



Cluster analysis of wind turbine wakes measured through a scanning Doppler wind LiDAR

Romit Maulik, Vishwas Rao and S. Ashwin Renganathan
Argonne National Laboratory, Lemont, IL-60439, USA

Stefano Letizia and Giacomo Valerio Iungo*
Wind Fluids and Experiments (WindFluX) Laboratory, Mechanical Engineering Department, The University of Texas at Dallas, 800 W Campbell Rd, Richardson, TX, 75080

Wind turbine wakes are responsible for power losses and added fatigue loads of wind turbines. Providing capabilities to predict accurately wind-turbine wakes for different atmospheric conditions and turbine settings with low computational requirements is crucial for the optimization of wind-farm layout, and for improving wind-turbine controls aiming to increase annual energy production (AEP) and reduce the levelized cost of energy (LCOE) for wind power plants. In this work, wake measurements collected with a scanning Doppler wind LiDAR for broad ranges of the atmospheric static stability regime and incoming wind speed are processed through K -means clustering. For computational feasibility, the cluster analysis is performed on a low-dimensional embedding of the collected data, which is obtained through proper orthogonal decomposition (POD). After data compression, we perform K -means of the POD modes to identify cluster centers and corresponding members from the LiDAR data. The different cluster centers allow us to visualize wake variability over ranges of atmospheric, wind, and turbine parameters. The results show that accurate mapping of the wake variability can be achieved with K -means clustering, which represents an initial step to develop data-driven wake models for accurate and low-computational-cost simulations of wind farms.

I. Introduction

Power generation through wind turbines encompasses an energy-transformation process from the wind kinetic energy into mechanical rotation of the moving parts of the turbines and, finally, transformation into electricity through the generator. The residual wind kinetic energy past a turbine rotor is typically characterized by higher turbulence intensity and requires a downstream fetch of about one order of magnitude larger than the rotor diameter to recover an energy potential comparable to that of the incoming wind field [1]. The near-wake flow features are mainly affected by the rotor aerodynamics and control settings, which can be characterized through the rotor thrust coefficient and incoming wind speed at hub height [2]. In contrast, the far-wake evolution is mainly governed by turbulent mixing and, in turn, the main wake variability is induced by the characteristics of the incoming wind, such as turbulence intensity, shear, and veer [3]. Being able to predict accurately wake variability for different incoming wind field and turbine settings to optimize wind farm layout and improve wind turbine controls is as important as to keep the computational costs of the used wind farm models very low considering the hundreds of configurations to be simulated for a wind power project [4].

In this work, measurements were performed with a scanning Doppler wind light detection and ranging (LiDAR) for wakes generated by utility-scale wind turbines deployed at an onshore wind farm. These are analyzed through cluster analysis to achieve an accurate and efficient mapping of the variability of wind turbine wakes for different atmospheric conditions, such as through the incoming turbulence intensity at hub height or the Bulk Richardson number [5], and operative conditions of the wind turbines, i.e. incoming wind speed at hub height. With this approach, we aim to provide a framework for the development of a new generation of data-driven models for predicting wind turbine wakes with low computational costs.

The remainder of the paper is organized as follows. The LiDAR experiment and the wind farm under investigation are described in Sect. II. Subsequently, the proper orthogonal decomposition (POD) analysis is presented in Sect. III, while the K -means cluster analysis is detailed in Sect. IV. Finally, concluding remarks are reported in Sect. V.

*Corresponding author; email:valerio.iungo@utdallas.edu

II. LiDAR measurements for an onshore wind farm

A LiDAR campaign was performed for an onshore wind farm located in North Texas consisting of 39 identical 2.3-MW wind turbines with rotor diameter, d , of 108 m, and a hub height of 80 m. The elevation map of the site is retrieved with a spatial resolution of 100 m from the U.S. Geological Survey [6]. By setting the offset altitude at the location of the LiDAR deployment, the standard deviation of the terrain is only 16 m, which allows considering this site as flat terrain. For the retrieval of the LiDAR data, the hub height of each turbine is corrected by taking the local altitude at the turbine locations into account.

The measurement campaign was conducted through various phases between August 2015 and March 2017 for a total of 236 days. Meteorological data ('met-data' hereafter) were provided as ten-minute averages and standard deviation of wind speed, wind direction, temperature, humidity, and barometric pressure, at heights of 36 m, 60 m, and 80 m. The atmospheric stability regime can be characterized through the Bulk Richardson number as follows [5]:

$$Ri_B(\bar{z}) = \frac{g \Delta T / \Delta z}{T(z_w) U^2(z_w)} \bar{z}^2, \quad (1)$$

where g is the gravitational acceleration, z_w is the met-tower height where the wind speed, U , is measured, ΔT is the temperature variation over $\Delta z = z_2 - z_1$ ($z_2=75$ m and $z_1=3$ m), and $\bar{z} = \sqrt{z_1 z_2}=15$ m.

Supervisory control and data acquisition (SCADA) data were provided for each turbine as ten-minute averages and standard deviation of wind speed, power output, rotor rotational velocity, and yaw angle. For more details on this dataset and the used quality-control process see El-Asha et al. [1], Zhan et al. [3].

The scanning Doppler wind LiDAR deployed for this experiment is a Windcube 200S manufactured by Leosphere, which emits a laser beam into the atmosphere and measures the radial wind speed, i.e. the velocity component parallel to the laser beam, from the Doppler frequency shift of the back-scattered light. The LiDAR system operates in a spherical coordinate system and measures the radial velocity defined as the summation of the three velocity components projected onto the laser beam direction. It features a typical scanning range of about 4 km with a range gate of 50 m, an accumulation time of 500 ms, and an accuracy of 0.5 m/s in wind speed measurements. More details on the LiDAR system and the field campaign are available in El-Asha et al. [1], Zhan et al. [3].

The LiDAR measurements were typically performed by using a range gate of 50 m, an elevation angle of $\phi = 3^\circ$, azimuthal range of 20° , a rotation speed of the scanning head of $2^\circ/\text{s}$, leading to a typical scanning time for a single planar-position indicator (PPI) scan of about 10 s [7, 8]. After rejecting LiDAR data with a carrier-to-noise ratio (CNR) lower than -25 dB, a proxy for the streamwise velocity is obtained from the LiDAR radial velocity, V_r , through the streamwise equivalent velocity $U_{eq} = V_r / [\cos\phi \cos(\theta - \theta_w)]$ [3], where θ is the azimuthal angle of the LiDAR laser beam and θ_w is the wind direction. The streamwise equivalent velocity is then made non-dimensional through the velocity profile in the vertical direction of the incoming boundary layer, which is also measured with the LiDAR. The reference frame used has x -direction aligned with the wake direction, which is estimated with the linear fitting of the wake centers at various downstream locations [3], y -direction in the horizontal transverse direction, and z -direction vertically and positive moving upwards.

A total number $N = 6,781$ of quality-controlled planar-position indicator (PPI) LiDAR scans of isolated wind turbine wakes have been processed to provide the non-dimensional wake velocity fields used for this study [3]. The wake velocity fields measured with the LiDAR are interpolated over the horizontal plane at hub height within the domain $1 \leq x/d \leq 7$ and $-1 \leq y/d \leq 1$ by leveraging the assumption of an axisymmetric wake when removing the distortion induced by the incoming boundary layer flow [3]. The wake velocity data have a spatial resolution of $0.1d$ and $0.05d$ in the x and y directions, respectively, generating for a single snapshot a data matrix of $[p \times q] = [61 \times 41]$ ($pq = 2,501$).

For grid points where the LiDAR data are not available, which can be due to the quality control process of the LiDAR data or a significant misalignment between the wake direction and the direction connecting the LiDAR and turbine locations, the velocity fields are interpolated through the `inpaint-nans` function available in Matlab [9]. The ensemble-averaged velocity fields calculated over the entire dataset are reported in Fig. 1 for both raw and interpolated data. The main data distortion due to the velocity interpolation occurs at the downstream corners of the spatial domain, which are the areas where the probability of missing LiDAR samples is higher. The interpolated velocity fields are only used for the POD, which does not allow for not-a-number (NaN) values over the spatial domain for the calculation of the eigenproblem of the velocity covariance matrix. In contrast, statistics of the wake velocity field are calculated with the original non-interpolated velocity fields by ignoring the NaN values.

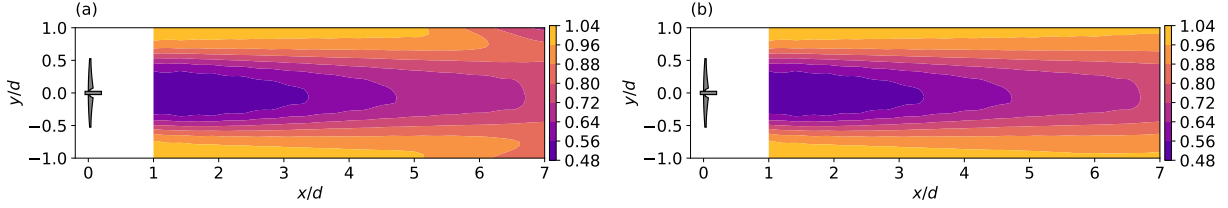


Fig. 1 Ensemble mean of the non-dimensional wake velocity fields calculated over the entire dataset: a) NaN values interpolated with the function `inpaint-nans`; b) interpolation-free data.

III. Proper Orthogonal Decomposition of the LiDAR dataset

Analyzing a large dataset of correlated variables to detect trends, data variability, and features can be very challenging and computationally prohibitive. Rather than analyzing the dataset in its entirety, it can be more convenient to represent the data on a lower-dimensional subspace with a suitable lower rank to enable a simplified analysis, which typically entails lower computational costs and a clearer interpretation of the results (see e.g. [10–12]). Specifically, proper orthogonal decomposition (POD) [13–16] allows generating an orthonormal basis, which is optimal for the reconstruction of the data variability. A snapshot of the wake velocity field measured with the LiDAR, u , can be represented through a linear combination of deterministic functions, which are referred to as POD modes ϕ_j :

$$u(x, y, z, t) = \sum_{j=1}^N \phi_j(x, y, z) a_j(t), \quad (2)$$

where t is time and N is the total number of snapshots. The POD modes, ϕ_j , represent the typical realizations in a statistical sense. In this study, the parameters $a_j(t)$ are coefficients that represent the amplitude of each mode and vary according to the collected snapshots. POD provides a modal decomposition that is completely a-posteriori and data-dependent, which does not neglect the non-linearities of the original dynamical system, even being a linear procedure. Furthermore, the POD basis is orthonormal and optimal in variance, i.e. among all linear decomposition techniques, it provides the most efficient detection, in a certain least-squares optimal sense, of the dominant components.

We introduce the POD procedure in the following. The LiDAR dataset, $\mathbf{U}$, with dimension $[pq \times N]$ can be approximated by computing the first r most energetic principal components through the singular value decomposition (SVD):

$$\mathbf{U} \approx \Phi \times \Sigma \times V^T, \quad (3)$$

where Φ ($[pq \times r]$) and V ($[N \times r]$) are orthonormal matrices and Σ is a diagonal $[r \times r]$ matrix with the first r singular values of $\mathbf{U}$ in descending order as diagonal entries [17]. Each diagonal element of Σ , σ_j , represents the energy contribution of the POD mode ϕ_j to the covariance matrix of the velocity snapshots. The POD modes are obtained as columns of the matrix Φ , while the principal components, a_j , associated with each POD mode are obtained by projecting the snapshot dataset onto the POD basis:

$$\mathbf{A} = \mathbf{U}^T \Phi, \quad (4)$$

where the principal components, a_j , are the columns of $\mathbf{A}$, whose size is $[N \times r]$.

We applied POD to the non-dimensional LiDAR wake measurements over the horizontal plane at hub height. The obtained eigenvalues of the POD modes, σ_j , and the respective cumulative energy reconstructed with an increasing number of POD modes are reported in Fig. 2. It is now crucial to select the smallest number of POD modes enabling an efficient reconstruction of the wind-turbine wake variability probed through the LiDAR measurements. The selection of POD modes can be done based on the energetic contribution of the various POD modes. In other words, by leveraging the energy optimality of the POD basis, the latter is truncated to reconstruct a certain percentage of the total energy of the velocity covariance matrix (see Fig. 2b). An alternative to this energetic approach has been used for the present work, which consists of visually inspecting the most energetic POD modes and selecting only POD modes whose spatial morphology indicates a clear physical feature.

The first 12 most energetic POD modes are reported in Fig. 3. POD mode 0 resembles the ensemble-average of the interpolated wake velocity fields, which is shown in Fig. 1a. Therefore, this POD mode has an evident physical

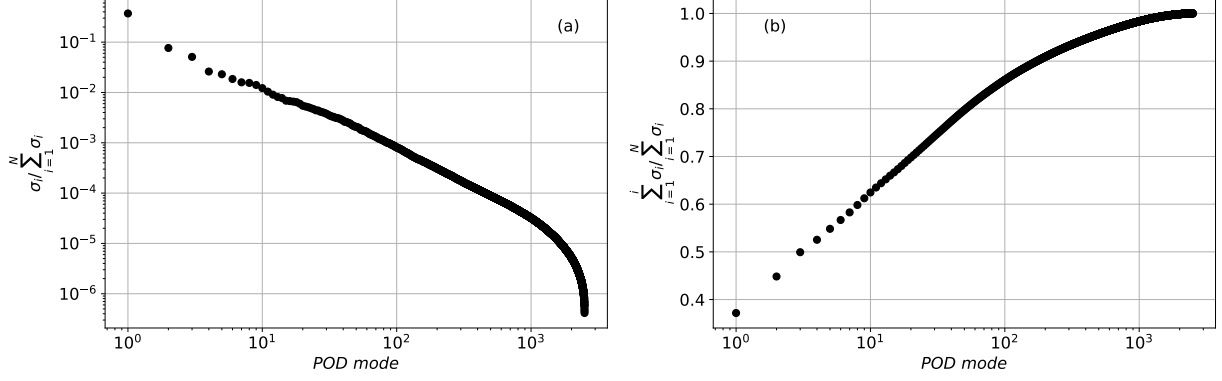


Fig. 2 Eigenvalues of the covariance matrix of the LiDAR snapshots, Σ : a) eigenvalues, σ_i ; b) cumulative POD energy.

contribution to the wake morphology. In contrast, POD modes 1 and 2 represent the corrections performed by the interpolation of the LiDAR data with the Matlab function `inpaint-nans` [9] and, thus, they only represent a numeric artifact rather than a physical feature to model. POD modes 3 and 4 seem to indicate a modulation in the transverse direction of the wake over the shape provided through the ensemble mean (POD mode 0). In other words, POD modes 3 and 4 can represent contractions or expansions of the wake in the transverse directions, and non-symmetric wake conditions that have been observed in the near wake for stable atmospheric conditions [2, 3]. Similarly POD modes 5 and 11 seem to indicate contractions or extensions of the wake in the streamwise direction. The remaining modes shown in Fig. 3 seem to indicate similar wake distortions, yet with slightly larger wavelengths, which can be considered as sub-harmonics of the above-mentioned wake modulations. Based on this qualitative analysis of the POD modes, which we understand is speculative rather than based on quantitative characteristics, the truncated POD basis selected for this work includes only POD modes 0, 3, 4, 5, and 11, which allows reconstructing 44.8% of the overall energy of the velocity covariance matrix.

Examples of the approximation performed on the LiDAR dataset through the truncated POD basis selected for this work are reported in Fig. 4. Two LiDAR snapshots collected during operations in Region 3 of the power curve (incoming wind speed normalized with turbine rated wind speed, $U_{hub}^* \geq 1$) and stable atmospheric condition, and Region 2 of the power curve ($U_{hub}^* < 1$) and convective atmospheric condition are reported in Figs. 4a and 4c, respectively. Experimental details of those measurements are reported in the caption of that figure. The corresponding POD-approximated wake velocity fields are reported in Figs. 4b and 4d, respectively. For both cases, a good approximation of the wake morphology is achieved, especially in the near-wake region, while noticeable differences are observed at the most downstream locations. In general, the wake morphology is smoothed throughout the spatial domain. This seems an affordable drawback of the POD data compression considering that the dimensionality of each snapshot is reduced from $pq = 2,501$ down to 5.

IV. K-means clustering

Clustering in general refers to a very broad set of techniques for finding subgroups in a dataset, which are referred to as clusters. The aim of clustering a dataset is to partition it into distinct sub-groups where samples sharing similar features belong to the same cluster and are segregated from those characterized by different features [18]. K -means clustering is a simple approach for partitioning a dataset into K distinct, non-overlapping clusters. To perform K -means clustering, the desired number of clusters, K , must be provided as input. Then, the algorithm assigns each observation to one of the clusters. The standard method used to perform K -means clustering is an iterative algorithm and it proceeds in two steps until convergence is achieved. Firstly, each observation is assigned a random integer (1 to K) as an initial cluster assignment. Next, for each of the K clusters, the centroid of the cluster is computed and each observation is assigned to the cluster that it is closest to. The second step is repeated until the assignment does not change anymore [17]. For our study, the inputs provided for the cluster analysis are the time-series of the principal components associated with the six POD modes selected to approximate the LiDAR dataset. The K -means outputs are the cluster centers,

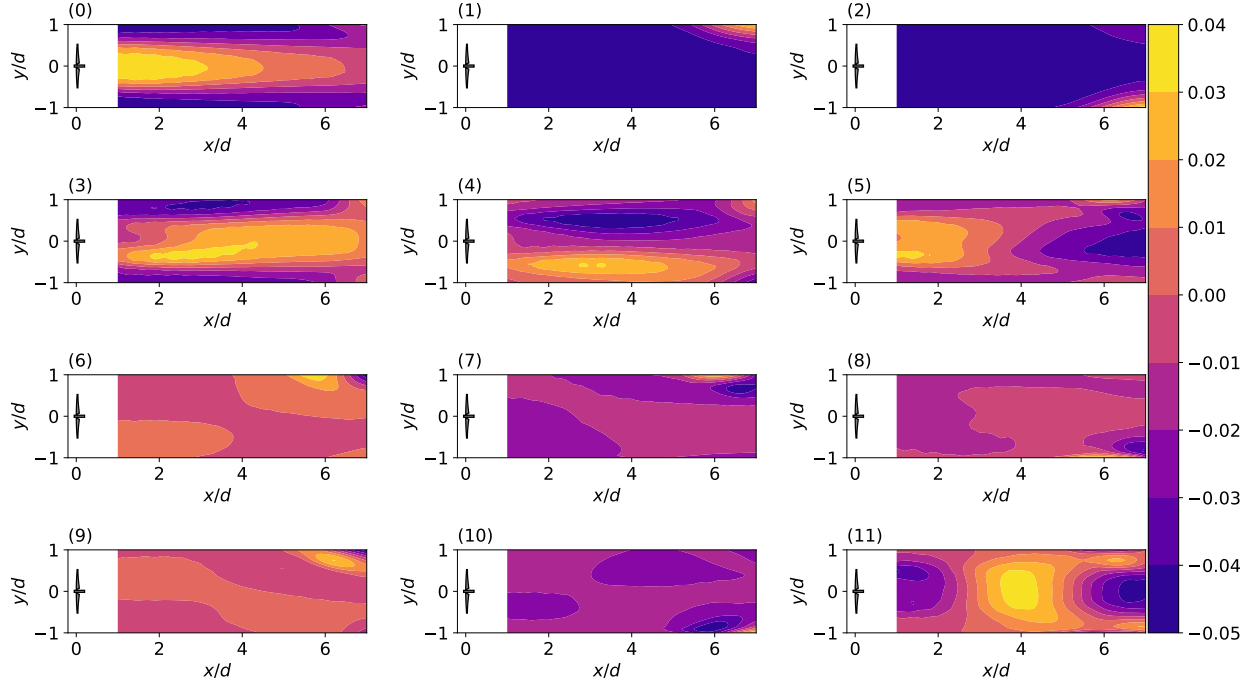


Fig. 3 First 12 POD modes obtained from the LiDAR dataset.

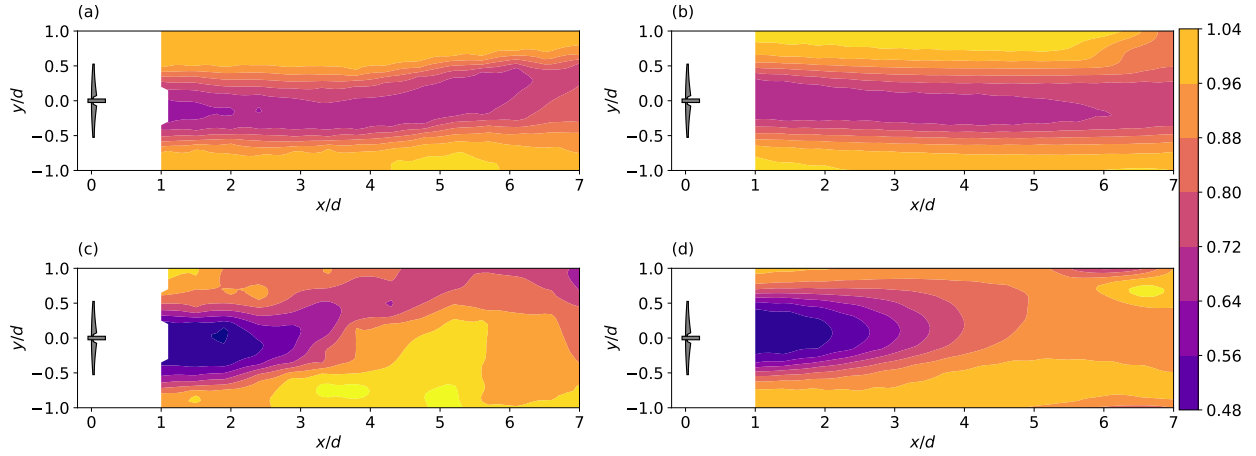


Fig. 4 Reconstruction of the LiDAR wake measurements through the selected truncated POD basis. The top row shows LIDAR data for $U_{hub}^*=1.08$, $TI=5.3\%$, $P_{norm}=1$, $Ri_B=0.005$ with the true (a) and POD-reconstructed (b) velocity fields. The bottom row shows LiDAR data for $U_{hub}^*=0.72$, $TI=15.4\%$, $P_{norm}=0.51$, $Ri_B=-0.003$ with the true (c) and POD-reconstructed (d) velocity fields.

namely the typical realization for each cluster, and labeling for each LiDAR snapshot to its respective cluster.

One of the challenges with using the K -means clustering algorithm is the choice of the number of clusters, K . This decision process is facilitated through the evaluation of the “inertia curve”, which shows the relevance of each cluster within the dataset. After a preliminary analysis, we selected a total of $K = 6$ clusters, whose inertia is reported in Fig. 5a. The generated clusters have a reduced relevance for the cluster analysis with an increasing number of clusters generated. However, the relevance of the clusters does not correspond to the occurrence of the data points within each cluster, as reported in Fig. 5b. It is observed that the cluster occurrence does not decrease with an increasing number of clusters, while the cluster with the largest occurrence (30.4%) is Cluster 2.

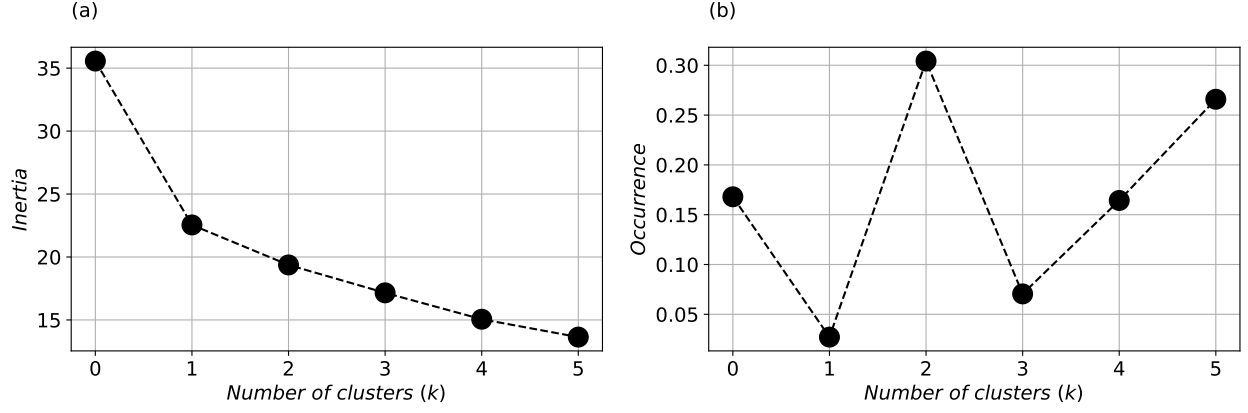


Fig. 5 Generation of the clusters: a) Inertia values of the clusters generated; b) Occurrence of the various clusters.

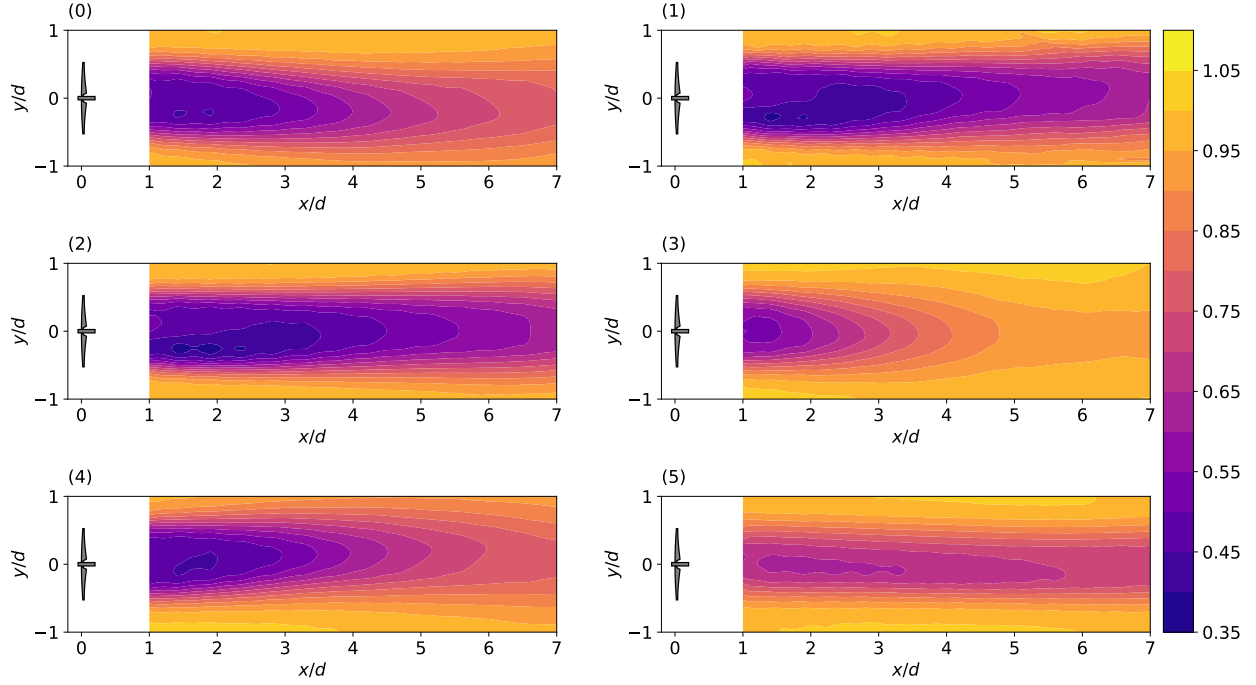


Fig. 6 Mean non-dimensional velocity fields associated with the various clusters.

For the sake of brevity, the cluster centers are not shown here, while the mean non-dimensional wake velocity fields calculated for each cluster are displayed in Fig. 6. This analysis shows that the clusters have been generated based on the intensity of the velocity deficit, recovery rate, and downstream extent of the wake. To provide a more quantitative analysis of the results obtained through the clustering procedure, for each cluster, the maximum velocity deficit, U_{min}^* , is estimated as a function of the downstream location, x , as reported in Fig. 7. Subsequently, U_{min}^* is fitted through the following power law for $x/d \geq 3$ [3, 19]:

$$U_{min}^* = 1 - A_U \left(\frac{x}{d} \right)^{-N_U}, \quad (5)$$

where A_U and N_U represent the velocity deficit at $x/d = 1$ and the wake-recovery rate, respectively. The values of A_U and N_U obtained by fitting the experimental data obtained for each cluster for $x/d \geq 3$ (Fig. 7) are reported in Table 1. First, we observe that the characteristics of Clusters 1 and 2 are very similar to each other. Analogously, Clusters 0 and 4 share similar flow characteristics. Cluster 3 has the largest values of the parameters A_U and N_U . This feature

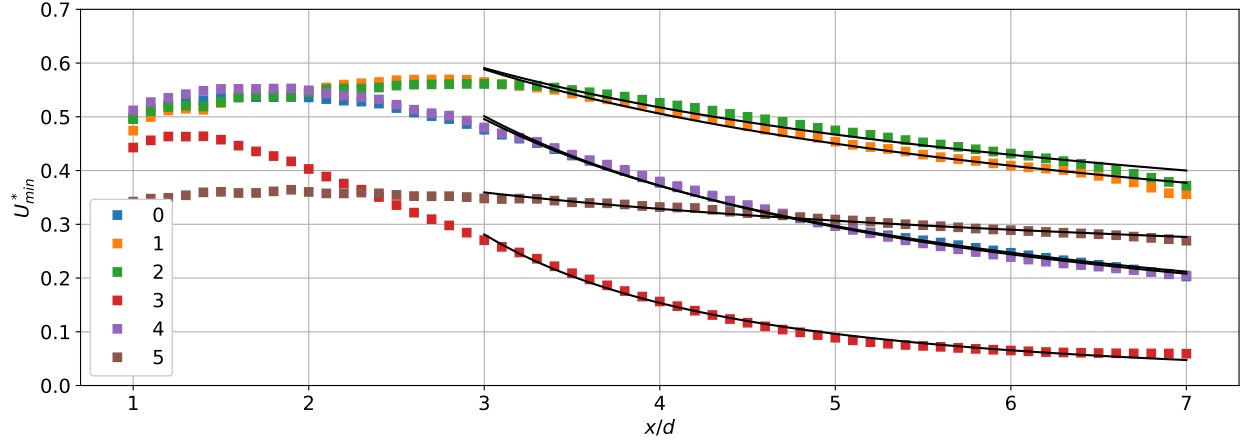


Fig. 7 Maximum velocity deficit, U_{min}^* , as a function of the downstream location, x , for the various clusters. Solid black lines represent their fitting with Eq. 5.

Cluster #	A_U	N_U
0	1.49	1.00
1	1.05	0.52
2	0.98	0.46
3	2.82	2.10
4	1.57	1.04
5	0.50	0.31

Table 1 Coefficients of Eq. 5 A_U and N_U fitted for the various clusters.

seems to be due to the short near-wake region and an anticipated wake recovery at around $x/d \approx 1.5$, as shown in Fig. 6(3). Indeed, besides Cluster 5, all the clusters have an initial velocity deficit at $x/d \approx 1$ of about $U_{min}^* \approx 0.5$. In contrast, Cluster 5 has the lowest initial velocity deficit ($A_U = 0.5$) and the lowest recovery rate of $N_U = 0.31$. Indeed, as can be observed in Fig. 6(5), the mean velocity field for this cluster entails a very shallow wake with a persistent downstream extent. Finally, Clusters 1 and 2 have a smaller value of $A_U \approx 1$ than for Clusters 0 and 4 ($A_U \approx 1.5$), and lower recovery rate ($N_U \approx 0.5$ and $N_U \approx 1$ for Clusters 1-2 and Clusters 0-4, respectively).

As mentioned above, the input provided for the K -means clustering analysis consists of the principal components selected for the truncated POD basis of the LiDAR measurements. Therefore, no input is provided about the incoming wind conditions or the settings of the turbine through the meteorological or SCADA data, respectively. We now investigate the effects of the cluster analysis on several meteorological and SCADA parameters through their histograms for the various clusters, which are reported in Fig. 8.

Starting from Cluster 5, we can state that the LiDAR measurements ascribed to Cluster 5 were collected for turbine operations in Region 3 of the power curve, namely for incoming wind speed larger than the rated wind speed of the turbine, which is equal to 11 m/s. Indeed, the histogram of the power for Cluster 5 has the largest occurrence for 2,300 W, which is the rated power of the turbines under investigation. Similarly, the pitch angle is typically larger than 0° , which indicates that the pitch controller of the turbine is activated to keep the generated power equal to the rated power by increasing the pitch angle for increasing incoming wind speed. These wind operations correspond to near-neutral atmospheric stability conditions, as highlighted by a Bulk Richardson number, Ri_B , close to zero for Cluster 5, and relatively low turbulence intensity with TI being typically between 5% and 12%. This cluster analysis justifies the relatively low-velocity deficit observed for the mean velocity field of Cluster 5 (Fig. 6(5)), which is due to a lower rotor thrust coefficient for blades with a higher pitch angle; indeed, $A_U = 0.5$ for Cluster 5. Furthermore, the relatively low TI of the incoming wind and smaller wake-generated turbulence associated with a reduced thrust coefficient, lead to a lower wake recovery rate ($N_U=0.31$).

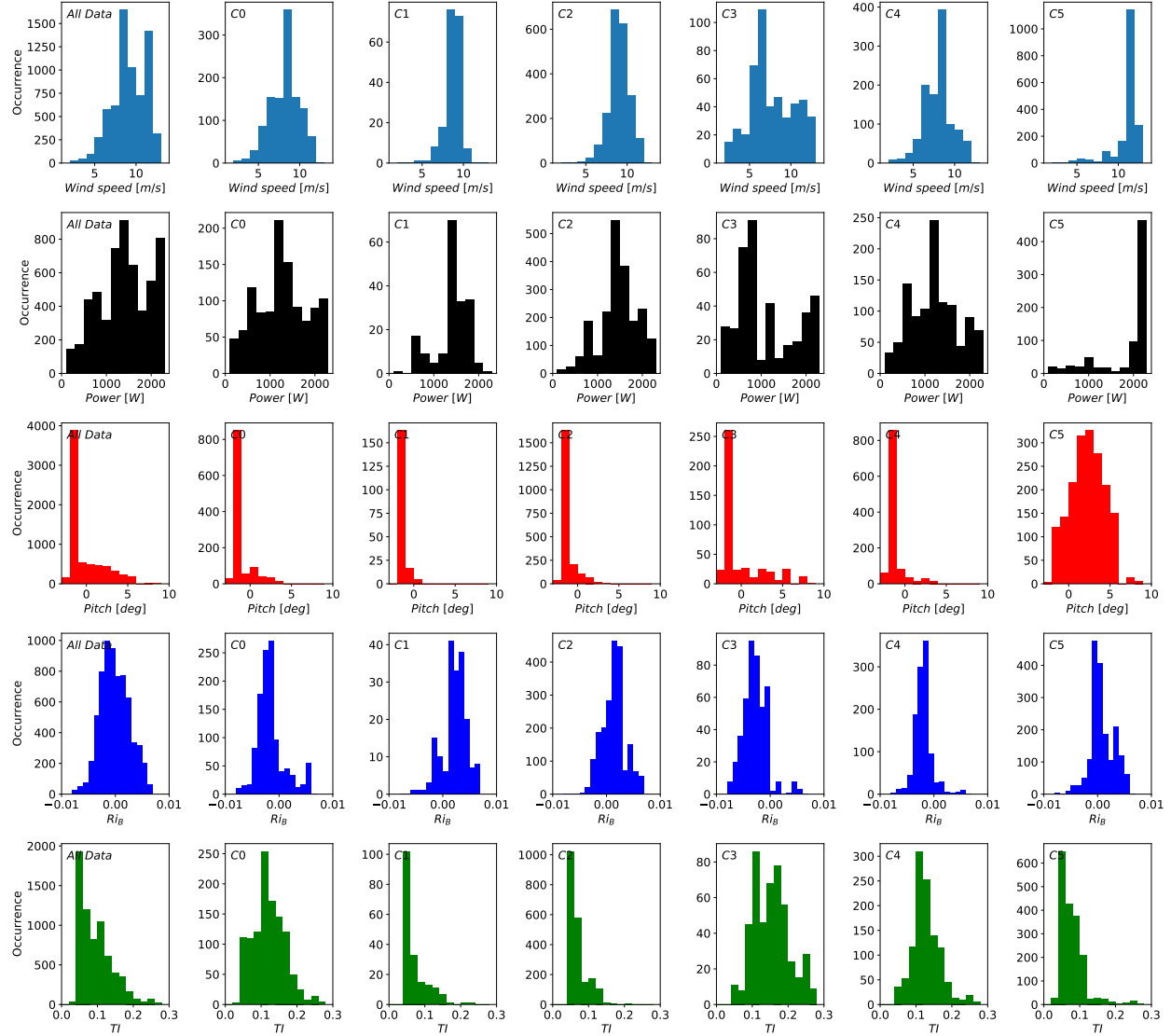


Fig. 8 Histograms of the hub-height wind speed, power, pitch angle, Bulk Richardson number, and hub-height turbulence intensity for the entire dataset and the various clusters generated.

The remaining 5 clusters are associated with operations in Region 2, namely with incoming wind speed at hub height lower than rated wind speed of 11 m/s, power capture lower than the rated power of 2,300 W, and typical pitch angle equal to the baseline position of -2° (Fig. 8). Clusters 1 and 2 are associated with operation under neutral-stable atmospheric conditions, indeed the Bulk Richardson number, Ri_B , is generally positive and TI generally lower than 15%. It is difficult to justify why the K -means analysis has generated two clusters for similar operative and atmospheric conditions. Indeed, the overall characteristics of the mean velocity field reported through Figs. 6, 7, and Table 1 are very similar for Clusters 1 and 2.

In contrast, Clusters 0 and 4 are related to operations in Region 2 of the power curve under neutral-convective conditions. Fig. 8 shows that the Bulk Richardson number, Ri_B , for Clusters 0 and 4 is typically negative, representing convective atmospheric conditions, and TI achieving larger values as high as 30%. The wake recovery for these clusters is faster than for Clusters 1 and 2, as confirmed by the larger values of the parameters A_U and N_U reported in Table 1.

Finally, Cluster 3 is associated again with operations under unstable conditions ($Ri_B < 0$). Furthermore, the levels of TI are even larger than for Clusters 0 and 4. Interestingly, the operations associated with Cluster 3 are characterized by a higher occurrence of wind conditions at the transition between region 2 and region 3, which is confirmed by the

higher occurrence of pitch angle values larger than 0° , which are typical for an activated pitch controller in Region 3 of the power curve. Therefore, this analysis suggests that Cluster 3 is associated with operations under convective conditions and for the so-called Region 2.5 of the power curve, namely when the pitch control starts operating for transitioning from Region 2 with maximum thrust coefficient and Region 3 with a fixed rated power. The combination of these operative conditions with highly turbulent convective wind conditions leads to the very narrow near-wake region and fast wake recovery observed through the LiDAR measurements.

V. Discussion and Concluding remarks

Wind LiDAR measurements collected in the wake of isolated wind turbines have been analyzed through K -means clustering to capture wake variability induced by different conditions of the incoming wind field and operative settings of the wind turbines.

First, the dimensionality of the LiDAR data has been reduced through a truncated basis generated with proper orthogonal decomposition (POD). Specifically, only five POD modes have been selected to efficiently reproduce the measured wake variability, while keeping the computational costs of the proposed procedure very low.

The K -means analysis has been performed by providing as input the principal components of the five selected POD modes. We generated six clusters from the LiDAR dataset, whose mean wake velocity fields differ in terms of wake velocity deficit, recovery rate, and downstream extent. It has been shown that one cluster is ascribed to turbine operations in Region 3 of the power curve, namely for incoming wind speed larger than rated wind speed and power capture equal to the rated power of the turbines of 2,300 W. For this cluster, the mean wake velocity field is characterized by a very shallow velocity deficit and slow wake recovery.

Two clusters have been ascribed to operations in Region 2 of the power curve under stable atmospheric conditions, while the other two clusters have been associated with operations in Region 2 of the power curve, but under convective atmospheric conditions. The latter are characterized by a faster recovery rate than for those connected with stable atmospheric conditions. Finally, the last cluster has been ascribed to transitional operations between Region 2 and Region 3 when the pitch controller kicks in under unstable conditions. The mean velocity field connected with this cluster is characterized by a very short near-wake region and a faster recovery rate.

Summarizing, it is noteworthy that by performing K -means clustering analysis of the only LiDAR data, namely without providing any information of the incoming wind field through the meteorological data and turbine operations through the SCADA data, it has been possible to reconstruct variability of the wake velocity field associated with utility-scale wind turbines. These results corroborate the concept that it is possible to develop fully data-driven wake models for which the variability of the wake velocity field will reflect effects induced by the turbine rotor and the characteristics of the incoming wind field.

Acknowledgments

This research has been funded by a grant from the National Science Foundation CBET Fluid Dynamics, award number 1705837. The Texas Advanced Computing Center is acknowledged for the computational resources. The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>).

References

- [1] El-Asha, S., Zhan, L., and Iungo, G. V., "Quantification of power losses due to wind turbine wake interactions through SCADA, meteorological and wind LiDAR data," *Wind Energy*, Vol. 20, No. 11, 2017, pp. 1823–1839.
- [2] Zhan, L., Letizia, S., and Iungo, G. V., "Optimal tuning of engineering wake models through LiDAR measurements," *Wind Energ. Sci.*, Vol. 5, 2020, pp. 1601–1622. <https://doi.org/https://doi.org/10.5194/wes-5-1601-2020>.
- [3] Zhan, L., Letizia, S., and Iungo, G. V., "LiDAR measurements for an onshore wind farm: Wake variability for different incoming wind speeds and atmospheric stability regimes," *Wind Energy*, Vol. 23, No. 3, 2020, pp. 501–527.

- [4] Santhanagopalan, V., Rotea, M. A., and Iungo, G. V., "Performance optimization of a wind turbine column for different incoming wind turbulence," *Renewable Energy*, Vol. 116, No. Part B, 2018, pp. 232–243.
- [5] Stull, R. B., *An introduction to boundary layer meteorology*, Vancouver, Canada: Springer Science, 1988.
- [6] "U.S. Geological Survey Website," <https://www.usgs.gov/>, 2020. Accessed: 2020-05-28.
- [7] Letizia, S., Zhan, L., and Iungo, G. V., "LiSBOA: LiDAR Statistical Barnes Objective Analysis for optimal design of LiDAR scans and retrieval of wind statistics. Part I: Theoretical framework," *Atmos. Meas. Tech. Discuss.*, 2020. <https://doi.org/https://doi.org/10.5194/amt-2020-227>, in review.
- [8] Letizia, S., Zhan, L., and Iungo, G. V., "LiDAR Statistical Barnes Objective Analysis for optimal design of LiDAR scans and retrieval of wind statistics. Part II: Applications to LiDAR measurements of wind turbine wakes," *Atmos. Meas. Tech. Discuss.*, 2020. <https://doi.org/https://doi.org/10.5194/amt-2020-228>, in review.
- [9] D'Errico, J., "Inpaint-nans," MATLAB Central File Exchange, 2004.
- [10] Iungo, G., and Lombardi, E., "Time-frequency analysis of the dynamics of different vorticity structures generated from a finite-length triangular prism," *J. Wind Eng. Ind. Aerodyn.*, Vol. 99, 2011, pp. 711–717. <https://doi.org/http://dx.doi.org/10.1016/j.jweia.2011.03.015>.
- [11] Iungo, G., and Lombardi, E., "A procedure based on proper orthogonal decomposition for time-frequency analysis of time series," *Exp. Fluids*, Vol. 51, 2011, pp. 969–985. <https://doi.org/http://dx.doi.org/10.1007/s00348-011-1123-1>.
- [12] Debnath, M., Santoni, C., Leonardi, S., and Iungo, G. V., "Towards reduced order modelling for predicting the dynamics of coherent vorticity structures within wind turbine wakes," *Phil. Trans. R. Soc. A*, Vol. 375, 2017, p. 20160108. <https://doi.org/http://dx.doi.org/10.1098/rsta.2016.0108>.
- [13] Lumley, J. L., *Stochastic tools in turbulence*, New York, NY: Academic Press, 1970.
- [14] Wold, S., Esbensen, K., and Geladi, P., "Principal component analysis," *Chemometrics and intelligent laboratory systems*, Vol. 2, No. 1-3, 1987, pp. 37–52.
- [15] Berkooz, G., Holmes, P., and Lumley, J. L., "The proper orthogonal decomposition in the analysis of turbulent flows," *Annual review of fluid mechanics*, Vol. 25, No. 1, 1993, pp. 539–575.
- [16] Holmes, P., Lumley, J. L., and Berkooz, G., *Turbulence, coherent structures, dynamical systems and symmetry*, Cambridge, UK: Cambridge University Press, 1996.
- [17] James, G., Witten, D., Hastie, T., and Tibshirani, R., *An introduction to statistical learning*, Springer, 2013.
- [18] Jain, A. K., Murty, M. N., and Flynn, P. J., "Data clustering: a review," *ACM computing surveys (CSUR)*, Vol. 31, No. 3, 1999, pp. 264–323.
- [19] Iungo, G. V., and Porté-Agel, F., "Volumetric lidar scanning of wind turbine wakes under convective and neutral atmospheric stability regimes," *J. Atmos. Ocean. Technol.*, Vol. 31, No. 10, 2014, pp. 2035–2048.