$] | Wind direction [$^\circ$] |
|-----------------|------------------------|--------------------------|
| Mean            | 0.08                   | 0.08                     |
| Std             | 0.55                   | 7.20                     |
| 25 %            | -0.27                  | -3.30                    |
| 50 %            | 0.08                   | 0.54                     |
| 75 %            | 0.43                   | 4.05                     |
| RMSE            | 0.55                   | 7.02                     |

#### Figure 5.
Example of two time series (left: four hours, right: one hour) of wind speed measurement at T60 (blue graph) and the flow prediction for $p = 8$ (orange graph) and $p = 2$ (green graph).

### 3.2. Flow prediction

For the evaluation of the flow prediction we use the data of all turbines up to a certain point in time and create an estimation of the wind speed and wind direction for a later point in time at the location of turbine T60 and compare these to the measurements at T60. We perform this evaluation for three prediction horizons $\Delta t = 30\text{s}, 60\text{s}, 120\text{s}$. Thus, the minimum age of the particles is the prediction horizon, so that the spatio-temporal distance cannot become 0 in this case either. The parameters are the same as for the flow reconstruction (see Section 3.1).

Qualitative excerpts of the prediction with a horizon of $\Delta t = 60\text{s}$ are depicted in Figure 5 with the Shepard parameter set to $p = 2$ and $p = 8$ on the same time series examples as before. Similar to the reconstruction, the prediction agrees well with the target wind speed, as shown in the left graph of Figure 5. Also, the right graph has very similar behaviour, although a slight delay of the sinusoidal wind speed fluctuations can be recognized, which will be discussed further in Section 4 of the manuscript.

The quantitative analysis of the prediction error of the wind speed and wind direction for $\Delta t = 60\text{s}$ is illustrated in the histograms in Figure 6 and a summary of all three prediction
Figure 6. (Left) Histogram of the wind speed error of the flow prediction with a time horizon of \( \Delta t = 60 \) s. (Right) Histogram of the wind direction error of the flow prediction with a time horizon of \( \Delta t = 60 \) s.

The error statistics of the predictions have an overall similar behaviour, which is in the same order of magnitude as the flow reconstruction. For larger prediction horizons, the RMSE growths, although this increase is relatively small. The mean error indicates systematic deviations. These could also be discovered when examining the sample time series. Here it is noticeable that the wind speed is sometimes over- or underestimated for longer periods. We are discussing this in more detail in the following Section 3.3.

3.3. Error correction
We have looked at the autocorrelation of the wind speed prediction error and found that up to a lag of 360 s the correlation value (Pearson correlation) is more than 0.4. This means that there are deterministic deviations or biases between the forecast and the target values for certain periods. This can be caused by a number of reasons. We believe that this is partly related to the calibration of the measurement systems. For example, we have observed that turbines with curtailed power output measure higher wind speeds on average. Another reason is the wake effect in the wind farm.

To cope with these biases, we are proposing an automated error correction. The autocorrelation of the error signal indicates persistence of a deterministic error, so we formulate
this into an assumption and can correct a prediction by considering the deviation of a previous prediction. In detail, we calculate the moving average of the prediction error and subtract the result from our next prediction according to Equation (8):

\[ U_{\text{corr}}(t) = U_{\text{est}}(t) - \left( \frac{1}{T} \int_0^T U_{\text{est}}(t - \Delta t - \tau) - U_{\text{target}}(t - \Delta t - \tau) d\tau \right) \] (8)

We set \( T = 360 \) s and applied this correction to our previous evaluation, achieving the results in Table 3.

### Table 3. Error statistics of the prediction.

| Error statistic | wind speed [ms\(^{-1}\)] | wind direction [°] |
|-----------------|---------------------------|-------------------|
| \( \Delta t \)  | 30 s | 60 s | 120 s | 30 s | 60 s | 120 s |
| mean | 0.00 | 0.00 | 0.00 | -0.01 | -0.01 | -0.02 |
| std | 0.46 | 0.48 | 0.52 | 6.72 | 6.90 | 7.11 |
| 25 % | -0.24 | -0.26 | -0.29 | -3.04 | -3.14 | -3.31 |
| 50 % | 0.00 | 0.00 | -0.00 | 0.25 | 0.25 | 0.24 |
| 75 % | 0.24 | 0.26 | 0.29 | 3.62 | 3.70 | 3.81 |
| RMSE | 0.46 | 0.48 | 0.52 | 6.72 | 6.90 | 7.11 |

The correction provides a significant improvement in the results (see RMSE), especially concerning the wind speed prediction. Furthermore, the results are now more symmetrical, i.e. the mean value is close to 0 and the quartiles are distributed symmetrically around 0, indicating that the deterministic error is considerably reduced. For wind direction prediction, only minor improvements have been achieved, but in this case, the proportion of deterministic errors is probably also lower.

### 3.4. Benchmark

To close the evaluation, we are comparing the corrected flow prediction with the persistence forecast method. This method consists of the assumption that the wind speed and wind direction stay persistent: \( U_{\text{est}}(t) = U_{\text{target}}(t - \Delta t) \). We calculated the error of the persistence forecast for prediction horizons of \( \Delta t \in [0, 180] [\text{s}] \) (see Figure 7): Figure 7 reveals that the presented prediction method performs better than the persistence forecast in terms of the RMSE.

### 4. Discussion

The presented algorithm together with the automated error correction can produce a robust minute-scale forecast for downstream turbines in a wind farm from turbine measurements. During the evaluation period, the wind direction changed significantly from easterly wind over southerly to westerly wind, which underlines the omnidirectional properties of the method. The Shepard parameter strongly influences the outcome of the forecast. A smaller Shepard parameter (e.g. \( p = 2 \)) results in a smoothed out but robust forecast for the average wind speed, while a higher value (e.g. \( p = 8 \)) delivers a flow, where the turbulence is almost preserved, but the uncertainty of the forecast is higher. This is important to consider for the application of the algorithm. For load calculation, for instance, a higher value can give a good prediction of the fluctuations. For predictive turbine control, on the other hand, a more robust prediction of the average wind speed can be more advantageous.
Additionally, the other parameters influence the quality of the flow reconstruction and prediction. We have tested different sets of parameters and decided to use those stated in Section 3.1, but we did not carry out an extensive study of the parameterization.

In the investigation, we observed several sources for uncertainties. First the inherent disturbances of the rotor on the sonic anemometer on the top of the nacelle of the turbine. This introduces noise on the wind direction measurement. Second, we noticed that there were offsets in the wind speed estimation of the turbines. Especially when the power output of a turbine was curtailed, the wind speed measurement average of that turbine was higher than on the neighbouring turbines. Third, the algorithm does not directly account for the wake effect, which is clearly visible in the wind speed averages, but often hard to detect at a certain point in time due to the turbulence in the wind speed measurement. The presented error correction mitigates these sources for uncertainty but cannot eliminate them.

Lastly, we noticed a small delay of the prediction in the demonstrated one-hour time-series. We believe the reasons for this are errors in the alignment and the wind vane measurements of the turbines. The propagation of the parcels amplifies these error in the wind direction which can result in delays.

5. Conclusions
This manuscript introduces a simple and intuitive spatio-temporal reconstruction and forecasting model for the flow inside a wind farm. The method is validated by one month of 1 Hz full-field wind farm data from the offshore wind farm Global Tech I and a simple error correction method takes deterministic deviations of the turbines into account. The results prove that the simplified Lagrangian advection can propagate the flow information through the wind farm and can deliver robust minute-scale forecasts in real-time, which outperform the persistence method.

This can be especially interesting for nowcasting, monitoring, identifying abnormal turbine behaviour and predictive wind farm control e.g. model predictive control (MPC). In the future, we plan to use this algorithm as data assimilation for the real-time flow solver presented in [11], in which the turbines and wake effects can be simulated and thus a more realistic wind farm flow can be determined in a higher fidelity framework.

Acknowledgements
This work was partially funded by the German Federal Ministry for Economic Affairs and Energy (BMWi) in the scope of the OWP Control project (FKZ 0324131A). Furthermore, we
acknowledge the wind farm operator Global Tech I Offshore Wind GmbH for providing SCADA data and their support of the work.

References

[1] Burger B 2020 Jährlicher Anteil der Windenergie an der Stromerzeugung in Deutschland "https://www.energy-charts.de/ren_share_de.htm?year=all&source=wind-share&period=annual" [Online; accessed 26-Februar-2020]
[2] Pinson P 2012 Journal of the Royal Statistical Society. Series C: Applied Statistics 61 555–576 ISSN 00359254
[3] Valldecabres L, Nygaard N G, Von Bremen L and Kühn M 2018 Journal of Physics: Conference Series 1037 ISSN 17426596
[4] Bossanyi E 2019 Journal of Physics: Conference Series 1222 ISSN 17426596
[5] Vera-Tudela L and Kühn M 2017 Renewable Energy 107 352–360
[6] Valldecabres L, Nygaard N G, Vera-Tudela L, Von Bremen L and Kühn M 2018 Remote Sensing 10 1701
[7] Knudsen T, Bak T and Soltani M 2011 Wind Energy 14 877–894
[8] Boersma S, Doekemeijer B, Gebraad P, Fleming P, Annoni J, Scholbrock A, Frederik J and van Wingerden J W 2017 A tutorial on control-oriented modeling and control of wind farms 2017 American Control Conference (ACC) (IEEE) pp 1–18 ISBN 978-1-5090-5992-8
[9] Simley E, Fürst H, Haizmann F and Schlipf D 2018 Remote Sensing 10 863
[10] Knudsen T, Bak T and Svenstrup M 2015 Wind Energy 18 1333–1351
[11] Rott A, Boersma S, van Wingerden J W and Kühn M 2017 Journal of Physics: Conference Series 854