https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Impact of wind farms on the occurrence and characteristics of low-level jets

Coleman Moss[1], Rupert Storey[2], Tanmoy Chatterjee[3], and Giacomo Valerio Iungo[1]

[1]Jonsson School of Engineering and Computer Science, The University of Texas at Dallas, Richardson, TX 75080
[2]Onshore Wind, GE Vernova, Berlin
[3]Advanced Research Center, GE Vernova, Niskayuna, NY 12309

Correspondence: Giacomo Valerio Iungo (valerio.iungo@utdallas.edu)

Abstract. The low-level jet (LLJ) is an occurrence of super-geostrophic wind speeds in the lower regions of the atmosphere, typically below 1 km. LLJs are highly relevant for wind energy as they can substantially modulate power production, wake recovery, and loads of wind turbines and wind farms. However, it is not clear how the presence of wind farms affects the occurrence rate and characteristics of LLJs, with some simulations suggesting that the presence of farms leads to fewer LLJs and to LLJs occurring at higher elevations. In this work, we develop methods to take advantage of unique long-term wind profile measurements to experimentally assess the impact of wind farms on the occurrence rates and morphologies of LLJs. We use more than 27 years of profiling radar wind measurements collected for the period August 1998–April 2026 at the Atmospheric Radiation Measurement Southern Great Plains site in Lamont, Oklahoma, with over ten large wind farms (each comprising up to 100 turbines) installed near the measurement site at ranges from 17 km to 67 km, with the closest (and largest) installations just to the east, west, and south. We apply both traditional LLJ detection methods (i.e., drop-off and shear) and a data-driven agglomerative-clustering method to hourly-averaged wind speed profiles. The aggregate LLJ occurrence rates are essentially indistinguishable pre- and post-construction (approximately 40% by the shear method and 36% by the drop-off method), with a transient reduction of approximately 3% (shear method) and 4.5% (drop-off method) during the construction era (2012–2020). Data-driven analysis further resolves sector- and time-of-day-specific shifts: in all wind sectors examined, including a sector with no upstream farms, clusters with strong near-ground shear seem to be replaced by clusters with reduced near-ground shear. However, LLJ rates are not significantly affected, although the results suggest a stronger modification at night (local time 0–6, reductions of around 50% for specific clusters). Given concerns about long-term background climatological variability, it is very difficult to ensure that these results are statistically significant. In contrast to previous simulations, we find no statistically significant evidence of wind farm modifications of LLJ rates.

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

1 Introduction

Winds in the atmospheric boundary layer (ABL), the region of the atmosphere closest to the Earth's surface, typically blow slowly near the ground and faster above. Under many conditions, wind speeds increase monotonically with height, resembling classical wall-bounded boundary layer flows. However, measurements across the world frequently detect cases of maximum wind speeds closer to the ground with slower wind speeds above (Doosttalab et al., 2020). These cases of super-geostrophic wind speeds at lower elevations, typically between 50 m and 2 km, are referred to in the literature as low-level jets (LLJs). As a meteorological phenomenon, the LLJ is important for aerosol transport, pollution transport, cloud formation, forest fire spread, and storm occurrence (Stensrud, 1996; Whiteman et al., 1997; Banta et al., 2002). In addition to these meteorological considerations, LLJs dramatically influence wind energy resources.

As wind turbine heights and rotor sizes consistently increase over time, modern rotor layers often intersect LLJ altitudes. Since LLJs are associated with increased wind speed, they have the potential to increase wind power production (Sheridan et al., 2024). Depending on the relative heights of the turbine hub and LLJ, the LLJ can enhance or damp the inflow turbulence within the rotor layer as well as wake recovery (Gadde and Stevens, 2021). Given the large positive and negative shear below and above the LLJs, LLJs usually increase power fluctuations in turbines and increase loads (Paulsen et al., 2025; Chatterjee et al., 2025). LLJs occurring above wind farms also increase the available momentum above the wind farm. Since the total power a wind farm produces is largely determined by the amount of energy it can extract from the atmosphere above the farm, LLJs have been suggested to increase the overall farm power by enhancing the downward momentum flux available for wake recovery and rear-row turbines (Meneveau, 2012; Doosttalab et al., 2020).

However, the increased footprint of individual wind farms and larger clusters of wind farms could also impact the occurrence and characteristics of LLJs. Conceiving of wind farms as roughness elements, several works have described the modulation of the atmosphere due to the presence of infinite and finite wind farms, typically under neutral atmospheric conditions (Frandsen et al., 2006; Calaf et al., 2010; Meneveau, 2012). Other works conceive of wind farms as a canopy layer (Markfort et al., 2012, 2018). In both cases, as the flow develops over the wind farm, the momentum extracted by the rotors increasingly comes from the atmosphere above the farm rather than from the farm inlet (Segalini and Chericoni, 2021). When LLJs exist over wind farms, the wind farms might reduce the intensity of the LLJ by extracting the increased momentum above the farm. Indeed, simulations of offshore conditions have suggested that LLJ occurrence rates decrease near wind farms and that nose heights increase (Quint et al., 2025). Figure 1 illustrates schematically how this interaction might look. While the schematic shows the modification of the LLJ over a wind farm, we also want to consider possible disruptions of the LLJ formation mechanisms by the presence of farms, leading to a larger-scale modification of LLJ characteristics beyond the farm footprint.

There are therefore two competing interactions. On the one hand, the presence of LLJs could increase wind farm power production. However, the presence of wind farms could reduce the occurrence of LLJs and shift the high-momentum region higher, perhaps reducing its beneficial impacts on power. When developing wind farms in regions with high LLJ occurrence, such as the central United States, it is critical to understand both of these interactions (Whiteman et al., 1997). To accurately estimate the power production of future farms in LLJ-impacted regions, we should develop a better understanding of these

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

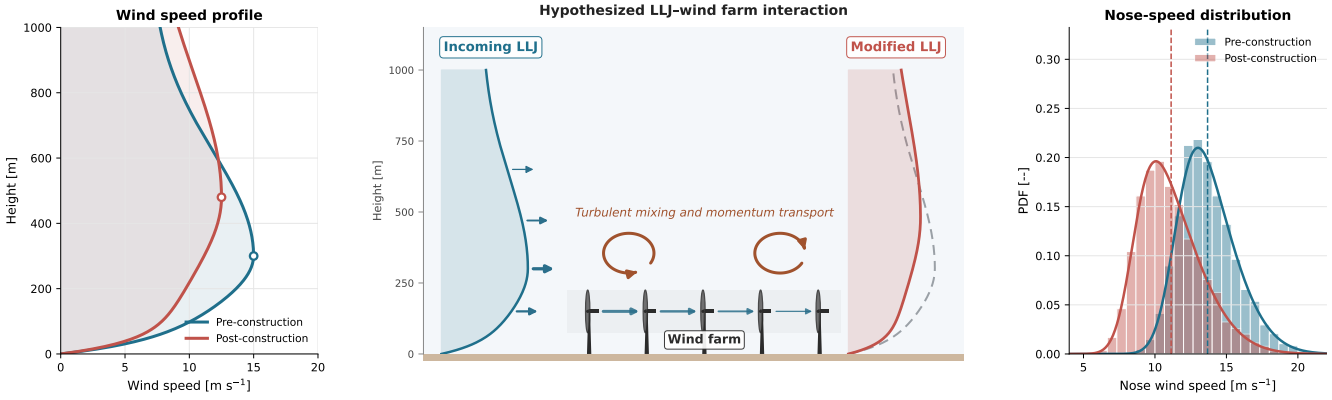

Figure 1. An illustration of the coupling between wind farms and LLJs resulting in reduced LLJ nose wind speeds and higher LLJ noses.

two interactions. If the LLJ increases the overall power but is reduced in frequency by the wind farm, then a pre-construction annual energy production (AEP) estimate that accounts for positive LLJ power impacts but not the negative farm effects will over-estimate the AEP (compared to AEP estimates using historical LLJ rates). Conversely, if LLJs decrease overall power and are reduced by wind farm presence, then a model that predicts AEP using historical LLJ rates and accounts for negative impacts on power will under-estimate the AEP, as the negative impacts will occur less often than expected. Although the experimental assessment of the impact of LLJs on wind farm power is ongoing, here we tackle the question of wind farm modulation of LLJs, which, to our knowledge, has not yet been considered experimentally.

Experimentally assessing the impact of wind farm construction on the occurrence rate and morphology of LLJs poses three main challenges. First, long-term historical data are needed, with careful attention to quality control and assessment of data variability, to separate statistically significant variations from background trends and noise. Second, the LLJ must be defined and a detection method developed. Third, a more detailed analysis is needed to study the morphology of the wind profiles beyond the simple LLJ detection methods. In this work, we tackle each of these challenges sequentially.

Given the importance of LLJs on meteorology and wind energy, it is surprising that there is a high degree of variability in the definitions and detection mechanisms of LLJs. In characterizing an LLJ, there are several features to consider. First, by definition, the LLJ is a local maximum wind speed at relatively low elevations. Therefore, there must be an identifiable wind speed maximum, often referred to as the LLJ nose (or core), with a measured decrease in wind speed above. As a result, measurements over limited heights (e.g., up to 1 km) with the wind speed maximum at the highest measurement height are typically not classified as LLJs under definitions requiring a decrease of wind speed above the nose of a certain magnitude. Having identified the nose, we can add additional characteristics. The height of the maximum wind speed is referred to as the nose height. If desired, constraints can be placed on the nose height. In particular, nose heights over the rotor area of wind turbines are especially important for wind energy applications. Having identified the nose, we can further identify the height and wind speed of the local minimum above the nose. The difference between the two wind speeds is often termed the drop-off (Debnath et al., 2023). Many LLJ definitions set a minimum threshold on the drop-off. If the drop-off is normalized by the

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

LLJ nose speed, the result is often termed the LLJ intensity, which can be a further constraint for LLJ identification methods. Finally, by virtue of being a local maximum in wind speed, wind shear will be positive below the nose and negative above. Constraints can be placed on the shear to ensure that this positive/negative pattern occurs with a certain magnitude.

Reviewing detection methods in the literature, we find a broad range of acceptable thresholds and combinations of characteristics. For example, early work split LLJs into categories, each category requiring a minimum LLJ nose speed and LLJ drop-off (Whiteman et al., 1997). Recent works have used a combination of limits on drop-off and intensity (Debnath et al., 2023). Some use only a drop-off limit (Sheridan et al., 2024; Quint et al., 2025). Aird et al. (2021) reviewed other common detection methods. Importantly, they showed that the LLJ statistics vary dramatically depending on the definition and detection method used. This makes cross-comparison between studies using different detection methods difficult. Finally, Hallgren et al. (2023) suggested that methods that require certain positive and negative shear below and above the nose are ideal for wind energy applications. They showed that these detection methods capture most cases of asymmetric drop-off above and below the nose. As these asymmetrical cases can drive high turbine loads, it is important to consider these cases in wind energy applications. Clearly, to make progress on understanding the LLJ/wind farm interactions, some effort also needs to be made to unify these disparate methods of LLJ detection and reach some level of standardization.

To achieve these goals, we start with an overview of the experimental site and the data we utilize. The Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) site and the profiling radar data collected there are discussed in Sect. 2, along with the pre-processing steps used to prepare the radar data. With radar data in hand, we can apply traditional LLJ detection methods, which use thresholds on the LLJ nose or shear above and below the nose to determine whether a vertical velocity profile indicates the occurrence of an LLJ or not. Section 3 introduces these methods along with their associated challenges, such as the sensitivity of the methods to the thresholds used. It also demonstrates the impact of wind farm construction on the occurrence rates of different velocity profiles observed at the site, along with the variation in the characteristics of LLJs, such as nose wind speed and height.

In contrast to traditional LLJ detection methods, in Sect. 4 we develop a data-driven LLJ detection method. In particular, we use agglomerative clustering to define characteristic wind speed profile clusters. We can then select clusters that are representative of LLJ occurrence without needing to specify thresholds on wind speed or wind shear. By comparing the frequency of LLJ clusters that occur before and after the construction of the wind farms, we can assess the impact of the farms on LLJ occurrence rates.

In Sect. 5, we compare the results from both the traditional LLJ detection methods and the new data-driven method and show broad similarities between the two. We also note how our results might be useful to wind farm designers who want to predict the occurrence of LLJs and their characteristics after the construction of wind farms by taking into consideration pre-construction LLJ data and the specifications of the wind farms, such as layout, turbine power and thrust curves, and turbine dimensions. Finally, we offer closing remarks in Sect. 6.

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

2 Experimental site and data

In this work, we utilize radar data collected at the ARM SGP C1 site in Lamont, Oklahoma, which are publicly available online (Coulter et al., 2026a). We chose this data stream specifically because, starting from 2012 and continuing up until 2020, several large wind farms were installed near the radar location. Figure 2 shows the location of the radar (black star symbol) relative to the nearby wind farms. Table 1 reports the year that each wind farm came online, as well as the distance and heading from the ARM site to the centroid of each farm. The radar collected data from August 1998 to the present time (April 2026 at the time of this study). As a result, it provides an excellent data set to characterize both LLJ occurrence rates and characteristics before, during, and after the construction of neighboring wind farms. In addition, three intermediate ARM sites had radar installed and started operating between 2014 and 2024. These sites, sites I8, I9, and I10, are also shown in Fig. 2. Since the radars at these sites began recording data after wind farm construction, we cannot perform long-term pre-construction/post-construction analyses using the data from these sites. However, we still use these sites to check for spatial variability in LLJ detections.

The radars installed at the ARM sites are 915 MHz Vaisala profiling radars. Using consensus averaging, they can estimate the wind profile with range gates between 0.06 and 1 km. At the SGP site, two different powers are used, resulting in two

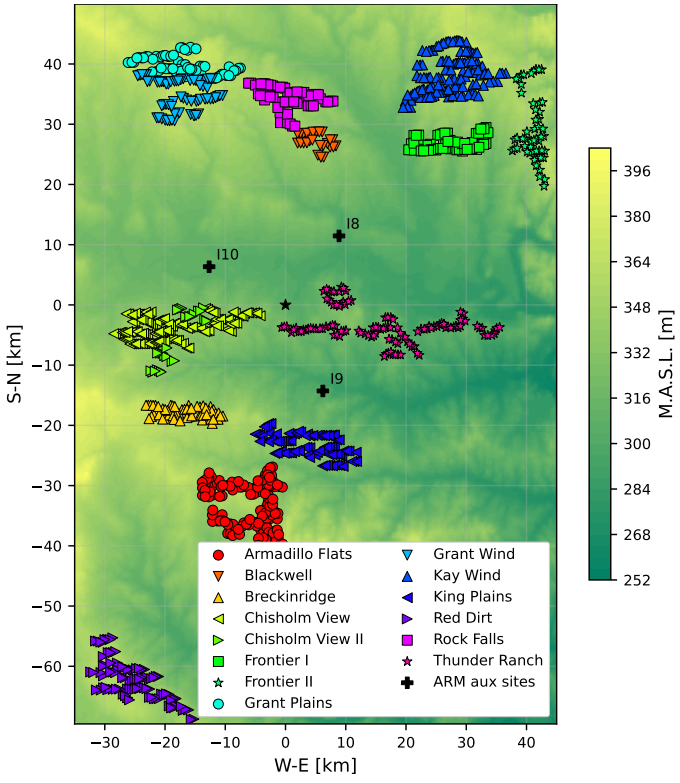

Figure 2. Map showing the location of the ARM SGP C1 radar (black star at the origin), intermediate ARM sites, and neighboring wind farms (positions relative to the radar site).

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Table 1. Year of commissioning, distance, and heading from the ARM SGP C1 site to the centroid of each surrounding wind farm. Typical rotor diameter, rated power, and total turbine count are also reported.

Wind Farm	Year Online	Distance (km)	Heading (°)	Rotor Diameter (m)	Rated Power (MW)	No. of Turbines
Blackwell	2012	27	11	108	2.30	26
Chisholm View I	2012	19	259	82.5	1.68	140
Breckinridge	2015	24	223	103	1.70	57
Kay Wind	2015	47	35	108	2.30	130
Chisholm View II	2016	19	253	107	2.40	27
Frontier I	2016	38	45	126	3.30	61
Grant Plains	2016	43	337	108	2.30	64
Grant Wind	2016	35	334	108	2.30	66
Red Dirt	2017	67	202	125	3.15	95
Rock Falls	2017	34	1	120	2.63	60
Thunder Ranch	2017	17	102	116	2.50	120
Armadillo Flats	2018	34	190	116	2.30	131
King Plains	2020	24	171	127	2.82	88
Frontier II	2021	50	55	149	4.80	53

measurement profiles: one with a coarser vertical resolution of around 200 m and reaching altitudes of 6 km, and another with finer vertical resolution of around 60 m and reaching altitudes of just over 2 km. Since we are most interested in LLJs with noses occurring in the lower 1 km of the atmosphere and need enough vertical resolution to detect the wind speed peaks and compute the wind speed shear, we only use the second power option. For this scan, the pulse repetition rate is about 43.5 kHz. A five-beam scan with four tilted beams and one vertical beam is used. The tilted beams have elevation angles of 74.5°. The range gate is approximately 60 m.

Basic data pre-processing is performed by the ARM team so that the available online data have already been temporally averaged to 1-hour profiles and have undergone basic data filtering to reject outliers. The pre-2019 record uses the `windconC1` consensus-averaged hourly product. After August 2019, this product was discontinued and replaced by the `windcnsC1` 5-minute consensus product, which we vector-average back to a 1 h cadence to match the legacy record. The two streams are spliced at 2019-08-19 with an instrument-driven gap of approximately 58 days bridged by NaN-padding before the filtering pipeline is applied.

After the ARM pre-processing, the data still contain outliers that might skew detection rates. For instance, an erroneously high wind speed could be detected as an LLJ. In addition, we want to preserve as much data as possible and interpolate to infill NaN values when possible. We apply a combination of two-dimensional interpolation in height and in time with a principal-component-analysis (PCA)-based filter to reject points with high suspected noise. For details on this procedure, the interested reader is referred to Appendix A. Here, we report salient details on the PCA. We restrict the data to heights between 125 m and

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

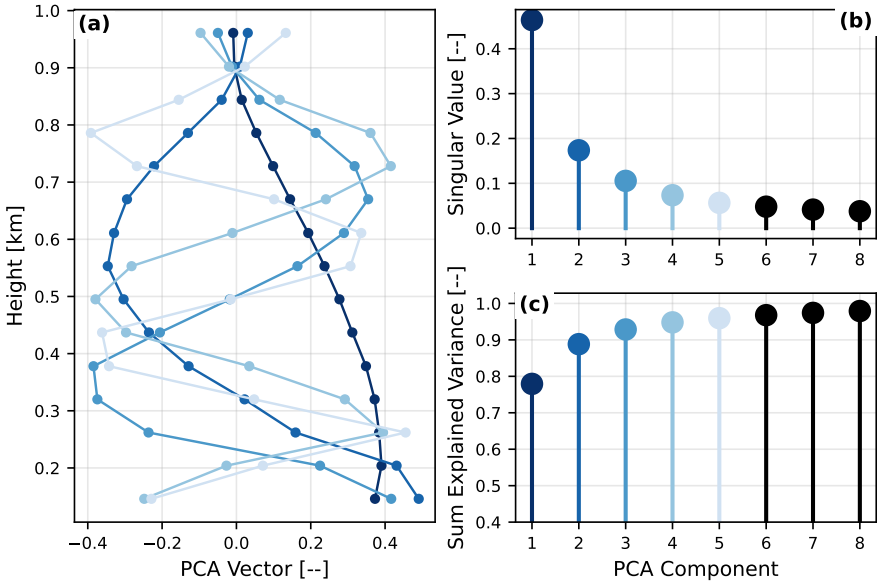

Figure 3. Singular vectors and values obtained from the PCA decomposition. The first five singular vectors are plotted in panel (a) with colors corresponding to the singular values and explained variance ratio plotted in panels (b) and (c), respectively.

1 km to avoid issues noted in near-ground measurements. We normalize data with the average velocity above 800 m, then fit the PCA. The number of principal components is chosen to reconstruct 97.5% of the variance, which results in 8 components for site C1. Figure 3 shows the singular vectors (only the first five), the singular values, and the total explained variance ratios of these 8 principal components, for the PCA fit on site C1. The neighbor sites each retain 6 principal components to reach the same 97.5% variance target (I8: 97.64%, I9: 98.04%, I10: 97.67%). We note here that the PCA already provides insight into expected characteristic wind speed profiles. The first vector (with the highest singular value) in Fig. 3(a), for instance, represents a case of negligible shear near the ground with monotonically decreasing speeds aloft. The next singular vector represents a case of maximum wind speed near upper elevations (between 700 m and 900 m), still with negative shear aloft. In other words, the two primary vectors from the PCA represent wind profiles that violate the canonical monotonically increasing wind speed with height. However, we should note that the actual measured wind velocity profiles are only obtained as a linear combination of the downselected PCA singular vectors.

The final number of accepted time stamps after the interpolating and filtering procedure is 168,857, which implies a total rejection of approximately 21% of the initial time stamps. The distribution of these good time stamps over the years is plotted in Fig. 4(a). Since farm construction began in 2012 and ended in 2020, we use those dates to split the data into pre-construction, during-construction, and post-construction periods. Since the LLJ can have strong seasonal and daily variability (Whiteman et al., 1997), we plot the distribution of data over month and hour in Figs. 4(b) and (c), respectively.

From Fig. 4(a), we see that the yearly data availability varies significantly from year to year. For this reason, we do not attempt to show annual variability in the LLJ characteristics. Instead, we will split the data into periods and consider only the

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

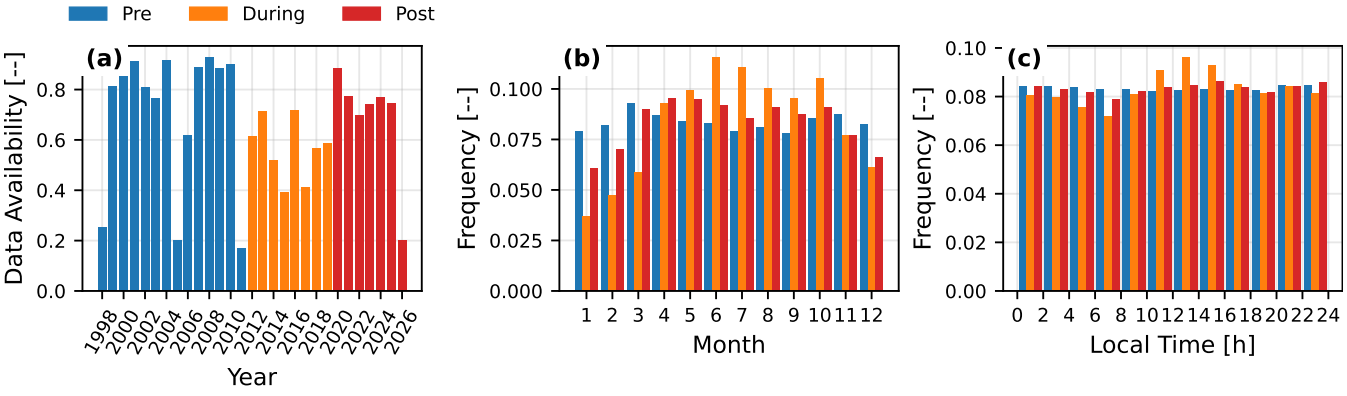

Figure 4. Data availability: across all years (a), months (b), and time of the day (in 2-hour wide bins) (c).

differences in these three data sets. We will apply this pre/during/post split in two ways. In the analyses when we are studying the impact of the farms on the site climatology as a whole, we will split the pre-, during-, and post-construction data sets using the years already mentioned. On the other hand, if we are interested in specific modulations of the LLJ morphology due to the presence of specific farms, then we will perform the pre/post split on an *ad hoc* basis, choosing the split year according to the construction dates of the relevant farms.

When performing these comparisons, we must keep in mind the seasonal and diurnal cycles in the LLJ occurrence rates, which might induce biases in observed results. When considering climatology as a whole, we balance the pre-, during-, and post-construction sets seasonally as follows. Defining four seasons following the four calendar quarters (e.g., Jan-Mar, Apr-Jun, etc.), we compute the maximum number of time stamps belonging to each season for each period. Then we randomly bootstrap additional samples from the seasons with populations less than this maximum so that all the seasons have equal populations. In this way, we balance seasonality while preserving the unique contributions of individual measurements in the highly populated seasons, as opposed to, for instance, downsampling the highly populated seasons to match the population of low seasons. Note that bootstrapping time series data that are autocorrelated can reduce the variance of those data (Künsch, 1989). To correct for this, bootstrapping is done in time blocks rather than sampling individual time stamps (Wilks, 1997). We perform a simple sensitivity analysis on wind speeds and LLJ detection rates (introduced later in Sect. 3) and find that 7-day blocks are appropriate and bootstrapped confidence intervals (CIs) plateau at this point. So, when bootstrapping to up-sample seasons here, or later in the paper as needed, we use moving block bootstrapping (MBB) with a 7-day block.

We report briefly here the results of up-sampling the pre-/during-/post-construction data sets. Pre-construction data start with 87,103 unique source points and are mostly uniformly distributed in season, needing minimal up-sampling. The final, up-sampled pre-construction data set has 89,685 time stamps, 3% more than the initial unique values and representing around 10.2 years of data. During-construction data are the most imbalanced and have been up-sampled by 23% from 39,620 unique values to 48,727 final values, representing approximately 5.6 years of data. Finally, post-construction data are up-sampled by 12% from 42,135 unique initial samples to 47,384 final samples, representing around 5.4 years of data.

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Finally, we consider here whether there is significant climatological variability that might obscure impacts on LLJ occurrence due to the construction of the farms. To do so, we introduce the concept of background conditions, which we will use later in several of our analyses. We assume that wind speeds and directions sufficiently far from the farms will not be impacted by the presence of the farms and therefore give a consistent estimate of the background atmospheric conditions pre-, during-, and post-construction. From the filtered radar data, we take the average of wind speed and direction measurements with elevation $\geq$800 m to be the background wind speed and direction. We apply the two one-sided tests (TOST) procedure to show equivalence in mean background wind speeds between the three periods within 0.5 m s^{-1} (Schuirmann, 1987; Lakens, 2017). TOST uses the two tests to define CIs on the mean wind differences. Provided both fall within $\pm$0.5 m s^{-1} of zero, we assume the two mean wind speeds to be identical. The mean differences between these periods are all below 0.5 m s^{-1}. However, 90% bootstrapped CIs for the comparison of the pre-construction and during-construction periods exceed the 0.5 m s^{-1} limit by about 0.09 m s^{-1}. The during-/post-construction comparison and pre-/post-construction comparisons both pass the TOST. As a result, we conclude that the three periods are effectively equivalent in mean background wind speed. Although there are other factors to consider, such as frequencies of stability regimes, drought/high moisture periods, rare and strong storm/frontal passages, and so on, we suggest that the averaging period is long enough to reduce the impact of these features. Our concern here is to show that there is no long-term climatological drift in the upper-elevation wind speeds that might influence later results.

To check for further potential background climatological variability, we compute the yearly, monthly, and hourly mean wind speeds. The yearly data are averaged over each calendar year, whereas the monthly and hourly data are grouped by period and then averaged across all years in the period. Figure 5 plots these results. The figure also reports the 90% MBB CIs computed from the bootstrapping used to balance the seasons. In Fig. 5, we see evidence of random climatological variability but no clear evidence for longer-term climatological trends. However, although the mean wind speeds have converged, the periods may not be of sufficient length for the monthly or hourly wind speeds to also converge. Although this does not prevent future analyses entirely, it is worth bearing in mind. For instance, we will not perform any analyses that split the data by season or month, so the lack of monthly convergence is not critical, provided that the mean wind speeds converge, as has already been shown. On the other hand, we will use bins in the time of day to study nocturnal LLJs. Thus, later analyses will need to account for potential climatological variability and ensure that wind farm effects are of sufficient magnitude to rule out random climatological fluctuations.

3 Traditional LLJ detection methods and derived results

We now compare the characteristics of the LLJ before, during, and after wind farm construction using traditional LLJ detection and characterization methods. We start with a gradient-based detection method, which has recently been shown to be well suited for wind energy applications (Hallgren et al., 2023). For this method, referred to as the shear method, the velocity profile must have a maximum wind speed at a height lower than 750 m, as we are interested in LLJs relevant for wind energy applications, thus not too far from the turbine rotor heights (which, for this site, tend to be between 30 m and 150 m). Furthermore, the

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

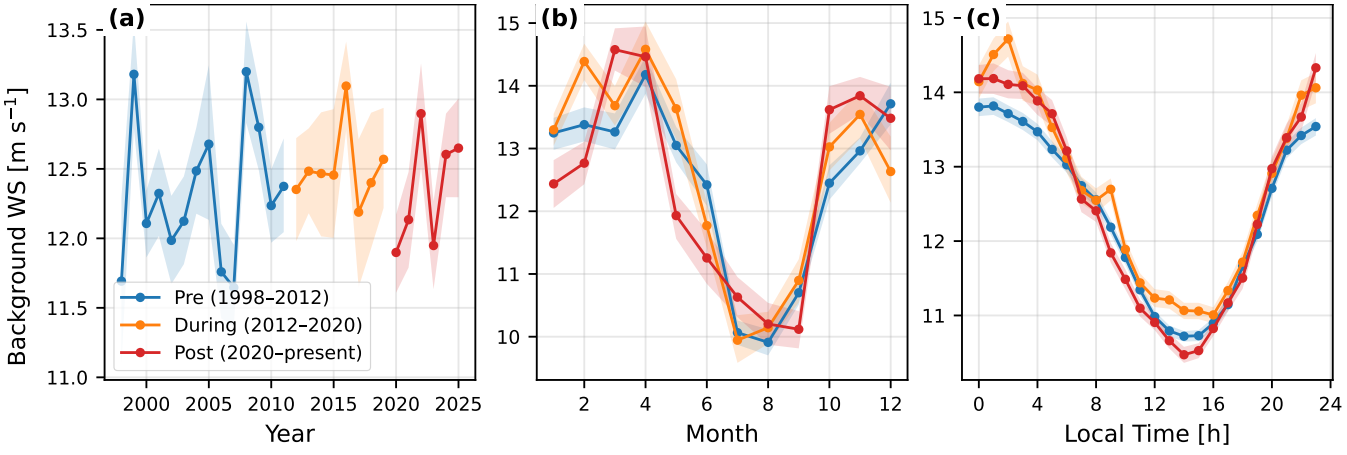

Figure 5. Mean yearly, monthly, and hourly wind speeds for the construction periods with 90% MBB CIs.

availability of radar data up to 1 km allows us to be confident of a velocity decrease above the maximum wind speed. For the shear method, above and below the LLJ nose, the absolute value of vertical wind shear should be greater than 0.01 s^{-1} (Hallgren et al., 2023). However, because we clip the lower-elevation data in the pre-processing step, we may have cases of high wind speed near the lower limit of observed data that would be rejected by this method, as we would not have sufficient data to observe the requisite speed shear. To account for this, we also compute the average shear between the lowest observed wind speed and the ground, assuming the speed to be zero at the ground. If this shear is greater than the maximum positive shear observed below the nose, we take it to be the below-nose shear.

As an alternative to the shear method, especially for profiles for which it can be difficult to compute the vertical velocity gradient due to missing data, we also consider a more typical drop-off-based method (Debnath et al., 2023). For this method, we require the LLJ nose to occur below 750 m and the difference between the wind speed at the LLJ nose and the local minimum of the wind speed above the nose, which is termed drop-off, to be at least 2 m s^{-1}. Furthermore, the drop-off normalized by the nose wind speed should be at least 0.1. While this method usually requires the wind speed maximum to occur with at least one measurement below it to validate the existence of a true nose, as mentioned above, we limit our lower observations, leading to several cases where the maximum is detected at the lowest height. Again, assuming the speed to go to zero at the ground, we can still assume these cases to be LLJs with nose speeds equal to or greater than the observed maximum wind speed. For this reason, we do not require that the maximum wind speeds be above the lowest observation height. In practice, the shear test is applied to wind-speed profiles whose missing heights have been linearly filled in, whereas the drop-off test is applied to the raw profiles (operating on observed heights via NaN-safe maximum and minimum operations), reflecting the respective sensitivities of each method to data gaps.

As an initial comparison, we compute the total fraction of samples that are considered to be LLJ profiles by the two methods described above for the pre-, during-, and post-construction data sets. Figure 6 illustrates the detection rates of the two methods for these different periods. Recall that MBB is used to generate bootstrapped samples that are uniform in season, so the seasonal

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

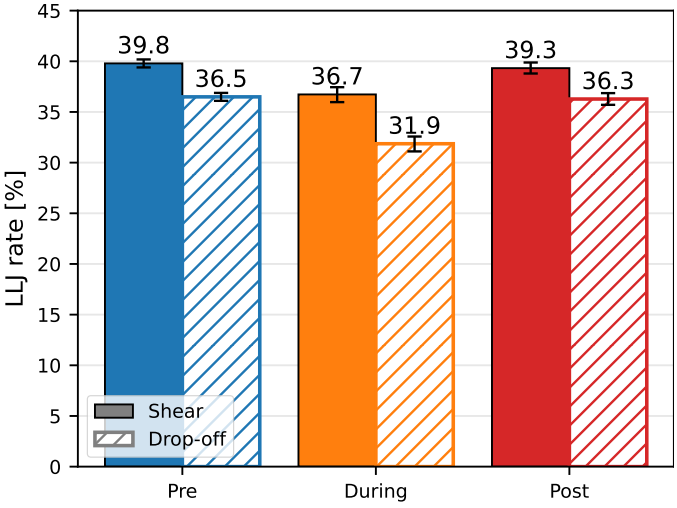

Figure 6. LLJ detection rates using the shear and drop-off detection methods during the different time periods with MBB CIs showing 90% range.

effects in LLJ occurrence are mitigated (Whiteman et al., 1997). The MBB also allows one to compute 90% CIs, which are plotted in the figure. We observe that rates have converged well, as indicated by the narrow CIs, and are roughly stable over the three periods. Pre- and post-construction detection rates are nearly identical and, in both cases, are approximately 3% lower for the drop-off method than for the shear method. The rates drop approximately 3% for the shear method and about 4.5% for the drop-off method during the construction period.

To consider in greater detail the differences in LLJ morphology between shear-detected and drop-off-detected LLJs, we apply a quadrant analysis as follows. Given two binary metrics, i.e., the drop-off and the shear detection methods, we can generate four categories: both methods detect an LLJ (case A), only the shear method detects an LLJ (case B), only the drop-off method detects an LLJ (case C), and neither method detects an LLJ (case D). The number of measurement periods ascribed to each case is denoted by a, b, c, and d, respectively. The total number of measurement periods is given as $n = a + b + c + d$. Using these values, we can employ several metrics to assess the agreement of the drop-off and shear methods before, during, and after the construction of the farms.

The first metric that we use is the overall agreement, given simply as $(a+d)/n$. If the two methods never agree, the value is 0, while it is 1 if they always agree. We can likewise define positive and negative agreement as $2a/(2a+b+c)$ and $2d/(2d+b+c)$, respectively, which give the fraction of agreement in the two methods considering only cases where an LLJ is detected (positive agreement) or not detected (negative agreement). Like overall agreement, these metrics both vary from 0 to 1. Finally, we also consider the Matthews correlation coefficient (MCC):

$$MCC = \frac{ad - bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}, \tag{1}$$

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

which varies from −1 to 1, indicating anti-correlation and correlation, with 0 indicating no correlation between the two methods. In Fig. 7 we report these agreement metrics for the pre-, during-, and post-construction periods. We also report the population of each individual quadrant.

Comparing the pre-, during-, and post-construction percentages reported in Fig. 7, we note that there is general agreement between the two methods under all construction conditions. MBB CIs reveal convergence of the metrics. Quadrant populations are nearly identical between pre- and post-construction periods. During-construction conditions have slightly fewer A and C occurrences and slightly more B and D occurrences, matching the results from Fig. 6, indicating a decrease in LLJs overall and a decrease in drop-off-detected LLJs. All agreement metrics are around 90%, except for the MCC, which sits above 80%. Negative agreement is marginally higher than positive agreement, revealing that the methods more often agree that there is no LLJ than that there is an LLJ. However, this difference is small. At least for this site, then, the shear and drop-off methods are very similar.

Although the broad characteristics of drop-off and shear might be little impacted by the presence of the farms and generally agree on the presence of LLJs, the average profiles may nevertheless show differences depending on the detection method. We next plot the average normalized wind speed profiles of each quadrant before, during, and after farm construction in Fig. 8 to check if the profile shapes change depending on the state of farm construction or the detection method used. From the figure, we note that the mean profiles for all cases are relatively similar. In every case, though, the post-construction profile generally has a larger wind speed relative to the upper-elevation wind speeds than the pre-construction profile. In cases B and D the during- and post-construction profiles are nearly identical. However, for cases A and C the profiles diverge. In case A, the during-construction profile lies between the pre- and post-construction profiles. Yet for case C it is generally slower than both profiles, except at the lowest elevations. MBB CIs are plotted in Fig. 8 but are barely visible, indicating that the mean profiles are well

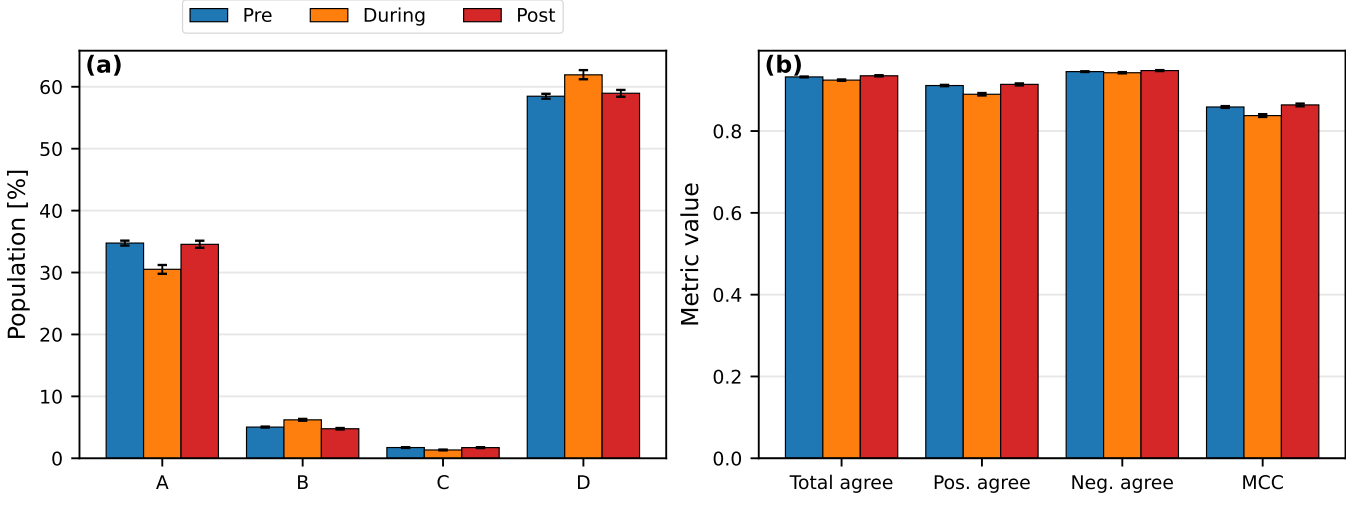

Figure 7. Populations of the four quadrants during the three construction periods (a) and the agreement metrics during the same periods (b). MBB CIs report 90% ranges.

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

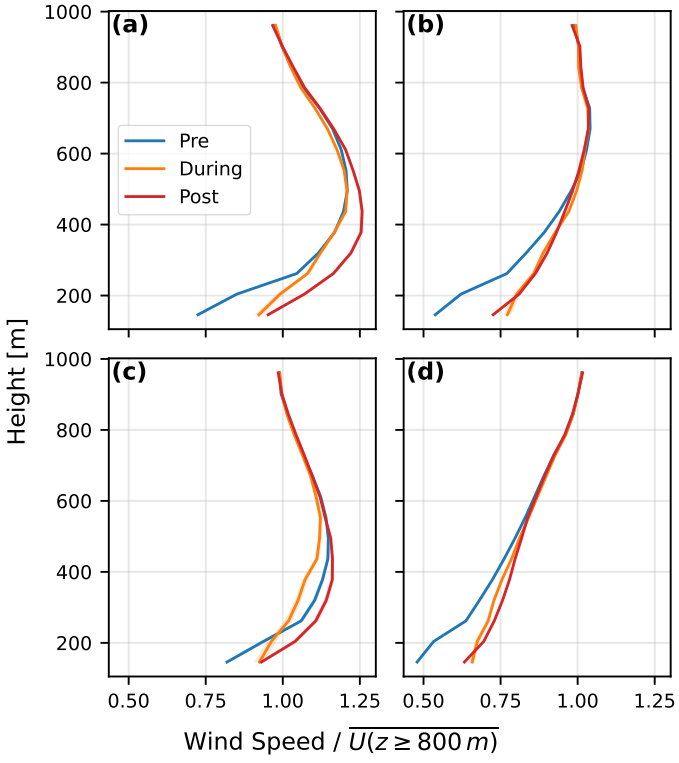

Figure 8. Mean wind speed profiles of all measurements belonging to each case for the quadrant analysis, computed before, during, and after the construction of the wind farms. Each profile is normalized by its own average wind speed at $z \geq 800$ m before the case-wise average is computed.

converged. We emphasize that each profile is normalized by its own upper-elevation ($z \geq 800\,\mathrm{m}$) average wind speed before averaging within a case, so apparent post-construction speed-ups at lower elevations are partially a relative-to-upper-winds effect. Potential variations in the upper-elevation winds will be analyzed in the following.

Considering the different cases further, we note that case A has the most clearly defined LLJs. Case B has a higher nose and, as a result, less of a drop-off above the nose. This explains why case B profiles are not labeled LLJs by the drop-off method: the nose is too high and therefore the flow does not have time to sufficiently decrease in speed above the nose to satisfy the drop-off criteria. On the other hand, case C profiles exhibit a clear LLJ nose but with less speed decrease below the nose. Finally, case D contains monotonically increasing boundary layer profiles that clearly are not LLJs by either criterion.

To further investigate the cases of the quadrant analysis, we also plot histograms of the relevant characteristics of LLJs (i.e., nose height and nose wind speed) for the cases where an LLJ is detected (i.e., cases A–C). We also consider the times of day at which the LLJs occur. We have no access to coincident surface flux measurements or other ways to quantify atmospheric stability, so time of day is our best proxy.

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

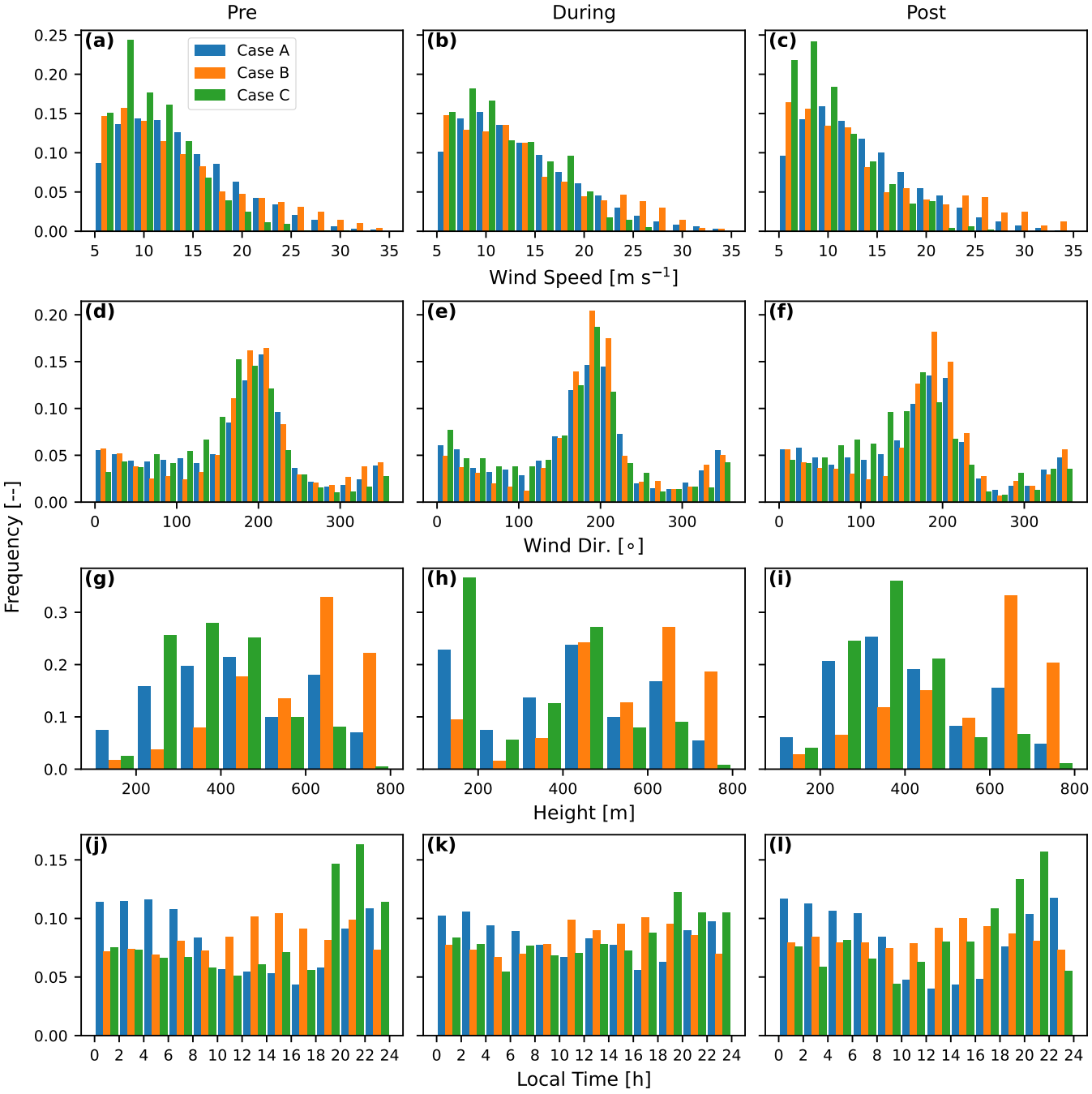

Figure 9. Histograms of LLJ nose speed (a–c), LLJ nose wind direction (d–f), LLJ nose height (g–i), and the time of day of the LLJ (j–l) for the different LLJ cases, plotted before and after the construction of the farms (a, d, g, j; b, e, h, k; and c, f, i, l representing pre-, during-, and post-construction, respectively).

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

From Fig. 9, we can draw several interesting insights. Starting with the wind speed results, we see that case C LLJs are markedly slower than case A or B LLJs during pre- and post-construction periods but not during-construction. Between cases A and B, there is no significant difference in the LLJ nose wind speeds. This suggests that the LLJs captured by case C of the quadrant analysis generally occur for relatively lower wind speeds ($\lesssim 8\,\mathrm{m\,s^{-1}}$). For those wind conditions, the presence of farms can induce enhanced turbulent mixing due to the presence of wind turbine wakes, bringing air parcels with higher momentum to lower heights and thereby slightly increasing wind speed and reducing wind shear, as shown in Fig. 8.

Next, considering the height of the LLJs, we see that cases B and C are opposite, with case B occurring more often at the upper heights and case C at the lower or middle heights. This makes sense given the mean profiles in Fig. 8. Case A exists somewhat in the middle, with a preference for mid-height LLJs and not extreme heights. There does not seem to be a large impact on the typical nose heights due to the presence of the farms in this analysis, although there is a marked increase in the low noses during construction.

Finally, considering the time of day, case A follows the expected daily cycle, with an increase starting around sunset and peak occurrence rates just after midnight (Whiteman et al., 1997). Case B, by contrast, occurs most often in the afternoon, with the most pronounced afternoon peak coming in the pre-construction period but persisting weakly in the during- and post-construction periods. Case C has roughly evenly distributed occurrence, with the exception of the hours just before midnight, when it peaks. We need to be careful in placing too much weight on these non-canonical occurrence patterns, since cases B and C cumulatively account for less than 10% of occurrence during all periods. Nonetheless, we speculate that this variability might give some insight into the generation of LLJs. To check this, we look through each period for the occurrence of case B and C time stamps. We then check how often each case is followed by a case A time stamp in the next 12 hours or, alternatively, a case B or case C time stamp, whichever is not currently being considered. That way we can get a sense of how these profiles evolve in time. Table 2 reports these results. During all periods, cases B and C are more likely to transition to the agreement case A than the inverted disagreement case. When the inversion occurs, it is more likely that case C produces case B than case B produces case C. In other words, both LLJs with lower noses and less below-nose shear and LLJs with high noses and less above-nose drop-offs are likely to transition to well-defined LLJs (case A). On the other hand, it is more likely for a low nose LLJ to move upward (case C to B) than for the opposite to happen (case B to C). The pre-midnight spike of case C LLJs likely transitions to the post-midnight maximum of case A LLJs. The afternoon occurrence of case B might help start the increase of case A LLJs during the evening and night, but probably does not contribute to the spike in evening case C LLJs.

Table 2. Probability that a case-X profile is followed by at least one case-Y profile within the next 12 h, with 95 % CI from MBB.

Transition	Pre	During	Post
$B \to A$	89.08 % [87.81, 90.38]	86.63 % [84.52, 88.53]	85.14 % [83.11, 87.08]
$B \to C$	17.18 % [15.57, 18.79]	18.22 % [15.94, 20.46]	16.72 % [14.31, 19.17]
$C \to A$	95.34 % [94.17, 96.40]	89.08 % [86.53, 91.61]	93.40 % [91.29, 95.43]
$C \to B$	39.91 % [37.10, 42.78]	40.94 % [36.92, 44.98]	35.30 % [31.59, 39.12]

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

In closing, we comment briefly on which detection method might be preferable for wind energy considerations. Clearly, most of the LLJ profiles will be identified by either method, which will agree with each other most of the time (more than 90% of the time during all periods). However, in the cases where the methods disagree, the shear method detects more LLJs. However, these shear-only LLJs appear to have noses at higher elevations far from rotor heights. On the other hand, drop-off-only LLJs seem to have noses very close to the ground. LLJs with lower nose heights would be especially important for wind energy applications, as they could produce the highest wind speeds over the rotor area of any LLJ condition. Thus, in contrast to Hallgren et al. (2023), the drop-off method might be better for wind energy applications at sites where measurements below the turbine top-tip are limited. Nonetheless, in contrast to other studies (e.g., Aird et al., 2021), we find little difference in the application of the two methods at this specific site and with this specific data set.

4 Data-driven LLJ detection

In this section, we consider data-driven alternatives to traditional LLJ detection methods discussed in the previous Sect. 3. We argue that, when LLJs occur with enough regularity and the associated wind-speed profiles are distinct enough from monotonically increasing or other classical profiles, unsupervised data-driven methods should be able to surface these profiles automatically, without the need to precisely define what constitutes an LLJ wind-speed profile. If this is the case, then we can avoid the issues that arise from using metric-based detection methods, such as the sensitivity to the type or value of the metrics used.

For such a task, unsupervised clustering algorithms are the natural choice of data-driven algorithms. These algorithms group similar observations into clusters. Most unsupervised clustering algorithms start with a user-defined number of clusters. The algorithm then tries to fit the data optimally into that number of clusters. A common example of this type of clustering is the k-means algorithm (Lloyd, 1982; McQueen, 1967). However, it may not initially be clear how many clusters are contained in the data set. Especially for high-dimensional data, such as wind speed profiles, when visualization is difficult, there is no easy way to determine the appropriate number of clusters needed. Furthermore, if the clustering is done all at once, in a single step, highly populated areas may skew clusters in non-physical ways. An alternative to address both of these problems is hierarchical clustering (Forina et al., 2002; Iungo et al., 2022). With this approach, clustering is done iteratively. The first step may be to cluster the data set into a few clusters, then take one of those clusters, and further subdivide into new clusters. It is not as critical to get exactly the right number of clusters at each step, since the clusters can always be refined by a further subdivision. Furthermore, iterative subdivision also avoids bias towards highly populated regions. Previous works have used k-means clustering in a hierarchical approach to successfully classify wind speed profiles (Moss et al., 2024, 2025).

Instead of using hierarchical k-means clustering, in this work, we utilize agglomerative clustering, which is another hierarchical unsupervised clustering algorithm (Ward, 1963; Lance and Williams, 1967). Although the results are similar to hierarchical k-means clustering, we choose agglomerative clustering for a couple of reasons. First, with k-means clustering, the user still needs to specify the number of clusters at each step. This means that different users might generate different clusters. Agglomerative clustering does not require a user-specified number of clusters. Second, the k-means algorithm is sensitive

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

to the starting point of the initial centroids. This sensitivity is typically mitigated by running the clustering several times in parallel using different randomized starting centroids. Nevertheless, k-means clustering is non-deterministic. Agglomerative clustering, on the other hand, always produces the same clusters. For both of these reasons, we choose to use agglomerative clustering to ensure that the clustering process is automatic, deterministic, and does not require user input.

For full details on the clustering method we apply, the interested reader is directed to Appendix B. Here, we note that, despite the dendrogram produced by the agglomerative clustering approach being deterministic, user choice is still required to determine whether a given cluster should be resolved into finer clusters. We split into sub-clusters when the mean normalized velocity profiles of the sub-clusters are sufficiently different with respect to the cluster variance. This results in 9 clusters. After constructing clusters with the agglomerative approach, some time stamps may be closer to the centroids of clusters other than the one to which they have been assigned. To ameliorate this mislabeling, we perform a single k-means relabeling pass. Each time stamp is assigned to the cluster with the nearest centroid where the centroids are the mean normalized velocity profiles of each cluster. Finally, it is important to note that we cluster all the data simultaneously without splitting into pre/during/post-construction periods or re-balancing in time. There is no need to add the re-balancing correction, as the agglomerative approach is designed to be insensitive to clusters of imbalanced sizes. Furthermore, if we were to cluster on time periods, there would be no easy way to compare cluster membership across time periods, as different periods would generate different clusters and, possibly, different numbers of clusters.

To illustrate the clustering procedure, we plot the dendrogram of sub-clusters determined by this approach. Figure 10 shows how each cluster is divided, when appropriate, into sub-clusters. The initial two clusters are the first two clusters generated by the agglomerative clustering and split the entire data set into two initial clusters. From there, the clustering continues iteratively, resulting in 9 final clusters. The clusters shown in Fig. 10 are from the agglomerative results after the k-means adjustment. The relabeling is illustrated in the confusion matrix plot in Fig. 11. As can be seen, relabeling occurs for many clusters, but is minor.

Having defined the clusters, we sort the clusters according to the following convention. We take the mean profile of all the normalized wind speed profiles belonging to each cluster and compute the differences in wind speeds between adjacent heights, approximating the shears of the profile. For each cluster, we record the maximum positive value and, if there are negative values, we record the minimum negative value. For all clusters with no negative value or a negative value greater than -0.02, we sort the clusters by decreasing positive value. For clusters with a negative value less than -0.02, we sort in decreasing negative value. Cluster labels are then assigned in increasing order, starting with the clusters in decreasing positive-value order and continuing into the clusters in decreasing negative-value order. This results in a smooth transition from strong shear profiles to relatively flat profiles to LLJ profiles as the cluster label increases. Figure 12 shows these results.

We can begin to investigate the impact of the farms on the clusters by splitting into pre-, during-, and post-construction periods. Figure 12 also shows the mean profiles split into these periods. In general, the mean profiles are very similar. As a result, we can primarily check the increase or decrease in occurrence rates to learn about the impact of wind farms on LLJs. Figure 12 therefore also reports the fraction of occurrence of each cluster in the period data sets along with the 90% MBB CIs.

From inspection, we can begin to understand the impact of the farms on LLJs. Most clusters that decrease in population post-construction versus pre-construction, for instance, belong to labels 1, 2, 7, and 9. The clusters that increase in population

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

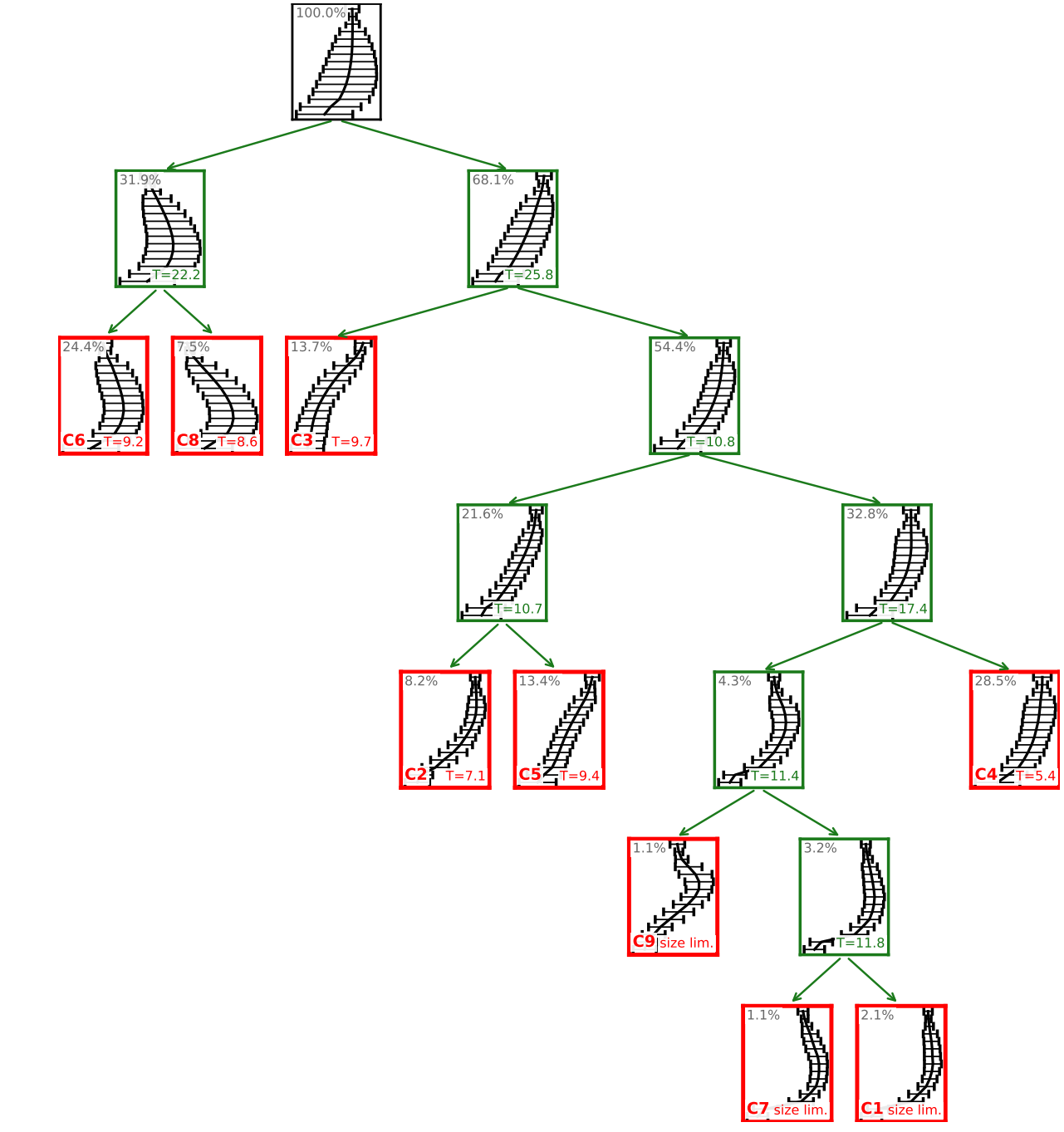

Figure 10. Dendrogram of sub-clustered results. Labels are from the agglomerative clustering pass. The percentages indicate the cluster population as a percent of the total. The value T in each cluster is the splitting criterion given in Eq. (B1) and Eq. (B2) and indicates no further sub-clustering is needed when less than 10.

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

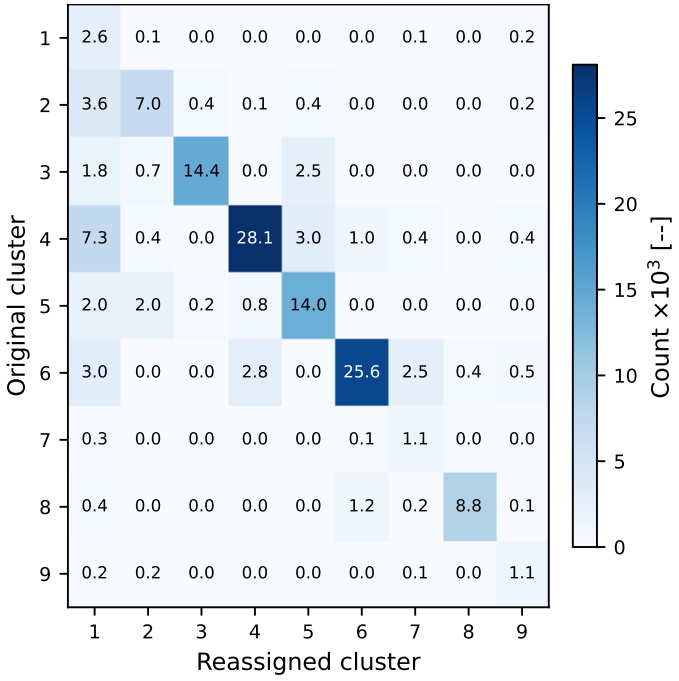

Figure 11. Confusion matrix showing initial agglomerative labels and relabeled results from the final k-means pass.

are 4, 6, and 8. The remaining clusters remain roughly consistent. The main trend, therefore, is a decrease in high-shear and LLJ profiles, and an increase in flatter, low-shear profiles. The only exception here is cluster 8. However, between clusters 7–9, this cluster has lower shear below the nose relative to the ground. Thus, if the general trend is towards reduced near-ground shear, an increase in this cluster is consistent. We note that there is some variability within this trend. For instance, cluster 1 increases by 6% from 17% to 23%, but then drops by 18% down to 5%. It is difficult to say for certain why the rate increases in the during-construction period, but farm effects in this period are nonetheless transient. It is reasonable, therefore, to accept the strong drop as due to the farm construction while the during-construction results could indicate transient effects or random variability.

To get a better sense of what is happening, we repeat the quadrant analysis from Sect. 3, but now on a cluster-by-cluster basis. Table 3 reports the populations of each of the four quadrants for the pre-, during-, and post-construction data sets.

Table 3. Per-cluster profile counts by quadrant case in the global tri-period balanced data sets. Each cell is pre / during / post-construction.

Case	C1	C2	C3	C4	C5	C6	C7	C8	C9
A	32/22/30	18/22/23	2/2/1	20/20/18	4/4/3	67/64/59	90/95/88	98/99/99	88/84/88
B	9/9/10	7/8/9	1/1/0	8/8/7	1/2/2	5/8/6	1/1/1	0/0/0	2/1/1
C	1/1/1	0/0/1	0/0/0	1/0/0	0/0/0	8/4/5	3/1/4	1/1/1	2/4/4
D	58/69/59	74/69/67	98/97/98	72/72/75	94/94/95	21/24/29	6/3/7	1/0/0	9/10/7

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

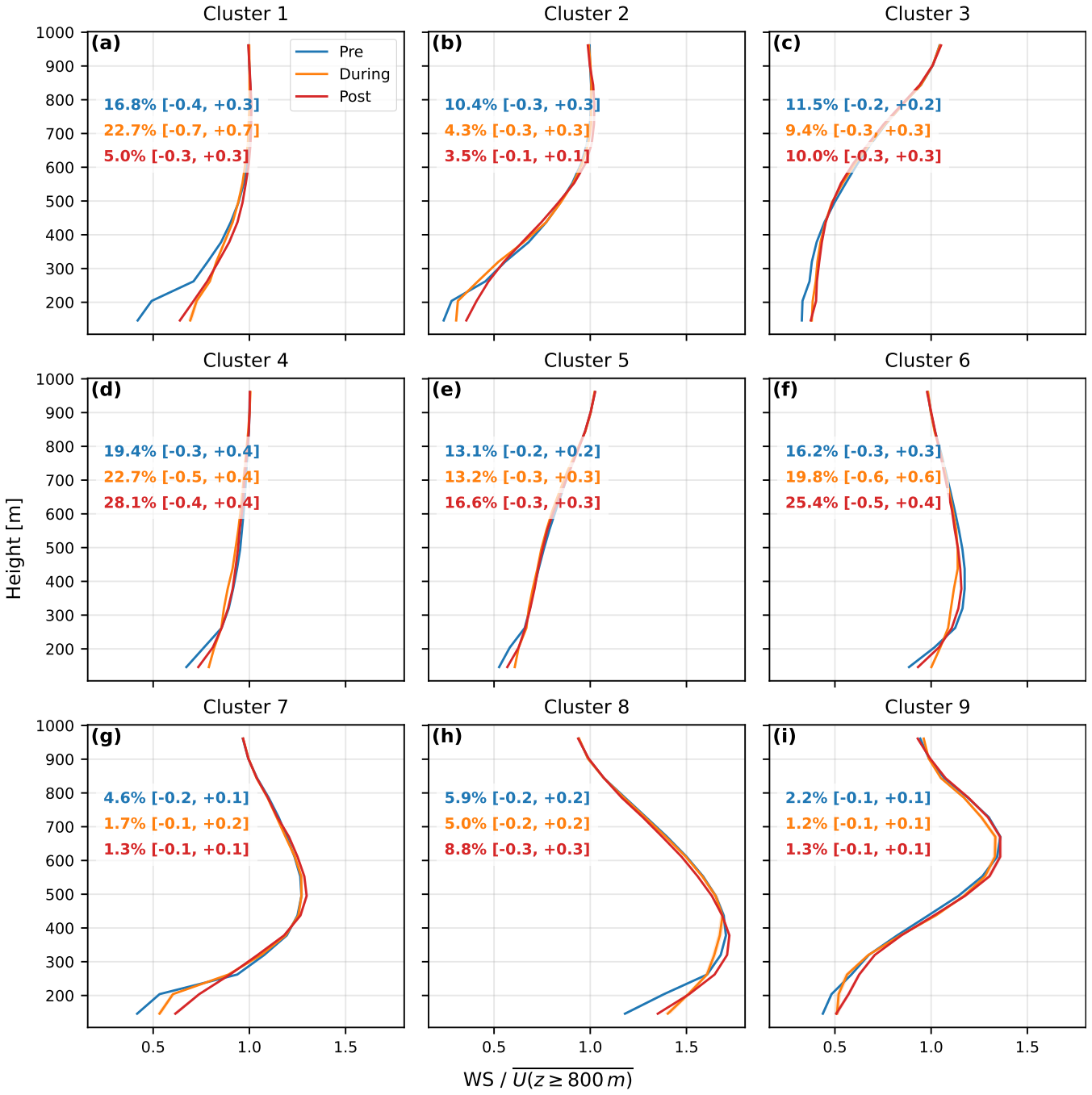

Figure 12. Mean wind speed profile for each cluster. Color indicates the period. Period populations are reported as percentages and the bracketed values indicate the upper and lower bounds of the 90% MBB CI.

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

If we consider the case D population to be the main indicator of whether a cluster can be considered an LLJ cluster or not, then we can see that the sorting method has done a good job of separating LLJ clusters from non-LLJ clusters. Clusters 1–5 are all non-LLJ, and clusters 6–9 are LLJ clusters. In particular, clusters 7–9 are very consistent LLJ clusters (more than 80%). In general, the difference in period cluster rates is small. Similarly, the disagreement rates are low. This confirms that our clustering procedure has produced clusters that are internally consistent, that is, consistently LLJ or non-LLJ, without specifying how the clusters are to be constructed. These clusters are therefore good candidates to study the impact of the farms on occurrence rates, since intra-cluster modifications of LLJ rates are small. However, note that clusters 1 and 2 are more marginally non-LLJ than clusters 3–5. Thus, increasing or decreasing these clusters is likely to have an impact on overall LLJ detection rates.

Next, to study the impact of the wind farms on the morphology of wind speed profiles, we continue to utilize these clusters, but now investigating pre- and post-construction cluster variability for specific conditions. In particular, we try to identify relevant background conditions to normalize the large-scale flow conditions between pre- and post-construction. In this way, we can isolate the impact of farms on the lower atmosphere by comparing changes in cluster occurrence rates for down-selected conditions. Note again that we no longer have a during-construction period, as we are focused on individual farms' constructions, and therefore we have a single construction date defining pre- and post-construction periods.

Having access to only the radar data, we can define background conditions only in terms of wind speed and direction, as well as the time of day, which we take to be a proxy for atmospheric stability. Background wind conditions are defined as the average values of speed and direction above 800 m, keeping the same procedure that we used to define the normalization wind speed for clustering. We define bins in the background wind speed, direction, and time of day with special focus on down-selecting wind sectors that are either directly upstream, downstream, or unimpacted by neighboring farms. In each bin, we compare the PDFs of the cluster occurrences before and after the construction of the wind farm. By comparing individual profiles that increase or decrease in likelihood of occurrence, as well as by computing the total difference in the two PDFs, we can quantify the total strength of the farm impact, as well as the typical modifications it makes on profile morphology.

To facilitate these comparisons, we first introduce the following concepts. Since the individual cluster occurrence rates all sum to 1, both pre- and post-construction, the difference in the pre- and post-construction PDFs sums to 0. However, the sum of the absolute differences will always be positive and will be greater as the difference in the two PDFs increases, to a maximum of 2. Thus, we define the strength of the difference between the PDFs as the sum of the absolute pre/post occurrence-rate differences, divided by 2 (equivalently, the total variation distance between the two PDFs, which ranges from 0 for identical distributions to 1 for fully disjoint distributions). We report this strength of difference as Δ. Further, we note that we can group all the clusters for a given bin into two groups: one group of clusters that increased in occurrence rate post-construction and one group that decreased in occurrence rate. The total amount that the increasing group increased is exactly equal to the amount that the decreasing group decreased, and it is also equal to the strength of difference between the PDFs. Therefore, we can express the modification of the PDFs due to the farm as the difference in the post- and pre-construction occurrence rates normalized by the strength of difference. The sum of all normalized differences that are positive is 1 and the sum of all normalized differences

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

that are negative is -1. This gives a convenient way to observe which clusters contribute most to the total difference in the pre- and post-construction PDFs.

In addition, in the following background-condition bin analyses, it is important to demonstrate that the bin populations are sufficiently large to make the differences in cluster populations statistically significant. Here, we continue to use MBB to estimate CIs on the reported quantities, retaining the 7-day block length used elsewhere in the paper for consistency. The cluster-significance permutation test (described below) is repeated at both 24-hour (canonical day-block) and 168-hour (week-block) block sizes; results are robust to this choice. The reported values come from the week-block analysis.

Finally, we want to ensure that the impacts of the farms are strong relative to random climatological variability and potential temporal drifts in cluster occurrence rates. We perform two checks to ensure that this is the case. For the first case, taking only pre- and post-construction conditions, we perform a permutation test with an α value of 0.05 to be 95% certain that the observed pre/post difference in cluster occurrence distributions (bootstrapped via MBB) is not random. We should reject comparisons where we cannot establish that the pre/post difference is significant compared to random fluctuations. However, just the permutation test is not sufficient. Suppose there is a long-term linear variation in a cluster occurrence rate. In this case, the permutation test would mark the pre/post difference as significant but, nonetheless, the difference would not be attributable to the farm construction and would instead be due to long-term background variability. To check for this type of drift, we split each period (pre/post) into early and late sub-periods, where all time stamps in a given sub-period are contiguous in time and the sub-periods are equal in population (that is, the split is performed at the median time stamp of the pre/post period, not the midpoint time stamp). We compute three PDF difference results: late pre-construction over early pre-construction, early post-construction over late pre-construction, and late post-construction over early post-construction. If there is a farm impact that is significant over and above any existing long-term climatological drifts, we would expect to see small intra-period drifts (intra pre/post periods) but a large inter-period drift (early post-construction over late pre-construction). This would confirm a unique farm impact.

To apply this technique, we start by considering the impact of a neighboring farm. We select wind directions between 240° and 270° to isolate the impact of the Chisholm View farm, which was operational for the entire period after 2012. To see the impact of the farm, we restrict wind speed to between 5 and 20 $\mathrm{m\,s^{-1}}$, which corresponds roughly to regions II and III of the turbine power curves at the site. Finally, we consider two time periods, one between 18:00 and 24:00 local time, representing the evening period during which inertial oscillation LLJs might form, and the other between 00:00 and 06:00 local time, representing the nighttime condition during which we expect to see a stable nocturnal LLJ (Whiteman et al., 1997).

In Fig. 13, we show the PDFs of occurrence rates before and after construction of the farm as well as the normalized difference between the two PDFs. We show these results for both the evening and night cases considered. We can draw the following conclusions. First, the impact of the farm is roughly the same in strength during the evening and night ($\Delta = 14$–$15\,\%$). However, detecting significant impacts on clusters is extremely difficult. Using just the permutation test, we see good evidence that clusters 2 and 9 decrease in occurrence during the evening, while cluster 5 increases. During the night, the permutation test indicates that clusters 2 and 7 decrease, while clusters 4 and 5 increase. Looking at the sub-period drifts for these clusters, though, none of these periods show shifts that are convincingly due to farm impacts rather than background

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

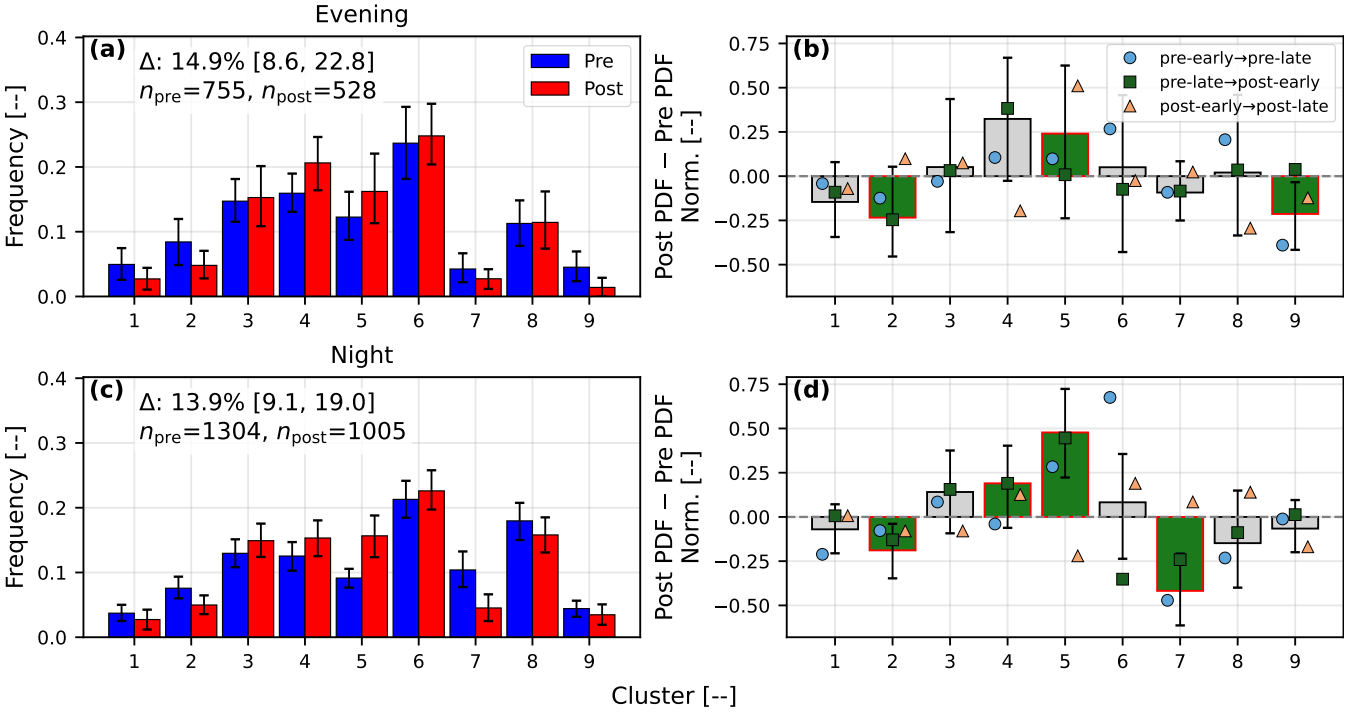

Figure 13. PDFs and difference in PDFs between pre- and post-construction data downstream of the Chisholm View wind farm. Wind speed and wind direction bins are kept constant ([5, 20] m s^{-1} and [240°, 270°]) and the time is varied to give an evening case (a, b) and night case (c, d). The error bars indicate standard deviation of the bootstrapped quantities. Highlighted bars in (b, d) indicate significant pre/post difference by the permutation test. Markers indicate the intra- and cross-period drifts.

drifts. The closest cluster is the impact on cluster 2 in the evening, where the intra-pre- and intra-post-period drifts are both less negative than the overall and inter-period drifts. However, the intra-pre-period drift is still close to the overall drift. Ignoring permutation and intra-period drift flags, though, we seem to see an increase in clusters 3–6 at the expense of clusters 1, 2, and 7–9. In particular, clusters 7–9 decrease dramatically at night. This can be explained as a decrease in near-ground shear profiles and LLJ profiles in favor of well-mixed/lower-shear profiles. Nonetheless, this explanation is speculative given the high uncertainty of the significance of the results.

Moving on, we next consider the impact of the farms on winds from the south in Fig. 14. Here, the wind speed and time of day bins are the same as when we considered the Chisholm View farm impacts. We choose the southern wind direction because this is the most common wind direction, is associated with high occurrences of LLJs, and has farm impacts (Whiteman et al., 1997). We use 2018 as the start of the post-construction period in this case to ensure that the Armadillo Flats wind farm was operational throughout the post-construction period. The impact of the farms is roughly similar in terms of Δ to the case downstream of Chisholm View. As before, though there are a few cases where the permutation test indicates a significant pre/post difference, the intra-period drift test rejects several of these cases. The only strong candidates for farm impacts are

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

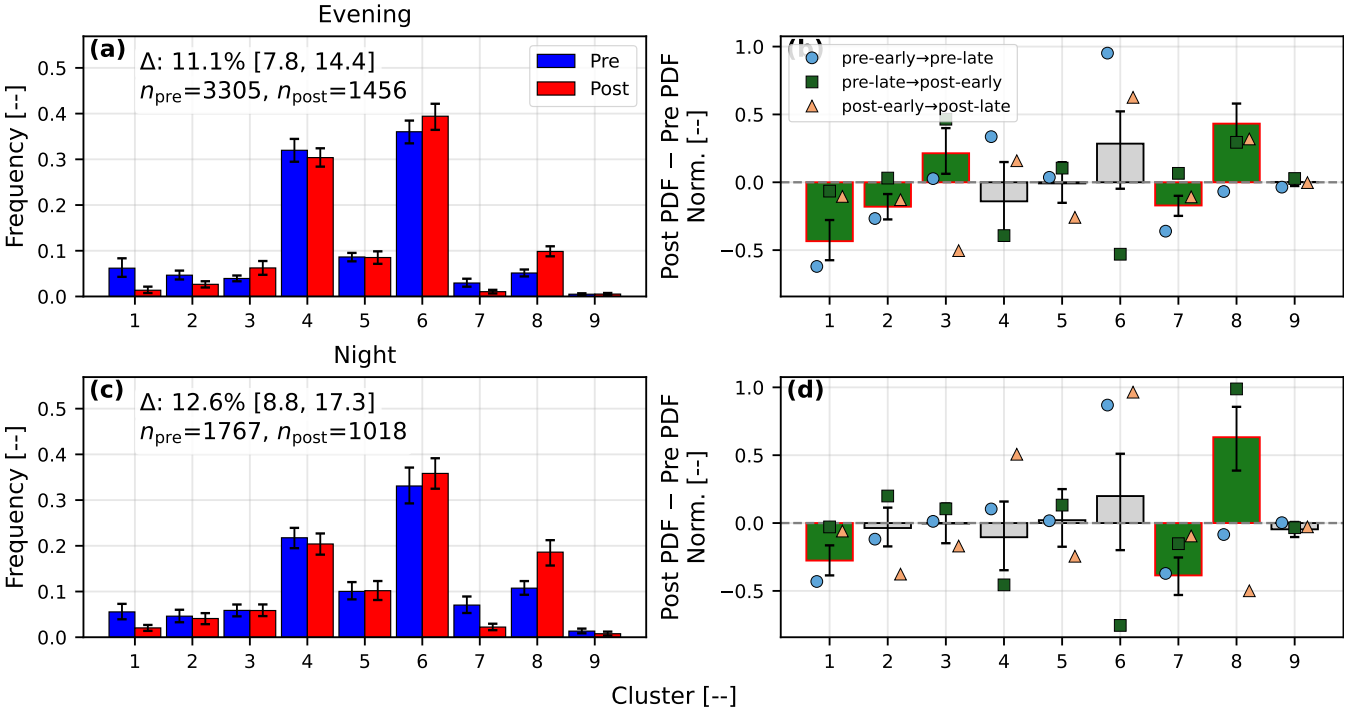

Figure 14. PDFs and difference in PDFs between pre- and post-construction data for wind directions from the south. Wind speed and time of day bins, infill coloring, and the red bars all follow Fig. 13 conventions. The wind direction sector is [180°, 200°].

the evening cluster 3 increase and the night cluster 8 increase. If we ignore the significance tests, there is a weaker impact on clusters 4–6 than in the previous case downstream of Chisholm View. Clusters 1 and 7 decrease, while cluster 8 increases. Taking cluster 8 to be an LLJ cluster that nevertheless has lower near-ground shear than other LLJ clusters, this is consistent with the suggestion in the Chisholm View case that the farms reduce near-ground shear. However, again the significance of these differences is suspect.

Finally, we consider wind directions from 270° to 330°. This sector contains no upstream wind farms (we refer to it as "unimpacted" in this directional sense, not as an assertion that no farm-induced effects can reach it) and allows us to test for any large-scale impact of the farms upstream or beside them. For this sector, we take 2016 to be the start of the post-construction period. This ensures that the two farms flanking the unimpacted sector, Chisholm View and Kay Wind, are both operational. Starting in 2017, the downstream farm, Thunder Ranch, was also operational. The total PDF difference is slightly smaller for this case in terms of Δ than for the previous cases (Fig. 15). Only a few clusters pass the permutation test for significance. Of these, only the increase in evening cluster 8 also passes the intra-period drift test. Ignoring the significance tests, we see a similar pattern to the previous cases: low-shear clusters like 4–6 and 8 tend to increase, especially at night, while high-shear clusters like 1, 2, and 7 tend to decrease. Establishing these impacts as significant remains difficult.

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

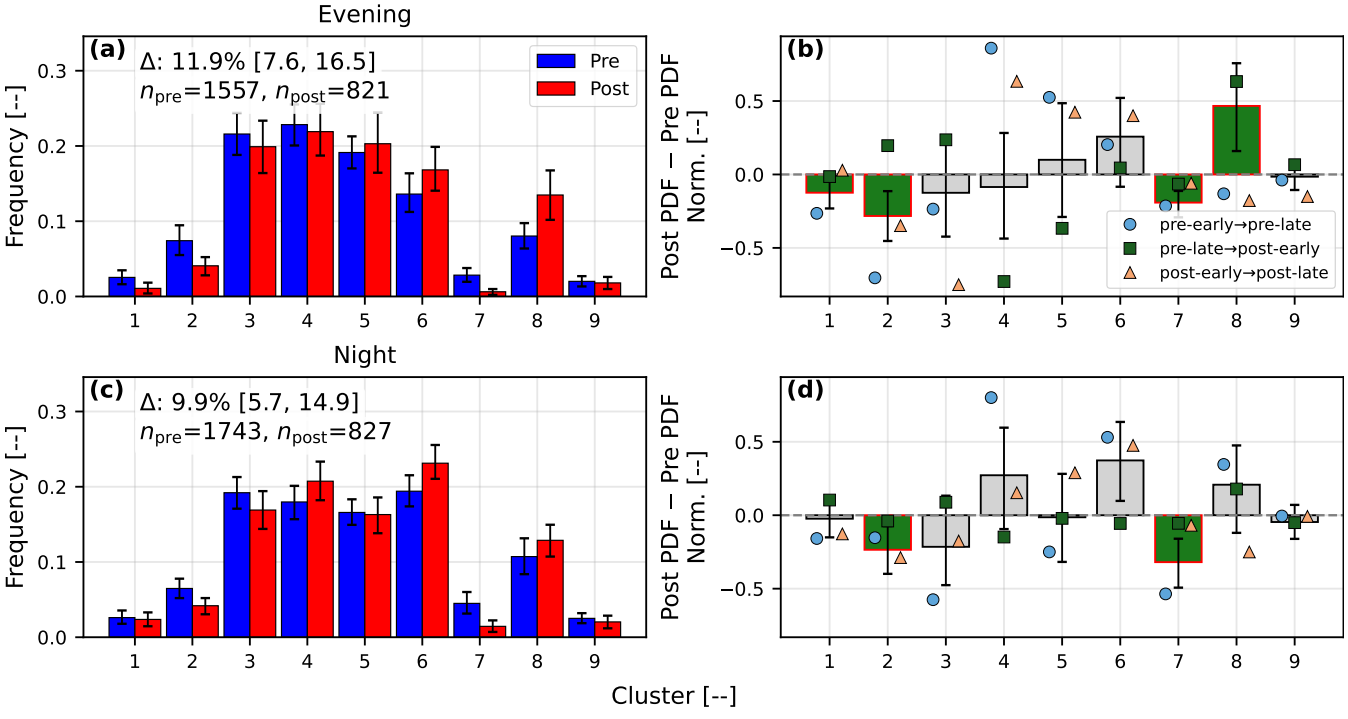

Figure 15. PDFs and difference in PDFs between pre- and post-construction data for wind directions from the northwest, which is unimpacted by upstream wind farms. Wind speed and time of day bins, infill coloring, and the red bars all follow Fig. 13 conventions. The wind direction sector is [270°, 330°].

In summary, we note that, in general, pre/post-construction variability in cluster occurrences suggests that clusters with high shear near the ground are replaced by clusters with low shear near the ground due to the presence of the wind farms. This happens with non-LLJ clusters and within LLJ clusters (i.e., LLJs with high near-ground shear decrease while LLJs with low near-ground shear increase). However, establishing these impacts as significant is extremely difficult. One possible confounder could be different distributions of background conditions for the different time periods. Since each background-condition bin defines a relatively narrow sector in wind direction, we only consider the background wind speed. Figure 16 plots the histograms of the background wind speed for all sectors of the wind direction and times of day.

Considering Fig. 16, it is difficult to conclude that the background conditions are significantly different pre-/post-construction for the sectors considered. Although there is some difference in the PDFs, this variation falls inside the error bars. Recalling the results from Fig. 5 and the fact that the average wind speeds in the pre-, during-, and post-construction periods were essentially identical, it seems better to conclude that the background conditions are roughly identical over the sectors and time periods considered. It remains unclear if trends in cluster occurrence could emerge despite identical background conditions (i.e., cluster occurrences are under-determined by background conditions). If this is not the case, then we can at least conclude that the random fluctuations in the background conditions (e.g., Fig. 5) are too large for us to consider the cluster impacts we

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

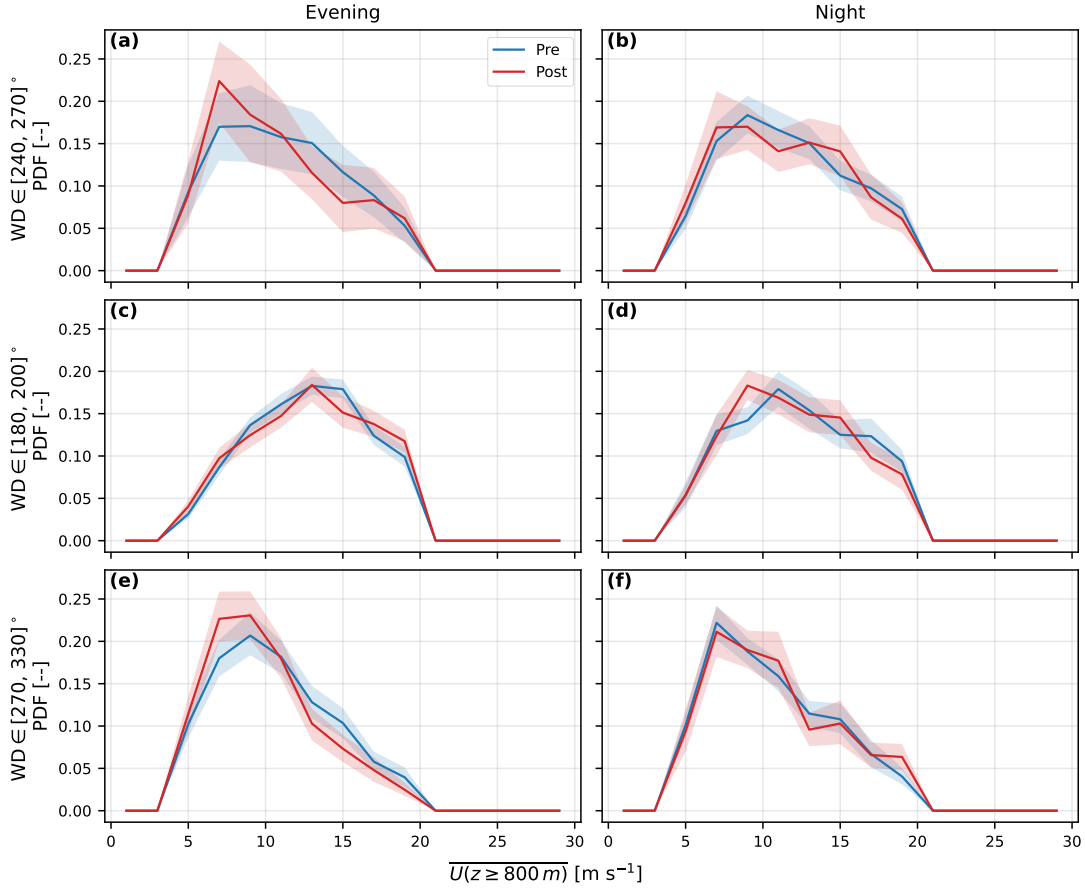

Figure 16. Histograms of background wind speed (averaged over $z \geq 800$ m) for each wind-direction sector and time-of-day bin used in the cluster PDF analyses, comparing pre- and post-construction periods. Bin colors and conventions follow Fig. 13. The shaded region indicates the 90% CI from MBB.

see here to be sufficiently converged. In other words, the magnitude of the farm impact seems to be less than the impact of random fluctuations in cluster occurrences, which are possibly linked to random fluctuations in background conditions.

Having noted the difficulties in separating temporally random fluctuations from the farm impacts, we close with a comparison of LLJs and clusters across the multiple auxiliary ARM sites. By comparing occurrences from these sites over overlapping
505 temporal periods, we remove the impact of background climatological variability plaguing the temporal analysis of a single site. We perform two analyses in this direction. First, we simply plot the LLJ detection rates against wind direction for the different sites. If there is a farm impact, we would expect a signature on the LLJ rates, as some sites experience the farm effect before or after others based on relative position. Unfortunately, we cannot compute agreement metrics, as these require simultaneous time stamps, and this condition heavily reduces the available multi-site data. Instead, we only consider LLJ

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

occurrence rates. Then we perform a single comparison between two sites that should be impacted and unimpacted based on farm positions. We compare the cluster occurrence rates to check for farm impact.

Starting with the multi-site LLJ occurrence rate check, we plot the site LLJ rates in Fig. 17. Here, the wind direction is the average background wind direction across the four sites. The LLJ rates are reported for wind direction bins 15° wide. The figure does not show any clear farm impacts. For many of the wind directions, all sites report similar LLJ rates. Site I9 tends to report fewer LLJs than the other sites but does so systematically for all wind directions and, therefore, could have data-related issues rather than a real site difference. Other divergences are difficult to explain relative to the farm layout. For instance, from wind directions around 120° to 180°, C1 reports fewer LLJs than I8 and I10 but more than I9. Yet referring to Fig. 2, I9 should be unwaked for this range. C1, I8, and I10 are all downstream of similar numbers of turbines. Thus, if there is a farm-related reduction of LLJs, we would expect to see similar, low rates at C1, I8, and I10 compared to I9. Similarly, I10 has fewer LLJs for westerly winds than C1 despite C1 being downstream of Chisholm View. In summary, then, there is no clear farm-related signature on the LLJ rates as a function of wind direction for these sites.

Next, we consider whether there is a difference in profiles between the sites that is masked by a simple comparison of LLJ rates. So, we compare cluster occurrence rates between C1 and I10 sites for westerly wind directions. In particular, we keep the same C1 sector downstream of Chisholm View to isolate a farm impact. For the I10 data, we select the same time of the day (TOD) and wind speed limits but set the wind direction between 250° and 330°. This is done to ensure that I10 is not waked by Chisholm View and to increase data availability. We assume that the slight difference in wind direction bins will not have an impact on the occurrence rates, which may not be the case. Nonetheless, a wide wind direction bin is needed to ensure I10 rates

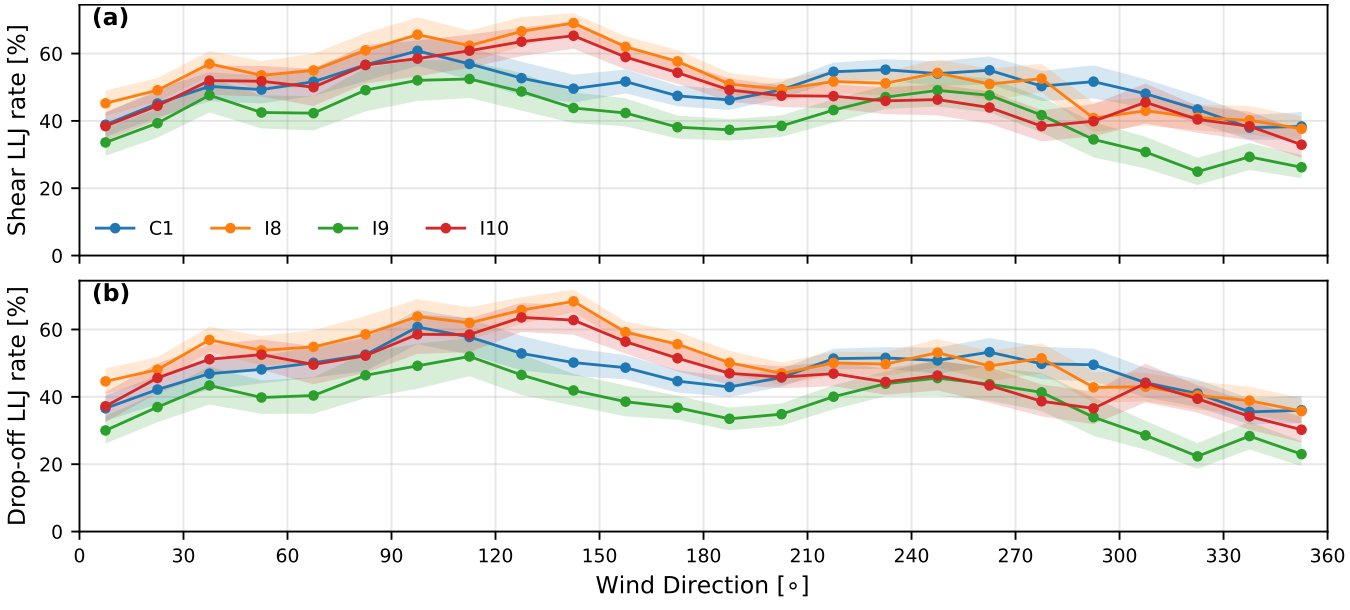

Figure 17. LLJ occurrence rates for all the ARM sites during the overlapping data period from 2020-2024 (all post-construction). Shaded regions indicate 90% MBB CIs.

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

converge. To assign cluster labels to I10 profiles, we select the cluster centroids from C1 closest to each time stamp. In other words, we apply a single k-means pass to the I10 data using C1 centroids. Figure 18 reports the difference in the occurrences of the two sites.

Referring to Fig. 18, we note that the difference between the sites is large ($\Delta \approx 25\%$). It seems consistent that clusters 1–3 decrease and clusters 6 and 8 increase. Cluster 4 also increases during the evening. This is broadly similar to the previous farm impacts we have observed—a decrease in high-shear clusters 1–3 and an increase in low-shear cluster 6. It is odd that cluster 8 increases so significantly and occurs with much higher absolute frequency than clusters 7 or 9. In other words, the increase in cluster 8 is not due to a decrease in clusters 7 or 9—a conversion of high-shear LLJs to a low-shear LLJ—but is due to some other mechanism. In summary, this comparison seems to offer some support to the claim that the presence of the farms reduces the occurrence of high near-ground shear profiles, but the impact on LLJs remains difficult to explain.

5 Discussion

Despite the recognized importance of LLJ–wind farm interactions, relatively few published works have utilized real-world data, and to our knowledge, none have considered the long-term meteorological impact of wind farms on LLJ occurrences using

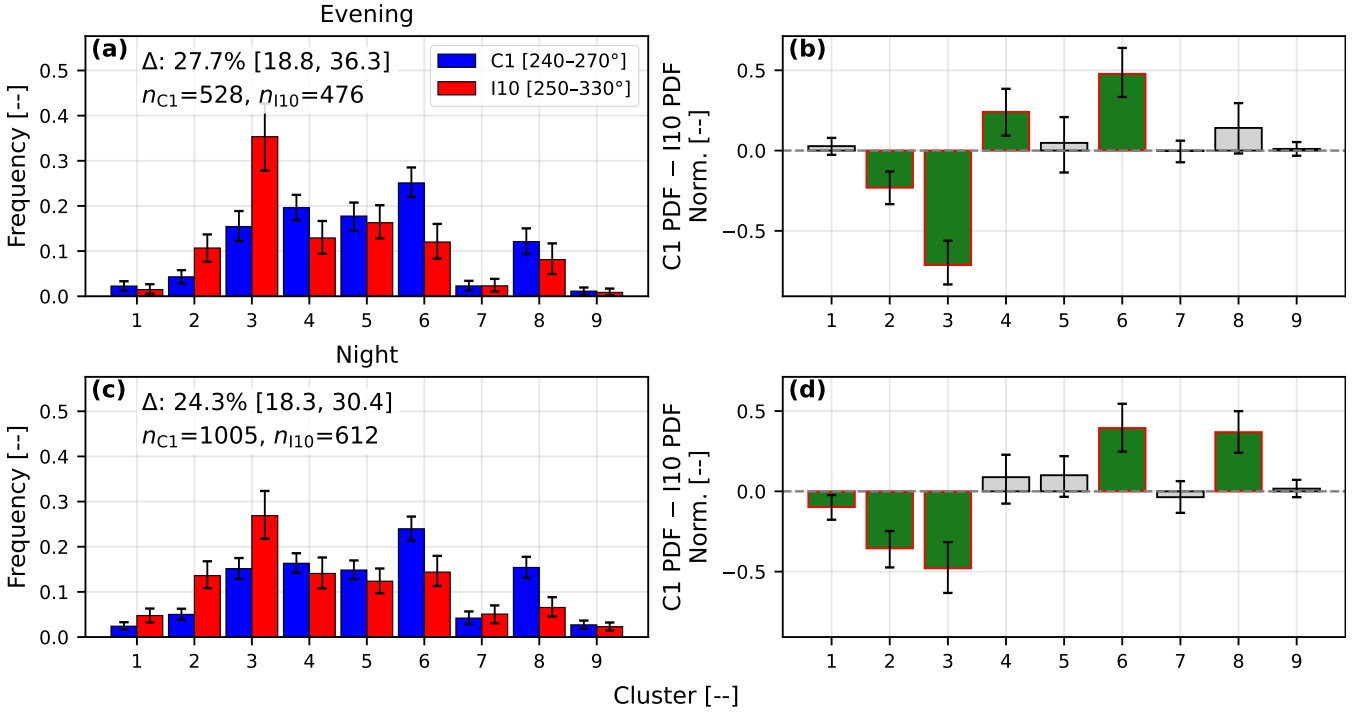

Figure 18. Comparison of the cluster occurrence rates at C1 for wind directions [240°, 270°] and I10 for wind directions [250°, 330°]. Wind speed and TOD bins are identical to Fig. 13.

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

radar observations of this duration. Here, we try to estimate the impact of the construction of wind farms on the occurrence rates and morphology of LLJs using long-term radar profiling data. Using traditional LLJ detection methods and novel data-driven techniques, we note that LLJ occurrences tend to remain broadly stable during and after the construction of farms. Specific wind speed profile morphology modulations are observable but difficult to establish as significant against random background variabilities.

First, the difference in LLJ detection methods, for this site and set of measurements, is not significant. Indeed, over 90% of the time, both the drop-off method and the shear method agree on whether an LLJ is occurring. Although previous results have suggested significant differences in LLJ statistics depending on the method (Aird et al., 2021) and others have suggested that certain methods are suitable for wind energy applications over others (Hallgren et al., 2023), we do not see evidence for those conclusions here. While the mean characteristics of shear-only or drop-off-only profiles are indeed somewhat different, these conditions account for a small portion of the data. Thus, for the purpose of simple LLJ detection, it seems like either method is useful.

Second, the LLJ occurrence rates computed using the shear and drop-off methods change little during or after farm construction compared to the pre-construction period. Supposing that these flags were sufficient to explain power variability in wind farms due to LLJ presence, then one would conclude that there is no significant impact on the LLJ-induced power variabilities for wind farms due to the construction of neighboring wind farms. Likewise, one would not expect the LLJ-induced power variabilities to shift before and after construction of a farm if LLJ measurements were only taken pre-construction. It may not be the case, however, that the binary shear and drop-off LLJ flags are sufficiently explanatory. Indeed, we note additional profile morphologies using the data-driven analysis that are concealed by a simple flag-based analysis. However, we did not find evidence of significant impacts of wind farm construction on LLJ detection rates determined by traditional detection methods.

This leads to the third point: there could be significant wind speed profile morphology modulations due to the presence of the wind farms that are not detectable by the simple binary traditional flags. From our data-driven analysis, we note that cluster occurrences can change dramatically between pre- and post-construction periods (in magnitude only, not necessarily in statistically significant ways). In general, these changes suggest a reduction in the occurrence of profiles with high near-ground shear, whether LLJ or not, and an increase in profiles with lower near-ground shear, again, whether LLJ or not. However, there is large background variability, and it is very difficult to establish these modulations as due to the farms rather than to random background fluctuations.

Fourth, setting aside concerns regarding the significance of the observed effects momentarily, it is generally the case that, of the LLJ clusters that decrease post-construction compared to pre-construction, the decrease is stronger at night than during the evening. Given that inertial oscillations, suggested as the main formation mechanism for the nighttime LLJ, should begin forming LLJs during the evening, this seems to suggest some relevant interaction between the wind farms and the mesoscale motions, which promote LLJ formation. Perhaps the farms disrupt the LLJ formation mechanism, leading to a reduction in the nighttime LLJs.

Fifth, assuming the presence of the wind farms reduces near-rotor shear (between the lowest valid radar gate at $\sim$125 m and the rotor tip) for comparable background wind speeds, the result is that wind speeds in the near-rotor layer increase

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

relative to pre-construction conditions. The radar does not measure sufficiently low to directly resolve the wake layer of the wind farm, so the speed-up immediately above the wake layer (potentially induced by continuity) is inferred rather than measured. Nonetheless, the lowest measurement gate is just above top-tip and the reduction in shear suggests a mixing effect that propagates above the wake layer.

6 Conclusions

In this work, we have investigated the impacts of the construction of several large wind farms on the atmospheric boundary layer (ABL), in particular, on the occurrence and characteristics of low-level jets (LLJ). Given the potential interactions between wind farms and LLJs, including modulation of turbine loads, wake recovery, and total farm power, when designing a wind farm it is important to understand the characteristics of LLJs that will occur after the farm's construction. However, it might be insufficient to rely only on historical data to characterize expected LLJs, as the presence of wind farms has been shown, under simulated conditions, to reduce the rate of occurrence of offshore LLJs (Quint et al., 2025). Here we have used long-term historical wind velocity profiles, for the first time using radar observations of this duration, to investigate whether a similar effect on LLJs can be detected in experimental data.

To investigate the impact of the farms, we have performed two main analyses. For the first, we contrast two common methods for defining and detecting LLJs and demonstrate the impact of the farms on the LLJs detected by these methods, as well as on the detection rates. Next, we have used data-driven methods to define characteristic wind profiles for the site and assigned each measurement to a cluster represented by these characteristic profiles. By comparing cluster occurrence rates before and after the construction of the wind farms, we can demonstrate the impact of the farms without the need for the precise definitions required by more traditional techniques.

In investigating the impact of the farms using more traditional methods, we have compared the drop-off method, which defines LLJ profiles as profiles with a sufficient difference between the nose wind speed and the minimum wind speed above the nose, and the shear method, which defines LLJ profiles as having sufficient positive and negative shear below and above the LLJ nose, respectively. Using a quadrant analysis, we have analyzed the four possible combinations of agreement and disagreement between these two detection methods. This analysis indicates that the methods generally agree (around 90% of cases are in agreement). By plotting the average wind profiles for each of the four cases pre-, during-, and post-construction, we have noted that the reason for the decrease in shear-detected LLJs seems to be an increase in wind speed near the ground and a corresponding decrease in near-ground shear. The cases in which only the drop-off method labels LLJs tend to have wind speed maxima at the lower measurement heights. Given that these LLJs would have noses closer to the typical rotor areas, they are highly relevant for wind energy applications. This leads us, in contrast to recent works, to suggest that the drop-off method could still be suitable for wind energy applications, at least in cases where there are limited measurements below the turbine top-tip (Hallgren et al., 2023). In any case, the very large agreement between the two methods suggests that choosing one method over the other will not have strong impacts.

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

As an alternative to LLJ studies using these traditional detection methods, which rely on tuned limits on drop-off and shear, we have introduced a data-driven method to study the impact of farms on wind profile morphology. In particular, we have used agglomerative clustering to hierarchically cluster all wind speed profiles, normalizing by the upper-elevation average wind speed. We then iteratively sub-cluster the data until clusters are either too small or have converged, only sub-clustering when the resulting mean sub-cluster profiles are significantly different with respect to the variance. These clusters provide a characterization of the site in a data-driven fashion.

By comparing the differences in the cluster occurrence rates before and after the construction of the relevant wind farms, we can demonstrate the impact of the farms on the morphology of the wind speed profiles. By defining the time of day and the average wind speed and direction from the upper elevations as our background conditions, we have used bins in the background conditions to isolate conditions directly downstream of wind farms and conditions for which no wind farms lie upstream of the radar. Statistical significance is difficult to achieve due to random background variabilities despite the absence of any clear evidence for a long-term trend in background conditions. Nonetheless, in all sectors, we see consistent suggestions of certain effects, such as the reduction of near-ground shear. Thus, farm effects might be relevant even when the measurement location is upstream of the nearest farms, which might suggest a broader impact of wind farms on the LLJ formation process and the general regional micrometeorology. When these effects occur downstream of a wind farm, they could be attributed to wakes. However, the fact that such effects might occur in unimpacted sectors suggests broader regional impacts also occur. Furthermore, the cluster analysis has suggested that the nighttime occurrence rates of the LLJs are more impacted than the evening rates. This suggests that the presence of the farms disrupts LLJ formation, which likely occurs during the evening, in such a manner that the disruption is only observed at night, when the LLJs are expected to be fully formed. However, these results are obtained in a totally data-driven fashion and therefore do not depend on the specific definition of LLJ.

Ultimately, the rather striking result of this analysis is the difficulty in establishing statistically significant farm impacts. From previous simulations (e.g., Quint et al., 2025), we expected to see strong farm impacts. Instead, we see no such strong evidence. In fact, we conclude that, if there is farm-induced LLJ variability, it is of the same order as random background fluctuations. As the data set we used already contains nearly 30 years of data, it seems that increasing the length of data collected will not produce more significant results. If wind farms tend to operate for around 20 years before repowering or other modifications, then longer time periods will be insufficient to establish significant effects. Instead, we note two main outcomes of this analysis. First, wind farm developers may not need to worry about wind farm modifications of LLJs. Given AEP estimates that incorporate historical LLJ characteristics, this study suggests that post-construction AEPs will not be strongly impacted by farm-induced LLJ modifications. Second, if longer temporal analyses are not useful in establishing significant impacts, then future analyses will need to focus on synthesizing experimental and simulated data to better understand the two-way coupling between wind farms and the atmosphere. Although this work does not establish significant LLJ impacts due to farm construction, it does seem that clear farm impacts occur beyond simple blockage/wake/sideways redirection, and these deserve further attention.

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Data availability. The data used in this work are publicly available on the ARM website (`https://adc.arm.gov/discovery/`) and can be found at the relevant citations (Coulter et al., 2026a, b, c; Muradyan and Ermold, 2026).

Appendix A: Data interpolation and filtering

Outliers can interfere with traditional LLJ detection methods, which rely on detecting peak wind speeds or gradients of certain magnitudes. A single outlier could meet the requirements for an LLJ even though the rest of the profile is a typical monotonic or log-layer profile. So, we apply an additional filtering step to reduce the impact of outliers. The full filtering and interpolation procedure described below is summarized as pseudocode in Algorithm A1. First, we restrict the radar data between 125 m and 1 km. The lower bound is applied since we observe that the data from the first measurement gate are often corrupted. The upper bound is applied since measurements from higher elevations tend to be corrupted and we are not interested in the upper atmosphere anyway. This results in 15 measurement heights, each spaced approximately 62 m apart. Because the first retained gate (125 m) is itself occasionally anomalous, we additionally null that gate on any profile in which it differs from the adjacent gate by more than $5\,\mathrm{m\,s^{-1}}$.

Next, we interpolate the data in two steps. Our goal is to preserve as much data as possible to achieve a statistically significant characterization of LLJ characteristics before and after farm construction. To this end, we first reindex the radar data to a uniform hourly time step, injecting full NaN profiles when a time stamp is missing. Then, for any NaNs in the data set, we check to see if they are neighbored by real values during the hours immediately preceding and following the time stamp of interest. If a NaN sample is bounded in this way, we linearly interpolate in time and infill with the interpolated value. If the NaN sample is also bounded immediately above and below in height by real values, we further interpolate in height and average the two values. These interpolated data are considered "real" and will be propagated into later analyses. Although such an interpolation approach will not smooth existing LLJs to the point where they will no longer be detectable, it could inject profiles during transition points (e.g., between an LLJ and a non-LLJ profile) that are non-LLJ. As such, it could potentially reduce the overall number of LLJs detected. Nonetheless, we consider this risk minor enough that we use this interpolation approach to maximize data availability.

Next, we use a simple PCA-based filter to reject points with high suspected noise (Lorenz et al., 1956; Hannachi et al., 2007; Uma Maheswara Rao et al., 2014). The main idea is to compress the data using a limited number of principal components, reconstruct a de-noised data set, and only keep the original data when it is close to the de-noised values. First, we interpolate all missing values in height, since the PCA algorithm we use cannot accept NaN inputs (Pedregosa et al., 2011). Next, we normalize the wind speed profiles so that the PCA focuses on the shape of the profiles rather than the magnitudes. Later, we will bin profiles based on the average wind speed above 800 m, so for consistency we normalize the profiles here by this average wind speed. Then, we fit the PCA to this data set. We choose to reconstruct at least 97.5% of the variance of the wind speed measured at each height throughout the data set. A unique PCA basis is fitted to each site to preserve local site characteristics and, as a result, each site individually determines the number of principal components to reconstruct at least 97.5% of the variance.

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

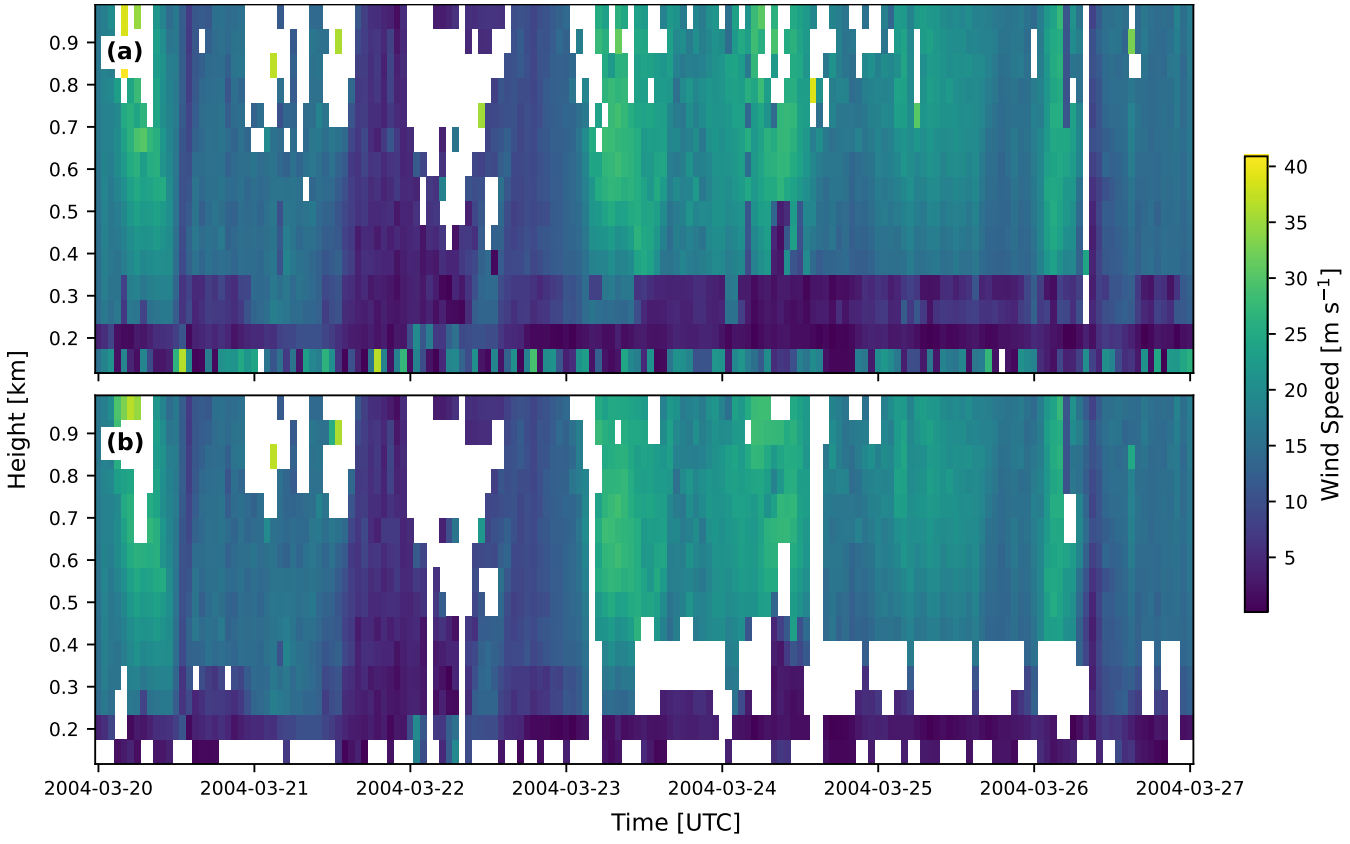

Figure A1. Example of the filtering procedure showing (a) unfiltered, raw data and (b) filtered data.

We reconstruct each wind speed profile as a linear combination of the PCA singular vectors multiplied by the respective PCA coefficient (namely the inner product of the initial wind speed profile and a PCA singular vector), then further multiplied by the average upper-elevation wind speed of the original profile. Measurements in which the initial value differs from the denoised value by more than 1.5 m s^{-1} in absolute value are rejected. The 97.5% variance threshold and the 1.5 m s^{-1} rejection threshold were empirically selected. Sensitivity to these choices is not deemed critical for the results generated and has not been exhaustively explored here. For samples that do not differ significantly from the denoised samples, we keep the original samples, rather than the denoised ones, to avoid smoothing, which might interfere with LLJ detection. After using the PCA filter to reject points with high suspected noise, we again interpolate in time and height as described above. To demonstrate the application of the filter and interpolating steps, Fig. A1 illustrates an example week of data before and after the filtering procedure is applied. In particular, note that near-ground wind speeds are often erroneous (constant near-zero values despite high upper wind speeds) and are correctly rejected.

Finally, we require that each profile has at least eight valid points to be considered. Now we summarize the filtering steps and report the impact on data availability of each step on C1 data. We start with 213,163 time stamps, 16,899 of which are

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

NaN profiles. Reindexing to a regular hourly time grid introduces 29,789 full NaN profiles, bringing the total time stamp count up to 242,952. Next, we interpolate in time and height. This infills 2.7% of the total NaN points. We then drop profiles that are either entirely NaN after interpolation or whose peak wind speed is below $5\,\mathrm{m\,s^{-1}}$, since near-calm profiles do not contain the LLJs of interest; this removes 58,859 time stamps (24%). A subsequent height-interpolation step additionally rejects 5370 profiles with fewer than five valid gates, as these are inadequate for PCA fitting or later consideration. The PCA filter follows, rejecting approximately 2.8% of the gates (over all times and heights). This is followed by another time-gap interpolation step, which imputes 2.0% of the remaining NaN points, and then the rejection of times with fewer than eight real points. This rejects an additional 9866 time stamps (4%).

Algorithm A1 Data filtering and interpolation procedure.

Require: Raw radar wind speed profiles $V(h,t)$

1: Restrict gates to $125\,\mathrm{m} \le h \le 1\,\mathrm{km}$ (15 gates).
2: **for all** profiles $V(\cdot,t)$ **do**
3: **if** $|V(125\,\mathrm{m},t) - V(187\,\mathrm{m},t)| > 5\,\mathrm{m\,s^{-1}}$ **then**
4: $V(125\,\mathrm{m},t) \leftarrow$ NaN.
5: **end if**
6: **end for**
7: Reindex time stamps to a uniform hourly grid; insert full-NaN profiles where missing.
8: For each NaN bounded in time by valid samples at $t \pm 1\,\mathrm{h}$, linearly interpolate in time; if additionally bounded in height by valid samples, average with the height interpolant.
9: Drop entirely-NaN profiles and profiles with peak wind speed below $5\,\mathrm{m\,s^{-1}}$.
10: Linearly interpolate remaining NaNs in height (required for PCA fitting); drop profiles with fewer than five valid gates.
11: Normalize each profile by $\langle V \rangle_{h \ge 800\,\mathrm{m}}$.
12: Fit PCA on the normalized profiles; retain the minimum number of components such that the reconstructed variance at every gate is $\ge 97.5\%$.
13: Reconstruct each profile V_{recon} from the retained components, re-scaling by $\langle V \rangle_{h \ge 800\,\mathrm{m}}$.
14: **for all** gates (h,t) **do**
15: **if** $|V_{\mathrm{orig}}(h,t) - V_{\mathrm{recon}}(h,t)| > 1.5\,\mathrm{m\,s^{-1}}$ **then**
16: Reject the gate (set to NaN).
17: **else**
18: Retain $V_{\mathrm{orig}}(h,t)$ (do not substitute the denoised value).
19: **end if**
20: **end for**
21: Repeat the time/height interpolation step on the post-PCA NaNs.
22: Drop any profile with fewer than eight valid gates.

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Appendix B: Clustering method

To briefly sketch the agglomerative clustering procedure, the algorithm starts by considering every observation as a unique cluster. It computes the distance between each pair of clusters. Usually, this distance is simply the Euclidean distance across all the features (in this case, distance in normalized wind speed). The two closest clusters are then merged into a new cluster. This is repeated, merging pairs of clusters one at a time, until there is a single cluster. The result is a dendrogram we can traverse in reverse order to resolve any given cluster into its component clusters. Agglomerative clustering can be very computationally expensive. We use a K-nearest neighbors connectivity matrix to improve computational efficiency. For more details, the interested reader is referred to the Scikit-learn documentation (Pedregosa et al., 2011). The complete clustering procedure used here, including the dendrogram-traversal stopping criterion described below, is summarized as pseudocode in Algorithm B1.

Since our goal is to characterize LLJs in a data-driven fashion and then compare pre-, during-, and post-construction features, we might start by trying to compute the full clustering dendrogram and then back out a set of suitably resolved clusters (i.e., clusters with unique mean wind speed profiles compared to other clusters). From these characteristic clusters, we could select LLJ clusters based on the mean wind speed profiles. We could then investigate the occurrence rates of these profiles before and after the construction of the farms to assess the impact of the farm constructions on LLJs. However, this approach has a problem. If we generate too many clusters at the beginning, then we will need to reintroduce some criteria to help us differentiate LLJ clusters and non-LLJ clusters. This takes us back to the problems of LLJ definition that we have already encountered with traditional detection methods. To take advantage of the metric-free data-driven approach, we need to generate as few clusters as possible.

For this reason, our approach is as follows. First, linearly interpolate all missing values in the wind speed profiles. This is simply because the Scikit-Learn Agglomerative Clustering package used cannot accept missing values (Pedregosa et al., 2011). Then, normalize all profiles by the average of upper-elevation wind speeds (≥ 800 m). This is to help the clustering focus on profile shapes rather than magnitudes. Neglecting this step would result in clusters differentiated by the wind speed rather than the profile shape and would make surfacing an LLJ cluster unlikely. Next, compute the agglomerative dendrogram on all data. Then, we step through the dendrogram one node at a time, further resolving sub-clusters. For a given cluster, we decide whether to further sub-cluster at that cluster or to terminate the branch at that cluster (i.e., keeping the current cluster as "final"), as follows. For a given cluster under consideration, we use the dendrogram to define the sub-clusters A and B. At each height, we compute the mean of each sub-cluster and the variance of the parent cluster (i.e., the variance of all data at that height across both sub-clusters A and B). For height i, T_i is defined as:

$$T_i = \frac{(\mu_{A,i} - \mu_{B,i})^2}{\sigma_i^2}, \tag{B1}$$

where $\mu_{A,i}$ and $\mu_{B,i}$ are the sub-cluster mean values at height i and σ_i^2 is the total variance at that height. Subsequently, for the entire profile, we define:

$$T_{\text{obs}} = \sum_{i=1}^{N_H} T_i, \tag{B2}$$

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

where N_H is the total number of heights. By setting a limit T_{thresh}, we can only sub-cluster the original target cluster into sub-clusters A and B when $T_{\text{obs}} \geq T_{\text{thresh}}$. In other words, we only separate the cluster into sub-clusters when the mean profiles of these sub-clusters are sufficiently different compared to the variance. T_{thresh} becomes the main user-determined hyper-parameter of this approach. We set it to 10 somewhat arbitrarily; we found that $T_{\text{thresh}} = 10$, which produces 9 clusters, allowed a good balance between many, detailed clusters and few, easily interpretable clusters. Nonetheless, different values of T_{thresh} would lead to different levels of detail. Note that over 15 heights, $T_{\text{thresh}} = 10$ equates to the average profile separation being approximately two-thirds of the variance. We also set a minimum population threshold of 1%. A cluster cannot be smaller than 1% of the total data set and, if sub-clustering would produce such a cluster, then sub-clustering is not permitted.

Finally, before using these clusters, we note that, due to the dendrogram being formed by iterative merging, it is often the case that an observation could be closer in the feature space to the centroid of a cluster to which it does not belong. To mitigate this issue of possibly mislabeled observations, we perform a single pass of re-labeling. We take the agglomerative cluster labels and use these to define cluster centroids, taken as the mean of all profiles of the same label. We then update the cluster labels so that each observation belongs to the cluster with the closest centroid. In other words, we perform a single k-means pass. In this way, we can avoid the issues of k-means by using the agglomerative approach to define cluster centroids and then refine with a single k-means pass to address mislabeled agglomerative values.

We note here that we can later split the data into pre-, during-, and post-construction data sets using either the bulk climatological approach or the farm-specific approach. It is not critical to perform these splits before clustering or to perform multiple sets of clustering on the multiple periods of data. In the first case, the only reason we might want to split into period data sets would be to normalize the probability density functions (PDFs) before recombining and clustering. We might hope to avoid bias in the clusters by first reducing biases in the data sets. However, the agglomerative approach is already well-suited for detecting important clusters in biased data sets. Thus, this resampling of the PDFs to remove bias is not needed prior to clustering. In the second case, performing multiple sets of clusters would result in a unique set of period labels. Inter-comparison would be impossible. Thus, we simply cluster all the data at once and then only apply splits or resampling later.

Author contributions. C.M. performed the data analysis and drafted the manuscript. T.C. and R.S. contributed to the conceptual development through regular meetings, provided guidance throughout the project, and reviewed and edited the manuscript. G.V.I. acquired funding, served as principal investigator, closely supervised the data analysis, and contributed to the preparation and revision of the manuscript.

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

Acknowledgements. This work was supported in part by the National Science Foundation under Award Nos. 1916715 and 1916776 (I/UCRC for Wind Energy, Science, Technology, and Research) and by the members of WindSTAR I/UCRC. Any opinions, findings, and conclusions

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

Algorithm B1 Agglomerative clustering with variance-ratio stopping criterion.

Require: Wind speed profiles $\{V_n(h)\}_{n=1}^{N}$; thresholds $T_{\mathrm{thresh}} = 10$, $p_{\min} = 0.01$

Linearly interpolate missing values in each profile.
Normalize each profile by $\langle V \rangle_{h \geq 800\,\mathrm{m}}$.
Build a K-nearest-neighbors connectivity matrix from the normalized profiles.
Compute the full agglomerative dendrogram using the connectivity matrix.
Initialize the working cluster set $\mathcal{C} \leftarrow \{\mathrm{root}\}$.
while any cluster in $\mathcal{C}$ has not yet been tested **do**
    Pop an untested cluster $C \in \mathcal{C}$; let A, B be its dendrogram children.
    Compute T_i at each height (Eq. (B1)) and T_{obs} (Eq. (B2)).
    if $T_{\mathrm{obs}} \geq T_{\mathrm{thresh}}$ **and** $|A| \geq p_{\min} N$ **and** $|B| \geq p_{\min} N$ **then**
        $\mathcal{C} \leftarrow (\mathcal{C} \setminus \{C\}) \cup \{A, B\}$.
    else
        Mark C as final (do not subdivide).
    end if
end while
For each final cluster, compute its centroid as the mean of member profiles.
Re-assign each observation to the cluster with the nearest centroid (single k-means pass).

Ensure: Final cluster label for each profile.

or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the sponsors. This research was also funded by the NSF CBET, Fluid Dynamics CAREER program, Award No. 2046160.

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

eawe european academy of wind energy
Open Access WIND ENERGY SCIENCE DISCUSSIONS

References

Aird, J. A., Barthelmie, R. J., Shepherd, T. J., and Pryor, S. C.: WRF-simulated Low-Level Jets over Iowa: Characterization and Sensitivity Studies, Wind Energ. Sci., 6, 1015–1030, https://doi.org/10.5194/wes-6-1015-2021, 2021.

Banta, R., Newsom, R. K., Lundquist, J. K., Pichugina, Y. L., Coulter, R. L., and Mahrt, L.: Nocturnal Low-Level Jet Characteristics Over Kansas During Cases-99, Boundary-Layer Meteorology, 105, 221–252, https://doi.org/10.1023/A:1019992330866, 2002.

Calaf, M., Meneveau, C., and Meyers, J.: Large Eddy Simulation Study of Fully Developed Wind-Turbine Array Boundary Layers, Physics of Fluids, 22, 015 110, https://doi.org/10.1063/1.3291077, 2010.

Chatterjee, T., Jayaraman, B., Yellapantula, S., and Quon, E.: Modeling Offshore Wind Farm Performance in Coastal Low-Level Jets Using Coupled Mesoscale-Microscale Large Eddy Simulations, PRX Energy, 4, 033 018, https://doi.org/10.1103/vj2s-gflw, 2025.

Coulter, R., Muradyan, P., and Martin, T.: Radar Wind Profiler (915RWPWINDCON), 1998-08-03 to 2019-08-19, Southern Great Plains (SGP), Central Facility, Lamont, OK (C1), https://doi.org/10.5439/1993735, 2026a.

Coulter, R., Muradyan, P., and Martin, T.: Radar Wind Profiler (915RWPWINDCON), 2013-10-24 to 2023-09-29, Southern Great Plains (SGP), Lamont, OK (NW radar wind profiler site, Intermediate / Auxiliary) (I10), https://doi.org/10.5439/1993735, 2026b.

Coulter, R., Muradyan, P., and Martin, T.: Radar Wind Profiler (915RWPWINDCON), 2014-06-20 to 2023-09-29, Southern Great Plains (SGP), Tonkawa, OK (NE radar wind profiler site, Intermediate / Auxiliary) (I8), https://doi.org/10.5439/1993735, 2026c.

Debnath, M., Moriarty, P., Krishnamurthy, R., Bodini, N., Newsom, R., Quon, E., Lundquist, J. K., Letizia, S., Iungo, G. V., and Klein, P.: Characterization of Wind Speed and Directional Shear at the AWAKEN Field Campaign Site, Journal of Renewable and Sustainable Energy, 15, 033 308, https://doi.org/10.1063/5.0139737, 2023.

Doosttalab, A., Siguenza-Alvarado, D., Pulletikurthi, V., Jin, Y., Bocanegra Evans, H., Chamorro, L. P., and Castillo, L.: Interaction of Low-Level Jets with Wind Turbines: On the Basic Mechanisms for Enhanced Performance, Journal of Renewable and Sustainable Energy, 12, 053 301, https://doi.org/10.1063/5.0017230, 2020.

Forina, M., Armanino, C., and Raggio, V.: Clustering with Dendrograms on Interpretation Variables, Analytica Chimica Acta, 454, 13–19, https://doi.org/10.1016/S0003-2670(01)01517-3, 2002.

Frandsen, S., Barthelmie, R., Pryor, S., Rathmann, O., Larsen, S., Højstrup, J., and Thøgersen, M.: Analytical Modelling of Wind Speed Deficit in Large Offshore Wind Farms, Wind Energy, 9, 39–53, https://doi.org/10.1002/we.189, 2006.

Gadde, S. N. and Stevens, R. J. A. M.: Effect of Low-Level Jet Height on Wind Farm Performance, Journal of Renewable and Sustainable Energy, 13, https://doi.org/10.1063/5.0026232, 2021.

Hallgren, C., Aird, J. A., Ivanell, S., Körnich, H., Barthelmie, R. J., Pryor, S. C., and Sahlée, E.: Brief Communication: On the Definition of the Low-Level Jet, Wind Energ. Sci., 8, 1651–1658, https://doi.org/10.5194/wes-8-1651-2023, 2023.

Hannachi, A., Jolliffe, I. T., Stephenson, D. B., et al.: Empirical orthogonal functions and related techniques in atmospheric science: A review, International journal of climatology, 27, 1119–1152, 2007.

Iungo, G. V., Maulik, R., Renganathan, S. A., and Letizia, S.: Machine-Learning Identification of the Variability of Mean Velocity and Turbulence Intensity for Wakes Generated by Onshore Wind Turbines: Cluster Analysis of Wind LiDAR Measurements, Journal of Renewable and Sustainable Energy, 14, 023 307, https://doi.org/10.1063/5.0070094, 2022.

Künsch, H. R.: The Jackknife and the Bootstrap for General Stationary Observations, The Annals of Statistics, 17, 1217–1241, http://www.jstor.org/stable/2241719, 1989.

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

Lakens, D.: Equivalence tests: A practical primer for t tests, correlations, and meta-analyses, Social psychological and personality science, 8, 355–362, 2017.

Lance, G. N. and Williams, W. T.: A General Theory of Classificatory Sorting Strategies: 1. Hierarchical Systems., Comput. J., 9, 373–380, 1967.

Lloyd, S.: Least Squares Quantization in PCM, IEEE Trans. Inform. Theory, 28, 129–137, https://doi.org/10.1109/TIT.1982.1056489, 1982.

Lorenz, E. N. et al.: Empirical orthogonal functions and statistical weather prediction, vol. 1, Massachusetts Institute of Technology, Department of Meteorology Cambridge, 1956.

Markfort, C. D., Zhang, W., and Porté-Agel, F.: Turbulent Flow and Scalar Transport through and over Aligned and Staggered Wind Farms, Journal of Turbulence, 13, N33, https://doi.org/10.1080/14685248.2012.709635, 2012.

Markfort, C. D., Zhang, W., and Porté-Agel, F.: Analytical Model for Mean Flow and Fluxes of Momentum and Energy in Very Large Wind Farms, Boundary-Layer Meteorol, 166, 31–49, https://doi.org/10.1007/s10546-017-0294-6, 2018.

McQueen, J. B.: Some Methods of Classification and Analysis of Multivariate Observations, in: Proc. of 5th Berkeley Symposium on Math. Stat. and Prob., pp. 281–297, 1967.

Meneveau, C.: The Top-down Model of Wind Farm Boundary Layers and Its Applications, Journal of Turbulence, 13, N7, https://doi.org/10.1080/14685248.2012.663092, 2012.

Moss, C., Puccioni, M., Maulik, R., Jacquet, C., Apgar, D., and Valerio Iungo, G.: Profiling Wind LiDAR Measurements to Quantify Blockage for Onshore Wind Turbines, Wind Energy, 27, 1268–1285, https://doi.org/10.1002/we.2877, 2024.

Moss, C., Letizia, S., Iungo, G. V., and Moriarty, P. J.: Realistic Noise Generation to Enhance Realism of Virtual Lidar Scans, Remote Sensing, 17, 2965, https://doi.org/10.3390/rs17172965, 2025.

Muradyan, P. and Ermold, B.: Radar Wind Profiler (915RWPWINDCNS), 2021-09-03 to 2023-09-28, Southern Great Plains (SGP), Billings, OK (SE radar wind profiler site, Intermediate / Auxiliary) (I9), https://doi.org/10.5439/1599097, 2026.

Paulsen, J., Schneemann, J., Steinfeld, G., Theuer, F., and Kühn, M.: The Impact of Low-Level Jets on the Power Generated by Offshore Wind Turbines, https://doi.org/10.5194/wes-2025-118, 2025.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., and Cournapeau, D.: Scikit-Learn: Machine Learning in Python, MACHINE LEARNING IN PYTHON, 12, 2825–2830, 2011.

Quint, D., Lundquist, J. K., and Rosencrans, D.: Simulations Suggest Offshore Wind Farms Modify Low-Level Jets, Wind Energy Science, 10, 117–142, https://doi.org/10.5194/wes-10-117-2025, 2025.

Schuirmann, D. J.: A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability, Journal of pharmacokinetics and biopharmaceutics, 15, 657–680, 1987.

Segalini, A. and Chericoni, M.: Boundary-Layer Evolution over Long Wind Farms, Journal of Fluid Mechanics, 925, A2, https://doi.org/10.1017/jfm.2021.629, 2021.

Sheridan, L. M., Krishnamurthy, R., Gustafson Jr., W. I., Liu, Y., Gaudet, B. J., Bodini, N., Newsom, R. K., and Pekour, M.: Offshore Low-Level Jet Observations and Model Representation Using Lidar Buoy Data off the California Coast, Wind Energy Science, 9, 741–758, https://doi.org/10.5194/wes-9-741-2024, 2024.

Stensrud, D. J.: Importance of Low-Level Jets to Climate: A Review, Journal of Climate, 9, 1698–1711, 1996.

Uma Maheswara Rao, D., Sreenivasulu Reddy, T., and Ramachandra Reddy, G.: Atmospheric radar signal processing using principal component analysis, Digital Signal Processing, 32, 79–84, https://doi.org/https://doi.org/10.1016/j.dsp.2014.05.009, 2014.

https://doi.org/10.5194/wes-2026-98
Preprint. Discussion started: 17 June 2026

Ward, J. H.: Hierarchical Grouping to Optimize an Objective Function, Journal of the American Statistical Association, 58, 236–244, https://doi.org/10.1080/01621459.1963.10500845, 1963.

Whiteman, C. D., Bian, X., and Zhong, S.: Low-Level Jet Climatology from Enhanced Rawinsonde Observations at a Site in the Southern Great Plains, Journal of Applied Meteorology and Climatology, 1997.

Wilks, D. S.: Resampling Hypothesis Tests for Autocorrelated Fields, Journal of Climate, 10, 65 – 82, https://doi.org/10.1175/1520-0442(1997)010<0065:RHTFAF>2.0.CO;2, 1997.