Correlations of power output fluctuations in an offshore wind farm using high-resolution SCADA data

Janna Kristina Seifert[1], Martin Kraft[1,], Martin Kühn[1,], and Laura J. Lukassen[1]

[1]ForWind, Institute of Physics, Carl von Ossietzky University Oldenburg, Küpkersweg 70, 26129 Oldenburg, Germany

Correspondence: Janna Kristina Seifert (janna.kristina.seifert@uol.de)

Abstract. Space-time correlations of power output fluctuations of wind turbine pairs provide information on the flow conditions within a wind farm and ~~on~~ the interactions of ~~the~~ wind turbines. Such information ~~plays an important role for the control of~~ can play an essential role in controlling wind turbines and short-term load or power forecasting. However, the ~~challenge to analyse space-time~~ challenges of analysing correlations of power output fluctuations ~~of wind turbine pairs in a free field~~ in a wind farm are the highly varying flow conditions. Here, we present an approach to investigate space-time correlations of power output fluctuations of streamwise-aligned wind turbine pairs ~~in free field~~ based on high-resolution SCADA data~~, which~~. The proposed approach overcomes the challenge of ~~highly~~ spatially variable and temporally variable flow conditions within the wind farm. ~~Using eight months~~ We analyse the influences of the different statistics of the power output of wind turbines on the correlations of power output fluctuations based on eight months of measurements from an offshore wind farm with 80 wind turbines~~,~~. First, we asses the effect of the ~~influences of different parameters on the correlation of power output fluctuations are analysed. Wind direction investigations show that~~ wind direction on the correlations of power output fluctuations of wind turbine pairs. We show that the correlations are highest for ~~streamwise aligned~~ the streamwise-aligned wind turbine pairs and decrease ~~towards spanwise pairs~~when the mean wind direction changes its angle to be more perpendicular to the pair. Further, ~~it is found that the correlation of power output fluctuations of streamwise aligned~~ we show that the correlations for streamwise-aligned wind turbine pairs ~~depends~~ depend on the location of the wind turbines within the wind farm ~~as well as the~~ and on their inflow conditions (~~free-stream~~ free stream or wake). ~~The main outcome~~ Our primary result is that the ~~correlation of streamwise aligned~~ standard deviations of the power output fluctuations and the normalised power difference of the wind turbines in a pair can characterise the correlations of power output fluctuations of streamwise-aligned wind turbine pairs~~can be characterised by~~. Further, we show that clustering can be used to identify different correlation curves. For this, we employ the data-driven k-means clustering algorithm to cluster the standard deviations of the power output fluctuations ~~and the~~ of the wind turbines and the normalised power difference of the wind turbines in a pair. ~~Evaluating these parameters with the data-driven clustering algorithm k-means~~Thereby, wind turbine pairs with similar power output fluctuation correlations are ~~grouped depending on these parameters and independent~~ clustered independently from their location. ~~These groups are here referred to as correlation states. With thisapproach~~ With this, we account for the highly variable flow conditions inside a wind farm~~which~~, which unpredictably influence the correlations~~in an unpredictable way. As a final result we shows that these parameters lead to clearly distinguishable correlation states~~.

1 Introduction

Wind energy continues to be a growing source of energy. In 2019, 15.4 GW of new wind power capacity was installed in Europe, ~~24 % thereof~~ with 24% thereof located offshore (Komusanac et al., 2020). Considering the offshore wind power in 2019, the capacity in Europe has increased by 3.627 GW ~~,~~ and a total of 7 wind farms were fully connected to the grid~~, while~~. Due to the increased size of the newly installed wind farms, the average size of ~~wind farms increased~~ offshore wind farms rose to 621 MW (Ramírez et al., 2020).

With the continuously increasing share of wind energy in the grid, the challenge of handling this highly fluctuating energy source becomes more important, as discussed in Ren et al. (2017). To convert wind energy into electrical energy, wind turbines are installed ~~,~~ generally in groups (wind farms) at onshore and offshore sites. Fluctuations in their power output ~~are the result of~~ result from environmental influences such as changes in wind speed or wind direction, influences from neighbouring wind turbines ~~, but also their own~~ and their state of operation. These power output fluctuations create challenges regarding the grid stability and ~~therefore are~~ are therefore an important field of investigation, (cf. Sorensen et al., 2007; Bossuyt et al., 2017b).

~~To achieve the maximum power output for a respective site, wind~~ Wind turbines within a wind farm are placed as efficiently as possible to achieve the maximum power output for a respective site. The spacing of wind turbines is determined by the terrain of the site and the influence of wind turbines ~~onto~~ on each other (their wake). Wakes cause energy losses through reduced wind speeds and, at the same time, greater power output fluctuations and loads through increased turbulence (Crespo and Hernàndez, 1996; Vermeer et al., 2003).

Wake and wind farm flow effects on different spatial and temporal scales are reviewed by Porté-Agel et al. (2020). Many studies do not ~~take into account~~ consider the power output fluctuations of wind turbines, which ~~have a major impact on~~ significantly impact the power output of a wind farm and ~~on~~ the electrical grid. Thus, for further improvement of wind turbine control strategies like active power control (Vali et al., 2019) and grid stability by minute-scale prediction of offshore wind farm power (Valldecabres et al., 2020), the occurrence of ~~wind turbine~~ power output fluctuations of wind turbines and their correlation within a wind farm are of great interest.

Andersen et al. (2017) investigated the influence of large coherent structures on the power output of wind turbines in large wind farms. ~~They~~ The large coherent structures were found to cause high correlations in the power output of ~~streamwise aligned~~ streamwise-aligned wind turbines. Research on wind speed correlations and power output correlations has shown that the wind turbines within a wind farm influence each other's power output fluctuations. Bossuyt et al. (2017a) ~~showed that for a~~ found significant correlations of the power output for a streamwise-aligned set up of a wind farm of 100 porous disc models in a wind tunnel~~, significant correlations of the power output can be found for a streamwise aligned set up of the discs~~. Next to an increased turbulence intensity throughout the wind farm, the correlation of the power output reduced with the increasing distance of the discs~~to each other~~. In an LES study by Lukassen et al. (2018)~~velocity~~ , space-time correlations of velocity fluctuations within a wind farm with periodic boundary conditions (modelling a periodic array of wind turbines) were analysed and modelled analytically. ~~The velocity~~ Velocity fluctuations, which are directly related to power output fluctuations, showed pronounced space-time correlations. Furthermore, the variance of the wind velocity and the mean velocity turned out

to be important parameters in the ~~modelling set up~~space-time correlation model. Stevens and Meneveau (2014) investigated ~~spectra of wind turbine~~ the spectra of power output fluctuations of wind turbines in LES of ~~finitely sized~~ finite-sized and infinitely large wind farms. The spectra were found to be ~~depended on non-trivial~~ dependent on the power output correlations of streamwise placed wind turbines. The ~~correlation of~~ power output correlation of the two wind turbines was significantly influenced by the wind direction, i.e. the lowest correlation for ~~spanwise placed~~ spanwise-placed wind turbines and highest correlation for ~~streamwise aligned~~ streamwise-aligned wind turbines. Dai et al. (2017) analysed ~~1 Hz~~ 1 Hz wind farm SCADA data ~~with respect to~~ concerning the influence of wind speed fluctuations around a mean velocity and wind direction fluctuations ~~on~~ around a mean wind direction on the wind turbine power output fluctuations of single wind turbines. They showed a direct relation between the wind speed fluctuations and power output fluctuations in the partial load regime. ~~By using~~ Using 10-minute averaged wind farm SCADA data, Braun et al. (2020) derived a stochastic model for the power time series of wind turbines ~~which was~~ based on the temporal autocorrelation of the power of single wind turbines.

~~In our work , we analyse~~ This work analyses 1 Hz wind farm SCADA data to describe the space-time correlations of the power output fluctuations of wind turbine pairs. In contrast to the wind tunnel measurements by Bossuyt et al. (2017a) and the LES analysis by Lukassen et al. (2018) mentioned above, the data set processed here includes unstable inflow conditions (varying wind speeds and wind directions), dynamically operating wind turbines as well as changing flow conditions within the wind farm. Furthermore, there may be potential measurement inaccuracies. The result is a large and highly complex data set. In this paper, we investigate the influencing factors on the correlation of power output fluctuations of wind turbine pairs and introduce parameters to distinguish different correlation curves, ~~here~~ herein called correlation states. ~~These parameters~~ A state defines a group of similar correlation curves. Note that the states found here refer to this specific wind farm and the considered time period. The parameters introduced to characterise correlation curves are then evaluated with a data-driven clustering algorithm ~~with the aim~~ to group the data according to the underlying correlation ~~states~~curves.

Starting with the description of the evaluated data set in Sect. 2, the processing of the data is explained in Sect. 2.1 and 2.2. The space-time correlation of power output fluctuations per wind turbine pair for time intervals of 600 s is introduced. ~~Using a filtered data set with less varying flow conditions, the~~ The correlation of wind turbine pairs is analysed ~~for different wind directions~~ in Sect. 3.1 for different wind directions using a filtered data set with less varying flow conditions. The correlation for wind directions with ~~streamwise aligned~~ streamwise-aligned wind turbines is evaluated in more detail. In Sect. 3.2, the location ~~dependency~~ dependence of the power output fluctuation correlation is determined by the comparison of wind turbine pairs located in different wind farm rows to confirm the findings of the wind tunnel measurements by Bossuyt et al. (2017a). With this and the results from the LES analysis by Lukassen et al. (2018), we identify relevant wind turbine ~~statistics which~~ power output statistics that influence the correlation. In Sect. 4, we use the straightforward *k*-means clustering approach (Lloyd, 1982) to group the data with respect to these statistical quantities~~and show that there~~ , which show that they are clearly distinguishable correlation states. ~~Conclusion~~ The conclusion and an outlook are drawn in Sect. 5.

2 Reference wind farm and data processing

The analysis performed in this work is based on measurements from the offshore wind farm Global Tech I (GT I). It is located in the North Sea, which is more than 100 km off the coast of Northern Germany. Its total capacity of 400 MW is provided by 80 wind turbines spread over an area of about 41 km^2. The wind turbines of type Adwen AD 5-116 have a rated power of 5 MW, a rated wind speed of 12.5 ms^{-1}, a hub height of 92 m and a rotor diameter (D) of 116 m. They are installed in a ~~grid like~~grid-like, slightly asymmetric pattern with a triangular shape towards the south (see Fig. 1).

The analysed data set was measured in ~~a period of~~ about eight months, from January 1st, 2019 until September 9th, 2019 and

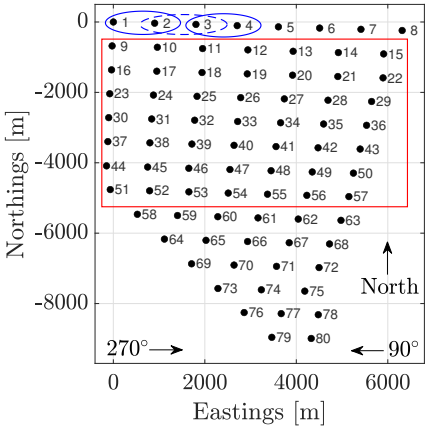


Figure 1. Layout of GT I. Each wind turbine is labelled with its corresponding number. The spacing of the wind turbines is inhomogeneous. The wind directions 90° and 270° (marked in the figure) will be analysed in detail in ~~the~~ subsequent sections. The red square depicts the set of wind turbines ~~which~~ that will later be used during the ~~location-dependency~~ location-dependence analysis in Sect. 3.2 due to their symmetric arrangement. In the clustering analysis in Sect. 4, the whole wind farm will be used. The blue ellipses exemplarily show the definition of the considered wind turbine pairs. ~~The definition of all pairs is listed in~~ Tab. A1 lists the definition of all pairs.

consists of 1 Hz wind turbine SCADA data. The processed signals include the generated power P, the azimuth angle of the wind turbines (i.e. the nacelle direction) θ, the ~~nacelle based~~ nacelle-based wind direction φ (measured relative to θ), the pitch angle β of each blade ~~,~~and a reconstructed wind speed U.

The reconstructed wind speed U is not directly measured but provided as a variable ~~which~~ that results from the measured power and control variables of the wind turbine (details on the reconstruction of U are not available). Due to ~~that~~this, U is considered as an approximated and idealised value ~~which~~ that does not include the wind speed independent power reduction, e.g. by a yaw misalignment of the wind turbine~~due to measurement errors of the wind direction. In the context of~~ . In this work, it can still be used ~~for assessing~~ to assess the effect of the wind speed on the correlations of power output fluctuations of wind turbine pairs~~which is~~ , which will be further discussed in Sect. 2.2.

The azimuth angle θ of the wind turbine refers to the direction it is facing in its preset reference system. This system does not necessarily ~~exactly match to~~ match the global geographical one due to the measurement inaccuracies of the azimuth angle and

a potentially inaccurate north orientation of the reference system of each wind turbine (cf. Bromm et al., 2018).

The nacelle based wind direction φ is estimated based on the measurements of two 2D sonic anemometers installed behind the rotor of each wind turbine. These measurements have to be treated with care as the measured flow behind the rotor is disturbed by the rotation of the rotor and the nacelle itself. Thus, it is only an estimation of the wind direction and yaw of the wind turbine. However, as shown by Dai et al. (2017), wind direction fluctuations at reasonable yaw angles (~~$< 45°$~~ $\leq 45°$) have only little effect on the power output fluctuations of wind turbines~~and thus~~. Thus, inaccuracies in φ have no major influence on the performed analysis. The combined measurements of θ_i and φ_i define the wind direction Φ_i at the i-th wind turbine.

~~For assessing an~~ To assess the average wind direction for the wind farm, we average over Φ_i of all wind turbines to reduce the influence of false measurements of single wind turbines. Due to the size of the considered wind farm, the wind direction is not expected to be consistent throughout the ~~whole~~ wind farm. Single wind turbines could be facing different wind directions compared to the average wind direction of the wind farm (cf. Schneemann et al., 2020; Sanchez Gomez and Lundquist, 2020). The wind direction of the wind farm that is averaged over all available wind turbines is defined as Φ_{av}.

2.1 ~~Filtering~~Data selection and filtering

Wind tunnel experiments and LES simulations as described in Sect. 1 ~~,~~pose controllable conditions for evaluating correlations. Such conditions cannot be met in a ~~free field~~ free-field wind farm. Next to temporally and spatially varying wind conditions, the ~~layout of the wind farm~~ wind farm layout leads to unequal conditions for wind turbines due to their positions, e.g. changing wind direction throughout the wind farm ~~,~~and asymmetric wind turbine spacing, especially for large wind farms. Further, each wind turbine ~~is operating independently from the~~ operates independently from other wind turbines including yawing, pitching, ~~start up or shut~~ starting up or shutting down. Next~~to this~~, single wind turbines can be set to operate in a down-rated state or be shut down due to maintenance or other reasons. ~~All these factors multiply to an order of unpredictable variability within a wind farm which causes highly dynamical flow conditions~~ The combination of all these factors causes highly dynamic flow conditions and thus, an unpredictable variability. To cope with these issues, the data set is filtered for different operation states creating a cleaned data set with comparable operation state conditions for all wind turbines. For each considered ~~600 s interval , the parameters~~ interval of 600 s, the conditions defined in the following have to be met, cf. Tab. 1.

In general, a wind turbine operating in partial load is not pitching, and the velocity in its wake is always below the rated wind speed. A wind turbine operating in full load aims at keeping a constant rotor speed and power by pitching its blades, where the wind speed in its wake can be larger than the rated wind speed. ~~To~~ The data set is limited to partial load to avoid the effects of pitching and the different wake behaviour on the correlation~~of wind turbines, the data set is limited to partial load~~. To further

Table 1. Filters applied to the raw data of each wind turbine within the wind farm.

Signal	Power	Pitch	Yaw
Settings	$0.5\ \text{MW} \leq P \leq 4.5\ \text{MW}$	$\beta < -1.3°$	$\theta = const.$

avoid effects from the transition from idle mode into operation or the transition from partial load to full load, only the data of the wind turbines generating power in the range of 0.5 MW and 4.5 MW is considered.

The previously defined ~~,~~ limited power range still includes derated wind turbines. For derating wind turbines, their controller is manually changed so that their maximum power is limited to a certain value, which is lower than their nominal power. Due to ~~that~~this, wind turbines might start pitching already in the previously defined load range as their newly set power limit is ~~reached already~~ already reached at lower wind speeds. Hence, to ~~fully~~ entirely exclude pitching wind turbines, the data is filtered for any pitching activity. Please note that for this specific data set, this implies that $\beta < -1.3°$.

Furthermore, yawing wind turbines are excluded from the analysis as well. The adjustment of wind turbines to the wind direction is managed by each wind turbine individually. Thus, wind turbines could be facing slightly different wind directions Φ_i and start yawing at different times. The yawing activity of a wind turbine transfers to its wake ~~,~~ i.e. changes its deflection (cf. Bromm et al., 2018). ~~This~~ Thus, yawing would affect the correlation for a pair of two wind turbines. To exclude yawing wind turbines, no change of θ is allowed in the regarded 600 s time interval: ~~$\theta = const.$~~ $\theta = const$.

To further filter the data for wind directions, the average wind direction Φ_{av} of all wind turbines is calculated for each time step of the regarded 600 s time interval. The average wind direction Φ_{av} has to fit the wind direction of interest within a tolerance of $\pm 10°$ for all time steps in the regarded 600 s. Note that the borders of the interval ~~are including~~ include the lower limit and ~~excluding~~ exclude the upper limit. Since this data filter only applies to the average wind direction Φ_{av}, individual wind turbines might have a slightly deviating relative wind direction for this specific time interval~~due to~~. This deviation could be caused by a false wind direction measurement, a yawing process ~~which~~ that has taken place asynchronously to the majority of ~~the~~ other wind turbines ~~,~~ or a wind direction deviation due to local changes over the area of the wind farm. This means there is no threshold for yaw misalignment within the 600 s intervals. As mentioned before, these effects have a limited effect on the power output fluctuations of the wind turbines.

As a summary, the overall data filtering procedure is as follows. ~~Each~~ The correlation analysis uses each time interval of 600 consecutive seconds where the two wind turbines of a wind turbine pair (as defined in Fig. 1) both ~~fulfil~~ pass all of the ~~above described filtering parameters~~above-described data filters, i.e. power range, pitch, yawing and wind direction~~, is used in the correlation analysis~~. This means that for different time intervals, a different set of wind turbine pairs is considered~~and that furthermore~~. Furthermore, wind turbine pairs can be considered for multiple time intervals.

2.2 Correlation of power output fluctuations

Power output fluctuations of individual wind turbines are defined as deviations of the instantaneous power from the average power of the regarded wind turbine i within a certain time interval Δt. We analyse time intervals of $\Delta t_{600} = 600$ s:

$$P'_{i,t_j}(t) = P_i(t) - \langle P_i(t) \rangle_{\Delta t_{600}} \tag{1}$$

where $\langle P_i(t) \rangle_{\Delta t_{600}}$ is the average of the measured power $P_i(t)$ over an interval Δt_{600}, including all 600 values for t in the discretised interval $[t_j, t_j + 599 \text{ s}]$. $P'_{i,t_j}(t)$ is the power output fluctuation within the interval Δt_{600} (the index t_j is omitted in the following). Depending on the data availability, the next interval of 600 consecutive seconds could go from

$[t_j + 1\ \mathrm{s}, t_j + 1\ \mathrm{s} + 599\ \mathrm{s}]$, and thus ~~,~~overlap the previous one up to 599 s. This does not result in significantly different findings compared to non-overlapping intervals as shown in App. B.

The selection of the interval size of 600 s is based on the layout of the wind farm and the considered power ranges or corresponding wind speeds. For example, considering the spacing of up to 9 D for westerly winds, ~~the~~ with a cut-in wind speed of $4\ \mathrm{ms}^{-1}$ and a rated wind speed of $12.5\ \mathrm{ms}^{-1}$, a particle moving with the undisturbed wind would take from about 84 s up to 261 s to travel from one wind turbine to its downstream neighbour. Taking a lower advection wind speed within the wind farm into account, a considered interval length of 600 s captures potential correlations of interest. ~~This means, each~~ Each time step followed by 599 consecutive time steps forms an interval, individually for each wind turbine. For all available intervals of all wind turbines, the power output fluctuations are ~~then~~ calculated based on Eq. 1.

To analyse the influence of wind turbines ~~onto~~ on each other, the space-time correlation is calculated ~~. This is done~~ using the Pearson correlation coefficient (Pearson, 1896)

$$r(\tau) = \frac{\langle P'_A(t) P'_B(t+\tau) \rangle_{\Delta t_{300}}}{\sqrt{\langle P'^2_A(t) \rangle_{\Delta t_{300}} \langle P'^2_B(t+\tau) \rangle_{\Delta t_{300}}}} \tag{2}$$

where $\langle ... \rangle_{\Delta t_{300}}$ is the average over an interval $\Delta t_{300} = 300$ s including all 300 values for t in the discretised interval $[t_j, t_j + 299\ \mathrm{s}]$, $r(\tau)$ is the Pearson correlation coefficient in dependence of a time lag τ, $P'_A(t)$ is the power output fluctuation of the upstream wind turbine A following Eq. 1 at a time t, $P'_B(t+\tau)$ is the power output fluctuation of the downstream wind turbine B at a time $t + \tau$ with a time lag τ.

The Pearson correlation coefficient is a value between -1 and 1, where 1 depicts the maximum possible linear correlation, -1 is the maximal linear anti-correlation and a value of 0 depicts no linear correlation. The correlation coefficient is evaluated for a fixed period of 300 s from $P'_A(t)$ to $P'_A(t + 300\ \mathrm{s})$ and likewise $P'_B(t+\tau)$ to $P'_B(t + 300\ \mathrm{s} + \tau)$. This allows a maximum time lag of $\tau = 300$ s for each considered 600 s interval.

Similar to Taylor's hypothesis (Taylor, 1938), we assume that ~~depending on the wind speed, wind~~ the wind structures responsible for the power output fluctuations measured at an upstream wind turbine A ~~, take some time to travel the~~ that travel a certain distance to the downstream wind turbine B with an advection speed that is similar to the average wind speed over that distance. But in contrast to Taylor's hypothesis, we do not assume frozen eddies but expect wind structures to change and thus decorrelate while travelling downstream. Further, as we have no access to the average wind speed over the distance between wind turbines A and B, we use the average wind speed measured at wind turbine B as a reference. Hence, to compare ~~correlations at different~~ the correlations calculated for intervals with different average wind speeds and different wind turbine distances, the time lag τ is normalised for each time interval starting at t_j individually

$$\tau_{norm,intv} = \tau \cdot \frac{\langle U_B(t+\tau) \rangle_{\Delta t_{300}}}{x_{AB}} \tag{3}$$

where $\tau_{norm,intv}$ is the ~~normalized~~ normalised time lag, $\langle U_B(t+\tau) \rangle_{\Delta t_{300}}$ is the average reconstructed wind speed from a certain (downstream) wind turbine B for a time interval $\Delta t_{300} = 300$ s for t in the discretised interval $[t_j, t_j + 299\ \mathrm{s}]$ and a certain lag τ. This means for a certain τ, the averaging interval of $\langle U_B(t+\tau) \rangle_{\Delta t_{300}}$ is $[t_j + \tau, t_j + \tau + 299\ \mathrm{s}]$. x_{AB} is the distance between wind turbine A and wind turbine B.

~~Due to this~~ Next, the correlation curves with the normalised lag $\tau_{norm,intv}$ are discretised using a histogram with a reference time lag of

$$\tau_{norm} = \tau \cdot \frac{U_{max}}{x_{AB,mean}} \tag{4}$$

where τ is the time lag (0 s to 300 s), U_{max} is an artificially introduced velocity that has to be at least equal to the maximum possible wind speed to fit all normalised curves ($U_{max} = 13\ \mathrm{ms}^{-1}$ for this case). This value is based on the wind turbine power curve characteristics, including a tolerance as the wind turbines considered here reach their rated power at 12.5 ms^{-1}. $x_{AB,mean}$ is the average distance between wind turbine A and wind turbine B of the considered wind turbine pairs. Note that $\tau_{norm,intv}$ is used for stretching and shrinking of the correlation curves. τ_{norm} is only a reference time lag that is only created for binning of the stretched or shrunk correlations and does not change the correlation curves.

Due to the definition of $\tau_{norm,intv}$ and τ_{norm} (see Eq. 3 and 4), the peak of the correlation curves is expected to be found ~~at $\tau_{norm} = 1$~~ around $\tau_{norm,intv} = 1$ if the advection speed of the wind speed fluctuations matches the wind speed affecting wind turbine B. Thus, in partial load situations where wind turbine B is in the wake of wind turbine A, the peak is expected to be at $\tau_{norm} > 1$. Here, the reduced wind speed in the wake recovers slowly, so that the wind speed affecting wind turbine B, i.e., U_B is already partly recovered and hence larger than the advection speed of the fluctuations. As mentioned before, U_B is reconstructed and might differ from the actual wind speed affecting the wind turbines. However, in the context of this normalisation, the effect on the resulting correlations curves is marginal as the correlation curves may only be slightly shifted due to the deviation ~~to~~ from the real wind speed. ~~In a next step, the correlation curves with the normalised lag $\tau_{norm,intv}$ are discretised using a histogram with a reference time lag of~~

~~$$\tau_{norm} = \tau \cdot \frac{U_{max}}{x_{AB,mean}}$$~~

~~where τ is the time lag (0 s to 300 s), U_{max} is an artificially introduced velocity which has to be at least equal to the maximum possible wind speed to fit all normalised curves (here $U_{max} = 13\ \mathrm{ms}^{-1}$). $x_{AB,mean}$ is the average distance between wind turbine A and wind turbine B of the considered wind turbine pairs. Note that $\tau_{norm,intv}$ is only used for stretching and shrinking of the correlation curves and that τ_{norm} is used only for binning of the stretched or shrunk correlations.~~

3 ~~Wind-direction dependency~~ Wind direction dependence and location ~~dependency~~dependence

As described in Sect. 1, ~~the aim of this work is~~ this work aims to study the influences of the ~~free field~~ free-field wind farm situation on the space-time correlations of the power output fluctuations. For ~~that~~this, we analyse the time intervals of a fixed set of 66 wind turbine pairs, namely those ~~which are streamwise aligned~~ that are streamwise-aligned for the wind directions 90° and 270° ~~,~~(see Fig. 1 and Tab. A1). Note that the pairs are the same for both wind directions, but the order of the evaluation for the ~~wind-direction dependent~~ wind direction-dependent correlation differs (i.e. the upstream and downstream wind turbine position of a pair is reversed).

In the following, we average correlations over a wind direction interval of 20° and all available time intervals of the considered wind turbines (either all wind turbines or a selection of wind turbines). We consider 20° intervals due to a 10° tolerance in the wind direction measurements of the wind turbines. The averaged correlation is ~~noted as~~ denoted by $R(\tau_{norm})$. In Sect. 3.1, the average correlation ~~of~~ for the 66 wind turbine pairs is analysed for each wind direction interval separately. Further, ~~in Sect. 3.2,~~ the location-dependent correlations ~~of~~ for the wind turbine pairs are evaluated, and wind turbine statistics ~~which characterise~~ that characterise the power output fluctuation correlations are investigated in Sect. 3.2.

3.1 ~~Wind-direction dependent~~ Wind direction-dependent space-time correlation

After applying the data filters described in Sect. 2.1 to the intervals of the 66 wind turbine pairs, the average correlation per wind direction is determined. For each wind turbine pair, the power output fluctuation correlations are averaged over the wind direction intervals of 20°, which are applied to steps of 10°, i.e. the interval for the wind direction 90° corresponds to the directions $80° \leq \Phi_{av} < 100°$ and the consecutive interval for the wind direction 100° is $90° \leq \Phi_{av} < 110°$. For the 10°-wind direction steps from 0° to 170°, we treat the pairs according to Tab. A with ~~reversed~~ a reverse order and for the 10°-wind steps from 180° to 350°, we treat the pairs with the given order. This means ~~even for~~ that even for the wind directions where both the wind turbines of a pair are parallel to the wind direction, the ~~'~~'upstream' wind turbine A is chosen according to the table. ~~Afterwards~~Afterward, the results are averaged over all wind turbine pairs for each 10° step separately. Due to the different availability of each wind turbine pair, they influence the average correlation to a different amount. Figure 2 displays the amount of data of all correlation intervals of all wind turbine pairs per wind direction interval of 20°. The main wind direction is about 220° and shows the maximum occurrence, whereas 90° has about 20% ~~less~~ fewer data and 270° has about 45% ~~less~~lesser data. For wind directions from 350° to 20°, there was almost no data available.

Figure 3 displays the averaged power output fluctuation correlation per 10° wind direction step, which is averaged over the 20° wind direction interval and all the time intervals of all available wind turbine pairs. The averaged correlation coefficient is plotted as colour and the time lag τ_{norm} (Eq. 4) ~~as radius. For wind directions around 90~~is denoted as the radius. Due to the varying data availability per wind direction and the applied data filtering (see sec. 2.1), the average correlation curve per bin is based on a different number of data. It turns out that after filtering, no data is available for the bins around 330° ~~and 270~~to 10°~~the wind~~ . Wind turbines in a pair are ~~streamwise aligned~~streamwise-aligned for wind directions around 90° and 270°. Fluctuations take a certain time to travel from one wind turbine to the other, where the fluctuations are influenced by the upstream wind turbine. The highest correlation peak is ~~found to be~~ at $\tau_{norm} > 1$ according to the definition of τ_{norm} in Eq. 4. ~~For~~ At 90°, a correlation of about 0.16 is found~~whereas for 270~~, whereas a correlation of about 0.2 is noted at 270°. The maximum correlation around 0.2 may seem ~~rather low but~~ relatively low but it is reasonable considering the high variability in the flow and wind turbine dynamics in ~~free field measurements. As comparison, in~~ free-field measurements. These dynamics are most likely caused by the wind direction and wind speed changes and the individual operation of the wind turbines (yawing, limited power, shut off). Even though the correlation curves were adapted to the average wind speed per interval, the wind speeds were just an assessment and could change during the interval. Also, the wind direction is averaged over the whole wind farm, which means certain intervals could include data from wind turbine pairs facing a slightly different direction. Further,

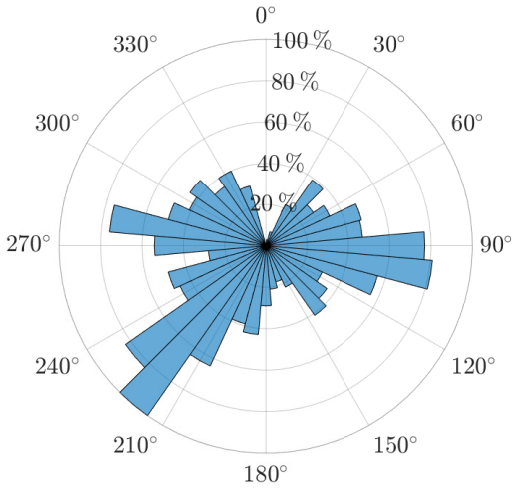


Figure 2. Availability of data per wind direction interval ~~normalized~~ normalised to the number of the available correlation intervals for 220°, namely 9,102,133 intervals. The outer labels depict the wind direction and the inner circles denote the percentage of availability.

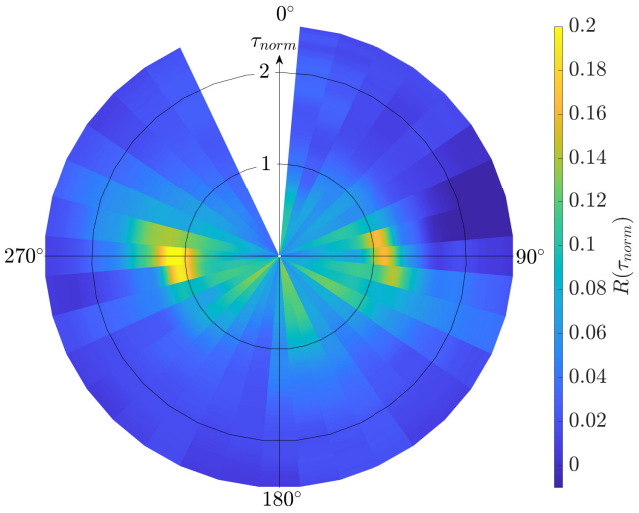


Figure 3. Average power output fluctuation correlation per 10° wind direction step of all available wind turbine pairs within the wind farm. Since the power output fluctuation correlation is averaged for wind direction intervals of 20° the intervals are overlapping by 10°. For a better visibility the intervals are visualised in 10° steps only, i.e. for 90° the interval goes from 80° to 100° but is visualised from 85° to 95°. The radius of the circle is the time lag τ_{norm}, i.e. $\tau_{norm} = 0$ is in the origin ~~,~~and $\tau_{norm} = 1$ is on the inner black circle. No data was available for the wind direction interval around 350° (cf. Fig. 2).

we only consider the intervals of wind turbine pairs that fit the data filter; however, other wind turbines could be yawing at the same time or start pitching. Thus, the flow within the wind farm could still be influenced by these wind turbines. In the

LES study of Lukassen et al. (2018), a maximum correlation coefficient of about 0.5 was found for the space-time correlations of wind speeds measured at comparable distances with comparable wind speed. In the wind tunnel experiments by ~~Bossuyt et al. (2017b)~~Bossuyt et al. (2017a), a maximum correlation of about 0.55 was found for the space-time correlation of the reconstructed power output of discs placed at comparable distances with comparable wind speeds. ~~Both, in~~ In both the simulations and ~~in the~~ experiments, the flow conditions are ideal compared to ~~free field~~ those in the free-field measurements. For wind directions approaching 0° and 180°, the wind turbines in a pair are oriented more perpendicular to the wind direction~~and~~ , and the fluctuations reach both wind turbines A and B at nearly the same time. This leads to a change in the expected position of the highest peak and the peak magnitude of the correlation curves. The found correlations are not as pronounced as those for the streamwise case ~~above~~ (i.e. around 90° and 270°), which confirms the simulation results by Stevens and Meneveau (2014). Thus, we will not investigate the spanwise correlations in further detail~~here~~.

Figure 4 shows the average power output fluctuation correlation around 90° and 270° in detail as cuts through Fig. 3. The absolute peaks are at 90° and 270°. For wind directions where the wind turbines in a pair are less ~~streamwise aligned~~ streamwise-aligned, the peak decreases and the correlation curve flattens. ~~In contrast to 90, the~~ The correlations for 270° are more defined and show slightly larger peak values compared to those for 90°. This may be due to the asymmetric wind farm layout (cf. Fig. 1). The deviation between the average correlation curves for wind directions around 260° and 280° could be ~~as well~~ caused by the not entirely horizontally aligned wind turbines and by the triangular shape at the lower part of the wind farm~~,~~ ; however, this ~~phenomena~~ phenomenon is not further investigated in this analysis.

3.2 Location-dependent space-time correlation

The location ~~dependency~~ dependence of the averaged power output fluctuation correlations is investigated for wind ~~directions~~ direction intervals around 90° and 270°. As mentioned before, ~~for these two wind directions~~ the wind turbines are~~streamwise aligned~~, on average, streamwise-aligned for these two wind direction intervals. The most northern wind turbines, 1 to 8, and wind turbines 58 to 80 in the lower triangle of the wind farm ~~,~~ do not follow the symmetric pattern of the square consisting of wind turbines 9 to 57. Thus, the following results are limited to this symmetric square as marked in Fig. 1.

Figure 5 displays the averaged correlations of the power output fluctuations for all wind turbine pairs included in the upper square of the wind farm for the wind ~~directions~~ direction intervals 90° and 270°~~, respectively. For 90, in total~~ . A total of 4,916,277 intervals and 3,329,333 intervals of 600 s are averaged ~~while for~~ for 90° and 270°~~the number of 600 s-intervals is 3, 329, 333.~~ respectively. Similar to Fig. 4, both correlation curves show ~~a comparable shape~~ similar shapes, whereas the correlation for 270° is ~~in general higher than~~ generally higher than that for 90°. ~~For 90, the~~ The maximum averaged correlation coefficient is about 0.16 ~~, for~~ and 0.21 for 90° and 270°~~it is about 0.21~~, respectively.

Further, the power difference (normalised by the average power output of the upstream wind turbine A) and the average standard deviation of the power output fluctuations of both wind turbines in a pair are determined to analyse the flow conditions. The results for 90° and 270° are listed in Tab. 2. For both wind directions, the averaged standard deviation of the power output fluctuations is larger for the downstream wind turbine B than for the upstream wind turbine A. ~~In general, however~~However, the averaged standard deviation of the power output fluctuations for 90° is smaller than that for 270°. ~~Also, the~~ The normalised

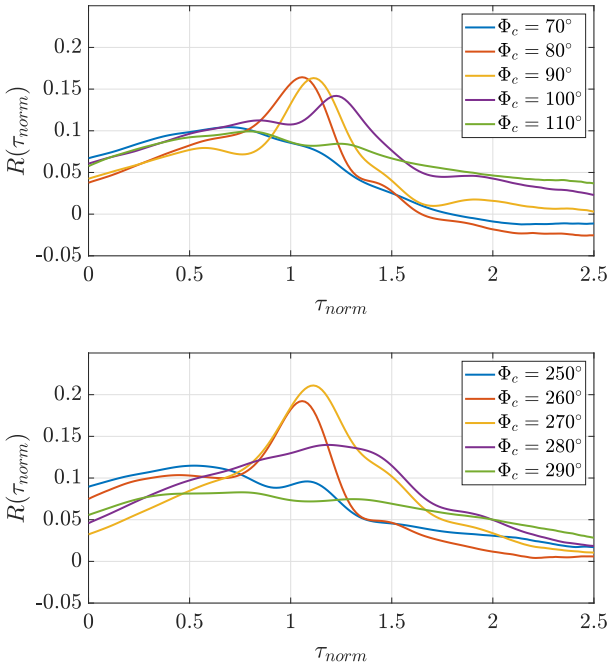


Figure 4. Average power output fluctuation correlations for wind direction intervals from around 70° to 110° and around 250° to 290° as radial cuts through Fig. 3. Φ_c depicts the centre of the wind direction intervals.

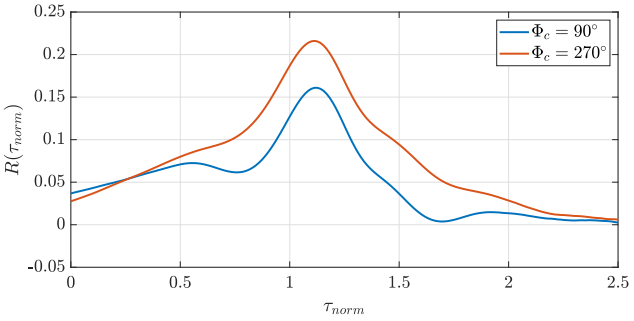


Figure 5. Average power output fluctuation correlation for wind direction intervals around 90° and 270° considering the wind turbines 9 to 57 in the symmetric square (cf. Fig. 1). Φ_c depicts the centre of the wind direction intervals.

power difference of the wind turbine pairs for 90° and 270° is about 12% and ~~for 270about~~ 8%, respectively. The different behaviour is likely to be caused by the distinct meteorological conditions, e.g. distribution of mean wind speed and atmospheric stability, between the two wind directions.

To further investigate the wind turbine location ~~dependency~~ dependence of the power output fluctuation correlations, the average correlation of wind turbine rows is calculated for both wind directions. Here, a wind turbine row consists of a line of wind turbines perpendicular to the incoming wind, as shown in the upper right corner of Fig. 6. For both wind directions, no

Table 2. Averaged wind turbine statistics computed for the wind direction intervals around 90° and 270° with A as the upstream wind turbine and B as the downstream wind turbine. $\sqrt{\langle P_A'^2 \rangle_{\Delta t_{600}}}$ is the standard deviation of the power output fluctuations of wind turbine A over 600 s intervals Δt_{600} (analogue for wind turbine B for the same 600 s intervals, respectively). $\langle P_A \rangle_{\Delta t_{600}}$ and $\langle P_B \rangle_{\Delta t_{600}}$ are the average power of wind turbines A and B over the same 600 s intervals. $\langle ... \rangle_{all}$ denotes the average of the statistics over all available time intervals of the wind turbine pairs. Note that Φ_c depicts the centre of 20° wind direction intervals, here from 80° to 100° and from 260° to 280°.

Φ_c	$\left\langle \sqrt{\langle P_A'^2 \rangle_{\Delta t_{600}}} \right\rangle_{all}$ [kW]	$\left\langle \sqrt{\langle P_B'^2 \rangle_{\Delta t_{600}}} \right\rangle_{all}$ [kW]	$\left\langle \frac{\langle P_A \rangle_{\Delta t_{600}} - \langle P_B \rangle_{\Delta t_{600}}}{\langle P_A \rangle_{\Delta t_{600}}} \right\rangle_{all}$
90°	212	222	0.12
270°	247	260	0.08

sharp correlation is found for the first row (turbine A in the first row, turbine B downstream of A, dark blue curves). It ~~has to be marked~~ should be noted that the upstream wind turbine A is standing in ~~free stream~~ the free stream, while the downstream wind turbine B is affected by the wake of the upstream wind turbine. Thus, the two wind turbines have very different inflow conditions. For wind turbine pairs located further downstream, both wind turbines are standing in the wake of the upstream wind turbines. Here, a clear correlation is found. For the correlation curves of the second to last row, the peaks become more defined as their width decreases.

As described by Bossuyt et al. (2017a), the turbulence intensity increases with the flow towards the back of the wind farm.

Table 3. Averaged wind turbine statistics per wind farm row computed for wind direction intervals around 90° and 270° with A as the upstream wind turbine and B as the downstream wind turbine. $\sqrt{\langle P_A'^2 \rangle_{\Delta t_{600}}}$ is the standard deviation of the power output fluctuations of wind turbine A over a 600 s interval Δt_{600} (analogue for wind turbine B for the same 600 s intervals~~, respectively~~). $\langle P_A \rangle_{\Delta t_{600}}$ and $\langle P_B \rangle_{\Delta t_{600}}$ are the average power outputs of wind turbines A and B over the same 600 s intervals. $\langle ... \rangle_{row}$ denotes the average of the statistics over all available time intervals of the wind turbine pairs in a row. Note that 90° and 270° again refer to the 20° wind direction intervals from 80° to 100° and from 260° to 280°.

	$\left\langle \sqrt{\langle P_A'^2 \rangle_{\Delta t_{600}}} \right\rangle_{row}$ [kW]		$\left\langle \sqrt{\langle P_B'^2 \rangle_{\Delta t_{600}}} \right\rangle_{row}$ [kW]		$\left\langle \frac{\langle P_A \rangle_{\Delta t_{600}} - \langle P_B \rangle_{\Delta t_{600}}}{\langle P_A \rangle_{\Delta t_{600}}} \right\rangle_{row}$	
Row	90°	270°	90°	270°	90°	270°
1	114	166	164	216	0.27	0.19
2	186	232	222	258	0.05	0.02
3	224	254	242	265	0.09	0.06
4	243	269	241	268	0.10	0.06
5	250	276	234	274	0.08	0.08
6	232	280	220	278	0.06	0.04

Furthermore, as indicated above in Lukassen et al. (2018), the variance of wind speed fluctuations plays an important role in

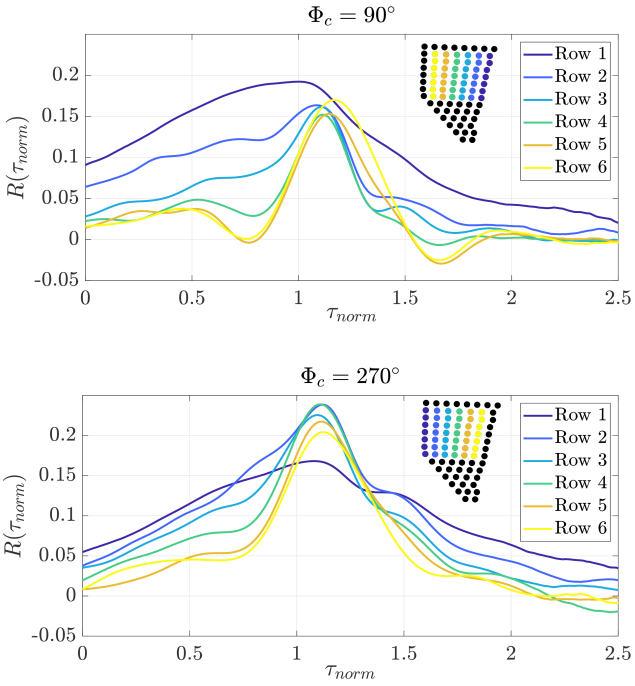


Figure 6. Averaged power output fluctuation correlation per wind farm row for wind direction intervals around 90° (top) and 270° (bottom) considering all wind turbines in the symmetric square (cf. Fig. 1). Φ_c depicts the centre of the wind direction intervals. For the considered correlation curves, wind turbine A is located in the respective coloured row and turbine B is one row downstream of A. As the wind turbines are analysed in pairs of two, the last row of wind turbines is unlabelled, as these wind turbines do not have a downstream partner. For both figures, the numbering and the colours of the rows are identical with regard to the considered wind direction.

modelling the velocity space-time correlations. To evaluate the row-dependent conditions in the measurement data, Tab. 3 lists the average standard deviations of the power output fluctuations measured at the upstream and downstream wind turbine of all pairs, as well as the average normalised power difference of all pairs. For both wind turbines in a pair, the average standard deviations of the power output fluctuations show a clear increasing trend throughout the wind farm, ~~similar to~~ which is similar to that of the turbulence intensity in the wind tunnel results mentioned above. The normalised power difference is largest for the first row, which is caused by the previously described deviating inflow conditions of the upstream and downstream wind turbine. This was also found in the experiment by Bossuyt et al. (2017a) where the first row generates the maximum power, while the second and following rows show a significant reduction.

4 *K*-means clustering of power output fluctuation characteristics

~~The results~~ Results of section 3.2 reveal that the standard deviation of the power output fluctuations ~~as well as~~ and the power difference of the wind turbines change depending on the location of the wind turbine (pairs) within the wind farm. As explained

in section 2.1, conditions in a wind farm are never ideal due to the variety of influence factors such as wind direction and wind speed fluctuations or influences of surrounding wind turbines. Turbines within the wind farm ~~which~~ that are turned off or derated might create ~~free-stream~~ free stream like inflows for downstream wind turbine pairs. Such irregularities influence the standard deviations ~~and~~ of the power outputs and the normalised power differences calculated for wind turbine pairs. For example, ~~if~~ a wind turbine that is turned off for a certain time interval ~~, it~~ is not considered in the analysis~~, but~~. However, it still influences the flow conditions within the wind farm and the statistics or correlations of the surrounding wind turbine pairs. Thus, a considered wind turbine pair downstream of the non-operating wind turbine could show a different correlation than that if the upstream wind turbine would be turned on. To identify these locally abnormal conditions and the resulting deviations in the power output fluctuations and their correlations, the k-means clustering algorithm is used to sort the correlations based on ~~the~~ previously defined statistics, ~~the~~ standard deviation and the normalised power difference of the wind turbines in a pair.

k-means is an algorithm ~~which~~ that iteratively sorts data into k clusters. After choosing an initial centre for each cluster (centroids) within the data, all data points are assigned to their nearest centroid. Afterwards, the new centres of the clusters are calculated based on the assigned data points. These steps are repeated until a previously defined number of iterations is reached or when the centres of the clusters no longer change. Finally, the data is distributed into k clusters. The result of the k-means algorithm is dependent on the starting positions of the cluster centres. Thus, the algorithm can be repeated with changing starting points for the clusters to find the best possible solution.

In the following, we investigate the clustering results for the directions 90° and 270°. Here, the triangular shape of the lower part of the wind farm (wind turbines 58 to 80)~~as well as~~ , and the most northern wind turbines 1 to 8 are ~~incorporated now~~ now incorporated (cf. Fig. 1) to identify the flow conditions ~~there~~of the whole wind farm. This results in 6,985,830 considered time intervals for 90° and 4,914,448 considered time intervals for 270°. ~~The clustering~~ Clustering is performed using the k-means algorithm of MATLAB (MATLAB, 2019) based on Lloyd (1982) ~~,~~ using random sample points as initial centroids to find the best solution. ~~To~~ Clustering is repeated ten times to avoid the generation of local centroids~~the clustering is repeated ten times~~ , and the run with the clusters with the lowest sum of point-to-centroid distances within the clusters is chosen. As a distance metric for the clusters, the squared Euclidean distance is chosen. The maximum number of iterations is set to ~~300.~~ 300 and k is set to five clusters. This number was empirically chosen as the data was grouped into a reasonable set of groups with clearly distinguishable correlation curves (correlation states). A greater number of clusters lead to further clusters with similar correlation curves. The only difference found was in the standard deviation of the power output fluctuations of the wind turbine pairs. Here, the cluster ~~indicate~~ indicates a higher standard deviation of the power output fluctuations for the upstream wind turbine A instead of the downstream wind turbine B. This slightly abnormal behaviour is shown in more detail in ~~appendix~~Appendix C. Also, different orderings of the intervals have been tested, namely random sorting, data sorted for an increasing standard deviation of the power output fluctuations of the downstream wind turbine B, and ~~a~~ chronological sorting according to the available time intervals. The results ~~have been found to be equal~~are equal, including the first decimal place of the centroids for all cases~~, thus, the~~ . Thus, a random sorting is used in ~~the~~ further analysis.

Table 4 lists the determined centroids (centres of the clusters) for wind directions 90° and 270°. ~~The standard deviations of~~ Standard deviations of the power output fluctuations of both wind turbines A and B ~~are significantly decreasing~~ significantly

decrease while the normalised power difference of A and B is significantly increasing from ~~cluster~~ Cluster 1 to 5. To further investigate these findings, we analyse the correlation curves corresponding to the clusters. Figures 7 and 8 show the average correlations for both wind directions for each of the five clusters (upper plots) and the percentage frequency of each pair within each of the five clusters (lower plots). As expected from Fig. 4 and 5, the average correlations for 270° are higher than those for 90°. Cluster 1 includes nearly 6% of the data and has the highest correlation. This is a significant increase compared to the average correlation shown in Fig. 5. ~~From cluster 2 to 4 the correlation is decreasing~~ The correlation decreases while the amount of data per cluster increases ~~. For cluster 5 no~~ from Cluster 2 to 4. No correlation is found ~~. Looking~~ for Cluster 5. A clear trend is visible upon looking at the occurrence of wind turbine pairs within each cluster~~, a clear trend is visible. While cluster~~ . While Cluster 1 with the highest correlation is dominated by wind turbine pairs, where the upstream wind turbine is located towards the back of the wind farm, ~~cluster~~ Cluster 5 with no correlation is dominated by wind turbine pairs with its upstream wind turbine located in the first row of the wind farm. From ~~Cluster~~ Clusters 2 to 4, the dominating wind turbine pairs shift from the back rows towards the front rows, whereas the percentage frequency ~~become~~ becomes more balanced throughout the wind farm (i.e. more light green coloured turbines).

~~Comparison~~ The comparison of the results of Fig. 7 and 8 and Tab. 4 ~~clearly~~ depicts that the greater the standard deviations of the power output fluctuations and the smaller the normalised power difference of the wind turbines in a pair, the higher the correlation ~~of~~ for ´ the wind turbine pairs. The slight row dependence, which was already indicated in Tab. 3, can be confirmed here. This is illustrated by a colour coding of the frequency of occurrence of wind turbine pairs in each cluster in the lower subplot of Fig. 7 (respectively Fig. 8). The sum of all frequencies of all wind turbines within one cluster ~~add~~ adds up to 100%, meaning a yellow coloured wind turbine pair makes up about 3% of the respective cluster~~and a green marked~~ , and a green-marked wind turbine pair makes up about 1.5% of the respective cluster. For example, the correlation peak for

Table 4. Cluster centroids for wind direction intervals around 90° and 270° with A as the upstream wind turbine and B as the downstream wind turbine. $\sqrt{\langle P_A'^2 \rangle_{\Delta t_{600}}}$ is the standard deviation of the power ~~outputs~~ output fluctuations of wind turbine A over 600 s intervals Δt_{600} (analogue for wind turbine B for the same 600 s intervals, respectively). $\langle P_A \rangle_{\Delta t_{600}}$ and $\langle P_B \rangle_{\Delta t_{600}}$ are the average power output of wind turbines A and B over the same 600 s intervals. $\langle ... \rangle_{cluster}$ denotes the average of the statistics over all available time intervals of the wind turbine pairs within a cluster. Note that 90° and 270° again refer to 20° wind direction intervals from 80° to 100° and from 260° to 280°.

	$\left\langle \sqrt{\langle P_A'^2 \rangle_{\Delta t_{600}}} \right\rangle_{cluster}$ [kW]		$\left\langle \sqrt{\langle P_B'^2 \rangle_{\Delta t_{600}}} \right\rangle_{cluster}$ [kW]		$\left\langle \frac{\langle P_A \rangle_{\Delta t_{600}} - \langle P_B \rangle_{\Delta t_{600}}}{\langle P_A \rangle_{\Delta t_{600}}} \right\rangle_{cluster}$	
Cluster	90°	270°	90°	270°	90°	270°
1	513	535	527	540	0.02	0.05
2	367	381	387	405	0.05	0.06
3	253	283	271	298	0.10	0.06
4	157	197	174	213	0.13	0.07
5	73	117	86	133	0.14	0.11

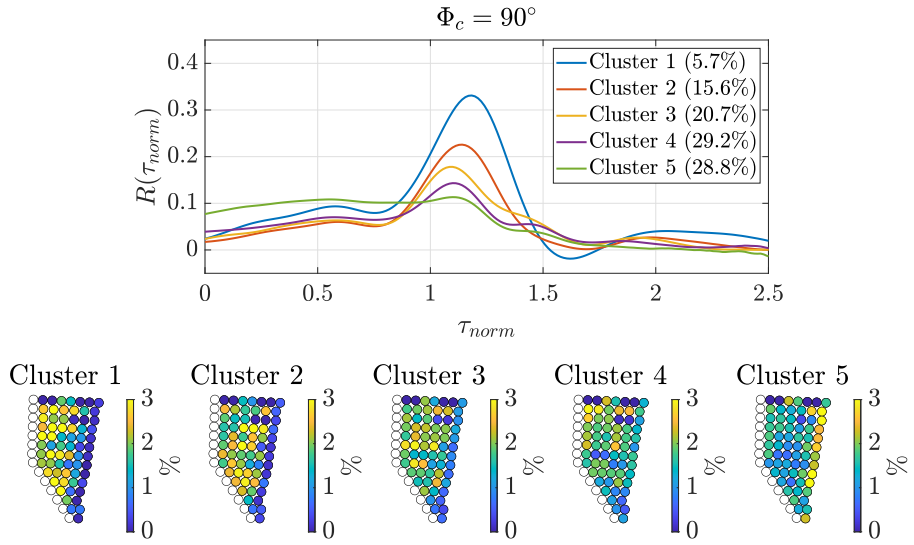


Figure 7. Clustering for wind direction interval around 90° with randomly sorted 600 s time intervals. Φ_c depicts the centre of the wind direction interval. The upper plot shows the average power output fluctuation correlation curve per cluster. The legend lists the share of data. The lower plot shows the percentage frequency of each wind turbine pair within the respective cluster given as colour. As the wind turbines are analysed in pairs of two, the last row of wind turbines is unlabelled, as these wind turbines do not have a downstream partner.

~~cluster~~Cluster 1 of more than 0.3 for 90° (respectively 0.4 for 270°) partly includes pairs with the upstream turbine in the last row ~~but also~~ and some turbine pairs in the rows before. This is considerably larger than the correlation curve of row 6 of Fig. 6.

5 Conclusions

We presented an approach to analyse the correlations of power output fluctuations of wind turbine pairs for 600 s time intervals based on 1 Hz SCADA data, which copes with the challenge of highly variable flow conditions in the measurement data and the identification of correlation states. Further, we investigated different influences on the correlation of power output fluctuations of wind turbine pairs. The investigation of the influence of different wind directions on the correlations of power output fluctuations of wind turbine pairs showed that ~~streamwise aligned~~ streamwise-aligned pairs are correlated~~while~~. In contrast, spanwise pairs show nearly no correlation of power output fluctuations. Thus, we focused our investigation on the streamwise wind turbine pairs.

Inspired by the findings of Bossuyt et al. (2017b), which showed an increasing turbulence intensity throughout the wind farm and the model for velocity space-time correlations by Lukassen et al. (2018), we introduced and evaluated parameters to characterise correlation states of power output fluctuations. The chosen parameters were the standard deviations of the power output fluctuations and the normalised power difference of wind turbines in a pair.

In general, we found ~~higher and more defined~~ that the averaged correlation curves of power output fluctuations for 270° with a

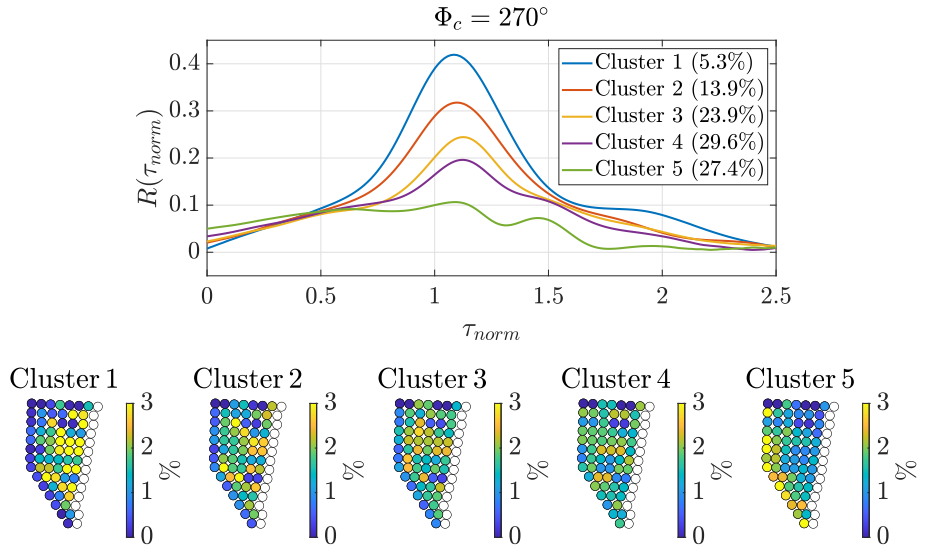


Figure 8. Clustering for wind direction interval around 270° with randomly sorted 600 s time intervals. Φ_c depicts the centre of the wind direction interval. The upper plot shows the average power output fluctuation correlation curve per cluster. The legend lists the share of data. Note that the values do not add up exactly to 100% due to rounding. The lower plot shows the percentage frequency of each wind turbine pair within the respective cluster given as colour. As the wind turbines are analysed in pairs of two, the last row of wind turbines is unlabelled, as these wind turbines do not have a downstream partner.

maximum correlation coefficient of 0.21 ~~in comparison to averaged correlations~~ have more defined (narrower) peaks compared to those of the averaged correlation curves for 90° with a maximum correlation coefficient of 0.16. ~~Regarding the introduced parameters~~Further, the standard deviation of the power output fluctuations of the wind turbines in a pair was larger for 270° ~~was found larger~~ than for 90°. This difference, together with the slightly asymmetric layout of the wind farm and different inflow conditions for 90° and 270°, are most likely the root causes for this deviation in the correlation curves. In the context of the considered highly varying flow conditions, peak correlations around 0.21 or 0.16 are still considered significant. The cause for these relatively low peak correlations lies in the varying flow conditions or noisiness of the flow within the wind farm.

The investigation of the average correlation ~~of~~ for wind turbine pairs per wind farm row strengthened our previous findings. We found different correlation curves for the rows of the wind farm, becoming more defined (more narrow peaks) towards the back of the wind farm. Wind turbine pairs, where the upstream wind turbine A is located in the first row and the downstream wind turbine B is located in the second row of the wind farm, show no correlation~~with~~. In addition, large normalised power differences of the wind turbines in a pair and small standard deviations of power output fluctuations were observed. This is most likely caused by the ~~free-stream~~ free stream inflow of the upstream wind ~~turbines~~ turbine A of the pairs. Most importantly, the analysis of the separate rows of the wind farm revealed a trend of increasing standard deviations of the power output fluctuations throughout the wind farm and ~~decreasing power difference. Due to the high variability in~~ a decreasing normalised power difference of the wind turbines in a pair. As mentioned before, the flow throughout the wind farm ~~,~~ is highly variable

due to the individual control and operation of the wind turbines. This means that not all wind turbine pairs in the same row are affected by the same flow conditions, ~~i.e.~~ as upstream wind turbines could be turned off, could be yawing or could be pitching. This further means that they show different correlation curves and ~~thus,~~ should be sorted ~~to~~ into different correlation states. Thus, to group data according to the underlying flow conditions~~which define~~ , which define the different correlation states, the introduced parameters (standard deviation of the power output fluctuations of wind turbines in a pair and the normalised power difference of wind turbines in a pair) ~~where~~ were combined with the clustering algorithm k-means. The clustering showed similar results for the wind directions 90° and 270°~~and the clusters showed clearly distinguishable parameters~~ . The clusters had distinguishable values in the standard deviation of the power output fluctuations and in the normalised power output differences of the wind turbines, which were directly related to the average correlation curve per cluster. Increased standard deviations ~~combined with small power differences~~ of the power output fluctuations combined with the small normalised power difference of the wind turbines in a pair showed the most defined correlations with the highest peak (Cluster 1). This combination was found for wind turbine pairs ~~with a position more~~ located further downstream in the wind farm ~~but also including~~ and some wind turbine pairs from rows towards the front. For 90°, the peak of the correlation increased via clustering from 0.16 to 0.32, and for 270° the peak of the correlation increased from 0.21 to 0.41. A value of 0.41 is close to the correlations found in the LES study by Lukassen et al. (2018) and experiments by Bossuyt et al. (2017b), which were between 0.5 and 0.55 for similar wind turbine spacing and similar wind speeds. In addition, for both wind directions, a cluster of non-correlated wind turbines ~~was found. It~~ (Cluster 5) was found, which mainly consists of wind turbine pairs in the first rows of the wind farm. ~~The remaining clusters~~ Clusters 2, 3 and 4 were not as significant as ~~the other but also showed clearly~~ Clusters 1 and 5 and showed distinguishable correlation curves with their peaks ~~in the range~~ ranging from 0.14 to 0.22 for 90° and from 0.2 to 0.31 for 270°.

Hence, we found that to analyse correlation states of power output fluctuations of streamwise wind turbine pairs in varying flow conditions, the standard deviation of the power output fluctuations of wind turbines in a pair as well as the normalised power difference of the wind turbines in a pair have been proven to be suitable ~~parameters~~ to identify correlation states. Furthermore, the data-driven k-means clustering approach enables an automated grouping of the data into correlation states based on these parameters. As an outlook, further analysis ~~on~~ of the space-time correlations within an offshore wind farm could help ~~in the control of~~ control wind turbines, e.g. ~~-~~for power output fluctuation management or active wake control. Also, knowledge about the correlation of wind turbine pairs allows short-term power output fluctuation forecasting within the wind farm ~~as well as~~ and interactive wind turbine control.

The presented findings can be enhanced in the future by ~~additional Lidar or Radar~~ adding Lidar or radar measurements to access independent wind direction and wind speed measurements. ~~Also~~Moreover, the analysis of correlation states might be extended to include the correlation of wind turbine pairs with multiple inter-turbine distances and the correlation of non-aligned wind turbine pairs. ~~The clustering~~ Clustering of correlation states can be further investigated by increasing the number of clusters to $k > 5$~~as the results~~ . Results for $k = 6$ indicated that the statistics of the upstream and downstream wind turbine of a pair ~~has~~ have a different influence on its correlation. ~~Also~~In addition, it is worth considering alternative clustering methods like

k-medoids (Kaufman and Rousseeuw, 2008), which is less sensitive to outliers ~~compared to *k*-means~~ or Density-Based Spatial Clustering of Applications with Noise (DBSCAN) (Ester et al., 1996) which is also less sensitive to outliers and has no fixed cluster shapes and ~~no predefined~~ fixed number of clusters~~like *k*-means~~. Using these algorithms could improve the clustering of the intervals and more defined correlation curves or could identify further clusters. Furthermore, measurements on the boundary layer conditions help ~~to~~ assess the influence of wind turbine wakes on the space-time correlations of power output fluctuations with the additional knowledge on the atmospheric stability.

Appendix A: Wind turbine pairs

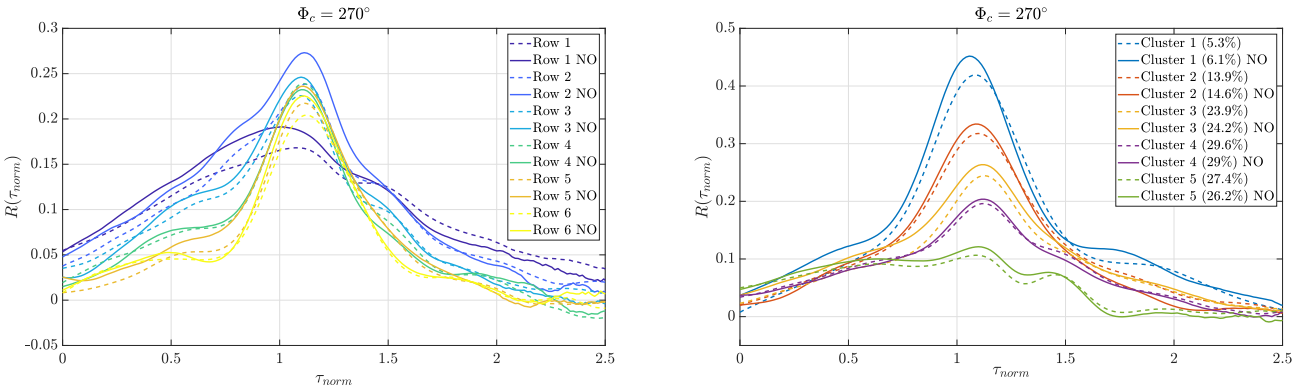
~~In order to~~ To calculate the power output fluctuation correlation, wind turbine pairs are chosen according to the respective wind direction. In total, 66 wind turbine streamwise pairs can be defined. Table A1 depicts the definition of wind turbine pairs for wind ~~directions~~ direction 270°. For wind direction 90°, the same pairs are chosen but with switched wind turbine order. E.g. for pair 1 for 270°, wind turbine 1 is the upstream wind turbine and wind turbine 2 is the downstream wind turbine~~, respectively for~~ . For 90°, wind turbine 2 is the upstream wind turbine and turbine 1 is the downstream wind turbine.

Table A1. Definition of streamwise wind turbine pairs for wind direction 270°.

Pair	01	02	03	04	05	06	07	08	09	10	11	12	13	14
WTs	01, 02	02, 03	03, 04	04, 05	05, 06	06, 07	07, 08	09, 10	10, 11	11, 12	12, 13	13, 14	14, 15	16, 17
Pair	15	16	17	18	19	20	21	22	23	24	25	26	27	28
WTs	17, 18	18, 19	19, 20	20, 21	21, 22	23, 24	24, 25	25, 26	26, 27	27, 28	28, 29	30, 31	31, 32	32, 33
Pair	29	30	31	32	33	34	35	36	37	38	39	40	41	42
WTs	33, 34	34, 35	35, 36	37, 38	38, 39	39, 40	40, 41	41, 42	42, 43	44, 45	45, 46	46, 47	47, 48	48, 49
Pair	43	44	45	46	47	48	49	50	51	52	53	54	55	56
WTs	49, 50	51, 52	52, 53	53, 54	54, 55	55, 56	56, 57	58, 59	59, 60	60, 61	61, 62	62, 63	64, 65	65, 66
Pair	57	58	59	60	61	62	63	64	65	66				
WTs	66, 67	67, 68	69, 70	70, 71	71, 72	73, 74	74, 75	76, 77	77, 78	79, 80				

Appendix B: Statistical dependence of 600 s intervals

The analysed 600 s intervals are not statistically independent as they overlap by 599 s in the extreme case. Also, thinking of bigger gusts evolving through the wind farm, it is most likely that wind turbine pairs experience similar correlation states when being affected by the gust. To clarify the influence of the overlapping of the considered intervals, we performed the calculations again using only non-overlapping intervals. The following figure (Fig. B1) compares the results for non-overlapping and overlapping intervals exemplary for wind direction 270°. Figure B1a displays the comparison of the average correlation curves per wind farm row. Figure B1b displays the comparison of the average correlation curves per cluster. In general, the results of the non-overlapping intervals are similar to the results of the overlapping intervals and differ at most by about 10%. However, Fig. B1a shows that this data set is at the limit of representing the correlations as the average correlation curves start to wiggle for $\tau_{norm} > 2$ due to the low number of data points. In total, only 8121 non-overlapping 600 s intervals are available for 270°, which resemble a measurement time of about 56 days. For all wind turbine pairs, 11514 intervals are available as multiple wind turbine pairs are available in the same interval.



(a) Average power output fluctuation correlation per wind farm row.

(b) Average power output fluctuation correlation per cluster.

Figure B1. Comparison of the average power output fluctuation correlation for wind direction interval 270° of non-overlapping and overlapping intervals. For both plots, the average correlation curves for non-overlapping intervals are marked with 'NO' and plotted with a dashed line. The average correlation curves for overlapping intervals are plotted in both cases as a solid line.

Appendix C: Effect of the numbers of clusters

As mentioned in Sect. 4, the number of clusters chosen for the present analysis was k = 5. This decision was made based on the results for k = 6 presented in Figure C1 and Fig. C2. For wind direction 90°, six ~~clearly~~ separable correlation curves are found. Comparing Fig. C1 to Fig. 7, it shows that Cluster 2 of Fig. 7 seems to be separated into ~~to~~ two clusters (Cluster 2 and 3 of Fig. C1).

For wind direction 270°~~only 5~~, only five clearly separable correlation curves are found ~~whereas~~ where one is overlapped by a very similar one. Comparing Fig. C2 to Fig. 8, it shows that Cluster 3 of Fig. 7 seems to be separated into two similar clusters (Cluster 3 and 4 of Fig. C2). The new clusters also do not reveal any further characteristics.

Looking at the statistics of the correlation curves listed in Tab. C1~~it further~~, it can be found that for wind direction 90°, Cluster 2 shows a higher standard deviation of the power output fluctuations for wind turbine A instead of B, while Cluster 3 shows a higher standard deviation of the power output fluctuations for wind turbine B instead of A, similar to all other ~~Clusters. For~~ clusters. For the wind direction 270°, Cluster 4 shows a higher standard deviation of power output fluctuations for wind turbine A instead of B, while Cluster 3 shows a higher standard deviation of power output fluctuations for wind turbine B instead of A, similar to all other ~~Clusters~~clusters.

The correlation curves and statistics imply that ~~a~~ the further separation of the statistics with $k > 5$ does not reveal any correlation states ~~which~~ that are more significant than ~~the ones~~ those found for $k = 5$. ~~However, clustering~~ Clustering with $k > 5$ might ~~result in a further distinction of~~ further distinguish the flow states for wind turbine pairs based on the standard deviations of power output fluctuations of wind turbines A and B~~but are~~. However, this is not further investigated, as this effect is not included in the scope of the work presented here.

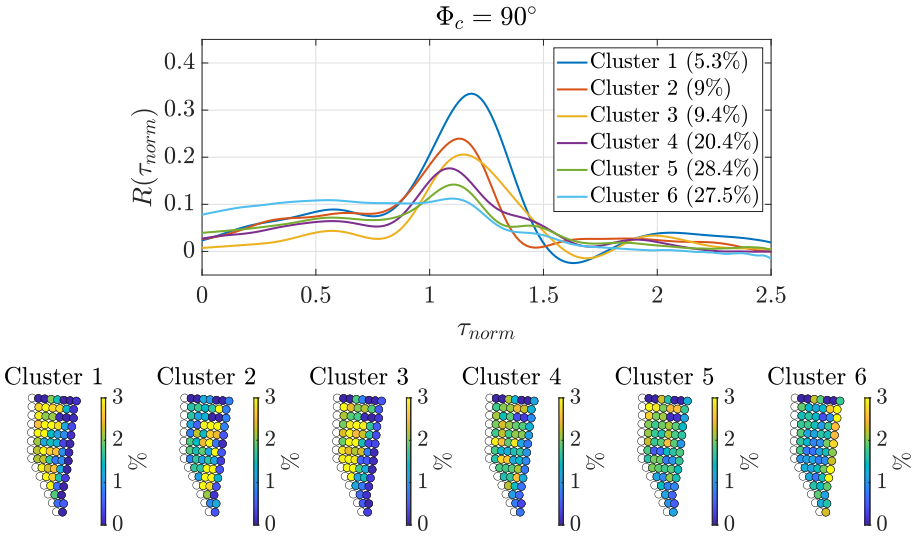


Figure C1. Clustering for wind direction interval around 90° with randomly sorted parameters and $k = 6$. Φ_c depicts the centre of the wind direction interval.

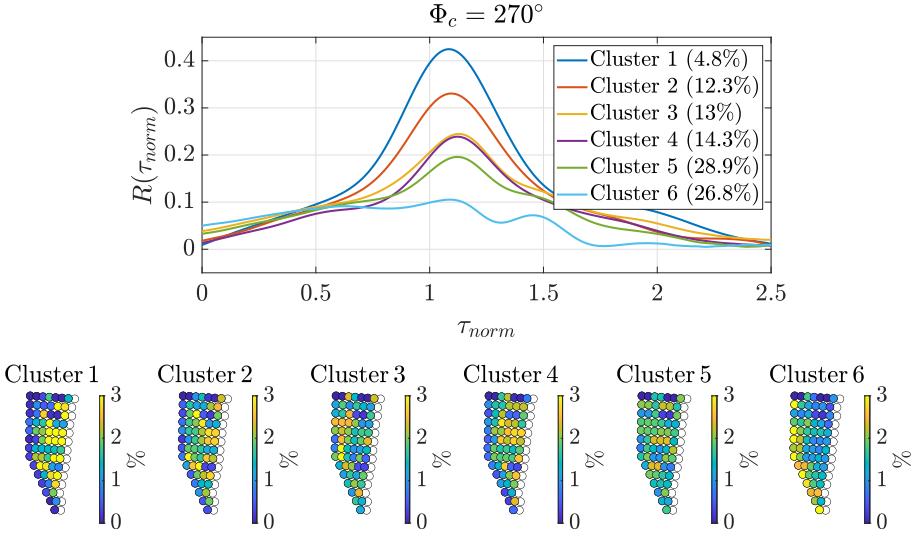


Figure C2. Clustering for wind direction interval around 270° with randomly sorted parameters and $k = 6$. Φ_c depicts the centre of the wind direction interval.

Table C1. Averaged wind turbine statistics computed for wind direction intervals around 90° and 270° and k = 6, with A as the upstream wind turbine and B as the downstream wind turbine. $\sqrt{\langle P_A'^2 \rangle_{\Delta t_{600}}}$ is the standard deviation of the power output fluctuations of wind turbine A over a 600 s interval Δt_{600} (analogue for wind turbine B for the same 600 s intervals, respectively). $\langle P_A \rangle_{\Delta t_{600}}$ and $\langle P_B \rangle_{\Delta t_{600}}$ are the average power of wind turbines A and B over the same 600 s intervals. $\langle ... \rangle_{cluster}$ denotes the average of the statistics over all available time intervals of the wind turbine pairs. Note that here 90° and 270° again refer to 20° wind direction intervals from 80° to 100° and from 260° to 280°.

	$\left\langle \sqrt{\langle P_A'^2 \rangle_{\Delta t_{600}}} \right\rangle_{cluster}$ [kW]		$\left\langle \sqrt{\langle P_B'^2 \rangle_{\Delta t_{600}}} \right\rangle_{cluster}$ [kW]		$\left\langle \frac{\langle P_A \rangle_{\Delta t_{600}} - \langle P_B \rangle_{\Delta t_{600}}}{\langle P_A \rangle_{\Delta t_{600}}} \right\rangle_{cluster}$	
Cluster	90°	270°	90°	270°	90°	270°
1	523	541	526	547	0.02	0.04
2	327	394	435	412	0.02	0.06
3	393	247	324	343	0.09	0.06
4	240	317	263	263	0.11	0.07
5	152	194	168	211	0.13	0.07
6	72	116	84	132	0.14	0.11

Author contributions. JKS developed the underlying method, performed the data analyses and wrote the paper. LJL provided intensive consultation on the development of the method and the scientific analyses. MKr and MKü provided intensive reviews on the scientific analyses. LJL had a supervising function.

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

Acknowledgements. We performed parts of the work within the research project ~~"OWP Control" ~~'OWP Control' (FKZ 0324131A) funded by the German Ministry of Economic Affairs and Energy basis of a decision by the German Bundestag. Parts of the computations were performed on the ~~high performance~~ high-performance computing system EDDY of the University of Oldenburg which is part of the project 'WIMS-Cluster' (FKZ 0324005) founded by the Federal Ministry of Economic Affairs and Energy. We acknowledge the wind farm operator Global Tech I Offshore Wind GmbH for providing SCADA data and their support of the work.

References

Andersen, S. J., Sørensen, J. N., and Mikkelsen, R. F.: Turbulence and entrainment length scales in large wind farms, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 375, 20160 107, https://doi.org/10.1098/rsta.2016.0107, 2017.

Bossuyt, J., Howland, M. F., Meneveau, C., and Meyers, J.: Measurement of unsteady loading and power output variability in a micro wind farm model in a wind tunnel, Experiments in Fluids, 58, 1–17, https://doi.org/10.1007/s00348-016-2278-6, 2017a.

Bossuyt, J., Meneveau, C., and Meyers, J.: Wind farm power fluctuations and spatial sampling of turbulent boundary layers, Journal of Fluid Mechanics, 823, 329–344, https://doi.org/10.1017/jfm.2017.328, 2017b.

Braun, T., Waechter, M., Peinke, J., and Guhr, T.: Correlated power time series of individual wind turbines: A data driven model approach, Journal of Renewable and Sustainable Energy, 12, 023 301, https://doi.org/10.1063/1.5139039, 2020.

Bromm, M., Rott, A., Beck, H., Vollmer, L., Steinfeld, G., and Kühn, M.: Field investigation on the influence of yaw misalignment on the propagation of wind turbine wakes, Wind Energy, 21, 1011–1028, https://doi.org/10.1002/we.2210, 2018.

Crespo, A. and Hernàndez, 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.

Dai, J., Cao, J., Liu, D., Wen, L., and Long, X.: Power fluctuation evaluation of large-scale wind turbines based on SCADA data, IET Renewable Power Generation, 11, 395–402, https://doi.org/10.1049/iet-rpg.2016.0124, 2017.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X.: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, in: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD'96, p. 226–231, AAAI Press, 1996.

Kaufman, L. and Rousseeuw, P.: Partitioning Around Medoids (Program PAM), chap. 2, pp. 68–125, John Wiley & Sons, Ltd, https://doi.org/10.1002/9780470316801.ch2, 2008.

Komusanac, I., Brindley, G., and Fraile, D.: Wind energy in Europe in 2019 Trends and statistics, https://windeurope.org/wp-content/uploads/files/about-wind/statistics/WindEurope-Annual-Statistics-2019.pdf, last access 21.01.2021, 2020.

Lloyd, S. P.: Least squares quantization in PCM, IEEE Transactions on Information Theory, 28, 129–137, https://doi.org/10.1002/9780470316801.ch2, 1982.

Lukassen, L. J., Stevens, R. J. A. M., Meneveau, C., and Wilczek, M.: Modeling space-time correlations of velocity fluctuations in wind farms, Wind Energy, 21, 474–487, https://doi.org/10.1002/we.2172, 2018.

MATLAB: version 9.7.0.1190202 (R2019b), The MathWorks Inc., Natick, Massachusetts, 2019.

Pearson, K.: Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity, and Panmixia, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 187, 253–318, https://doi.org/10.1098/rsta.1896.0007, 1896.

Porté-Agel, F., Bastankhah, M., and Shamsoddin, S.: Wind-Turbine and Wind-Farm Flows: A Review, Boundary-Layer Meteorology, 174, https://doi.org/10.1007/s10546-019-00473-0, 2020.

Ramírez, L., Fraile, D., and Brindley, G.: Offshore Wind in Europe Key trends and statistics 2019, https://windeurope.org/wp-content/uploads/files/about-wind/statistics/WindEurope-Annual-Offshore-Statistics-2019.pdf, last access 21.01.2021, 2020.

Ren, G., Liu, J., Wan, J., Guo, Y., and Yu, D.: Overview of wind power intermittency: Impacts, measurements, and mitigation solutions, Applied Energy, 204, 47–65, https://doi.org/10.1016/j.apenergy.2017.06.098, 2017.

Sanchez Gomez, M. and Lundquist, J. K.: The effect of wind direction shear on turbine performance in a wind farm in central Iowa, Wind Energy Science, 5, 125–139, https://doi.org/10.5194/wes-5-125-2020, 2020.

Schneemann, J., Rott, A., Dörenkämper, M., Steinfeld, G., and Kühn, M.: Cluster wakes impact on a far-distant offshore wind farm's power, Wind Energy Science, 5, 29–49, https://doi.org/10.5194/wes-5-29-2020, 2020.

Sorensen, P., Cutululis, N. A., Vigueras-Rodriguez, A., Jensen, L. E., Hjerrild, J., Donovan, M. H., and Madsen, H.: Power Fluctuations From Large Wind Farms, IEEE Transactions on Power Systems, 22, 958–965, https://doi.org/10.1109/TPWRS.2007.901615, 2007.

Stevens, R. J. A. M. and Meneveau, C.: Temporal structure of aggregate power fluctuations in large-eddy simulations of extended wind-farms, Journal of Renewable and Sustainable Energy, 6, 043 102, https://doi.org/10.1063/1.4885114, 2014.

Taylor, G. I.: The Spectrum of Turbulence, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 164, 476–490, https://doi.org/10.1098/rspa.1938.0032, https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1938.0032https://royalsocietypublishing.org/doi/10.1098/rspa.1938.0032, 1938.

Vali, M., Petrović, V., Steinfeld, G., Y. Pao, L., and Kühn, M.: An active power control approach for wake-induced load alleviation in a fully developed wind farm boundary layer, Wind Energy Science, 4, 139–161, https://doi.org/10.5194/wes-4-139-2019, 2019.

Valldecabres, L., von Bremen, L., and Kühn, M.: Minute-scale detection and probabilistic prediction of offshore wind turbine power ramps using dual-Doppler radar, Wind Energy, 23, 2202–2224, https://doi.org/https://doi.org/10.1002/we.2553, https://onlinelibrary.wiley.com/doi/abs/10.1002/we.2553, 2020.

Vermeer, L., Sørensen, J., and Crespo, A.: Wind turbine wake aerodynamics, Progress in Aerospace Sciences, 39, 467–510, https://doi.org/10.1016/S0376-0421(03)00078-2, 2003.