Taylor & Francis
Taylor & Francis Group

Cluster-based probabilistic structure dynamical model of wind turbine wake

Naseem Ali, Marc Calaf & Raúl Bayoán Cal

Published online: 19 May 2021.

Submit your article to this journal 🖉

View related articles 🖸

View Crossmark data 🖸

Taylor & Francis
Taylor & Francis Group

Check for updates

Cluster-based probabilistic structure dynamical model of wind turbine wake

Naseem Ali [a], Marc Calaf[b] and Raúl Bayoán Cal [a]

[a]Department of Mechanical and Materials Engineering, Portland State University, Portland, OR, USA;
[b]Department of Mechanical Engineering, University of Utah, Salt Lake City, UT, USA

ABSTRACT

For complex flow systems like the one of the wind turbine wakes, which include a range of interacting turbulent scales, there is the potential to reduce the high dimensionality of the problem to low-rank approximations. Unsupervised cluster analysis based on the proper orthogonal decomposition is used here to identify the coherent structure and transition dynamics of wind turbine wake. Through the clustering approach, the nonlinear dynamics of the turbine wake is presented in a linear framework. The features of the fluctuating velocity are grouped based on similarity and presented as the centroids of the defining clusters. Determined from probability distribution of the transition, the dynamical system identifies the features of the wakes and the inherent dynamics of the flow.

1. Introduction

Control models are crucial for performance optimisation of the wind farms [1,2]. Therefore, wind energy designers require cultivated models that efficiently imitate the flow dynamics [3–7]. Numerous mathematical models have been developed to try to capture the wake behaviour. These models for example include the steady-state models such as Gaussian deficit model and the Park model [8,9], and wake dynamic models such as dynamic wake meandering (DWM) [10]. However, the results are model and method dependent [11,12]. Alternatively, higher fidelity simulations can be used [13,14]. Although these models are more accurate to represent the flow dynamics, they are also computationally very expensive, and they are not useful for real-time control.

Wind energy designers require sophisticated models that accurately simulate and account of the complicated interactions of the turbines with the atmospheric boundary layer to obtain best use of wind energy resources [5,15–18]. To optimise the performance of wind energy resources, control and optimisation strategies are crucial. Model reduction has recently welcomed more concentration in the matter of investigation and development [19–21], where different types of reduced-order models (ROMs) are utilised such as the Petrov–Galerkin projection and data-driven ROM [22]. A model reduction framework consists of lowering the degree of freedom that defines a dynamical system to be able to

---

represent it through simplified models. Most model reduction methods are manipulated by the projection in the high-fidelity space as shown in the Petrov–Galerkin framework [22]. Proper orthogonal decomposition (POD) is an example of such an approach, which is commonly used to model complex fluid flows [23]. Based on the energy norm, POD is the optimal Galerkin expansion that minimises the averaged residual of energy. Further, the projection of the Navier–Stokes equation onto the POD modes defines the evolution equation. Different versions of the POD model are suggested in ROM, such as weighted POD [24] and adaptive POD [25]. To control a linear system, balanced POD is introduced based on the product of controllability and observability Gramians [23]. Another approach used to control flow instability is through the projection of linearised governing equations onto the global and adjoint global modes [26–29]. The POD approach is used for analysing the wake of the wind turbines and presenting a reduced-order model, see Hamilton et al. [30].

Based on a different framework, Kaiser et al. [28] introduced a cluster-based reduced-order modelling (CROM) approach for the study of a mixing layer. The CROM combines a cluster analysis and a Markov model, where the dynamical model depends on modelling the probability of the transition function between the obtained clusters with a Markov chain. Hence, in this approach, physical mechanisms embedded in the flow field can be clarified by highlighting transitions among clusters. In general, cluster analysis considers the integral part of machine learning and artificial intelligence methods that can be trained automatically from data. The clustering algorithm organises the snapshots to ensure that the inner-cluster similarity is maximised, while the inter-cluster similarity is minimised. The CROM technique demonstrates the power to specify the physical mechanisms of flow dynamics and therefore is applied for flow control [31,32]. The CROM has also been used to model the wake stabilisation mechanism downstream twisted and standard cylinders, predict the flame transition in gas turbine combustors, and model two-phase flow [33–35].

Liu et al. [36] compromised approach to cluster the turbines in the wind farm using K-means, a self-organising map, and spectral clustering (SC). The wind turbines were clustered into several groups by identifying similar characteristics of wind speed and output power. Thereafter, they introduced forecasting models based on correlation analysis. Wang et al. [37] introduced a deep belief network (DBN) model for wind power forecasting. The numerical weather prediction (NWP) data was used as the input of the proposed model and the k-means clustering algorithm was joined to cluster the NWP data. Wu and Peng [38] proposed neural network (NN) model with K-means clustering for short-term wind power forecasting. K-means clustering is applied to classify the samples into several groups, which contain the information of historical power data and meteorological conditions.

For wind farm controller analysis, high fidelity models are required for exploring the optimal possibilities of wind farm control. Due to the computational cost of these models, they are not convenient for online control. Therefore, new generations of models are required to solve this deficiency. Reduced-order models can provide coherent features of wake dynamics and solve part of the computational complexity, although these models are only suitable for specific conditions and are not proven for real cases. Linking between reduced-order model and parameter-varying control can provide new generation of the optimised wind farm control [39,40]. The wind plant wake flow studies generate massive amounts of data experimentally, numerically and field data. Different parameters have been independently studied such as the spacing, wind direction, orientation, to show the effect of

these on the power produced. Machine learning techniques offer a sophisticated approach to investigate these parameters together give a picture about the explicit interaction and hidden correlation between these parameters, and their effect on the power produced.

The current work seeks to describe the wake of the wind turbine and specify the dynamics essential to develop a reduced-order model. Unsupervised learning and Markov chain is used to introduce a low-order probabilistic model. The work is organised as follows: Section 2 outlines the proper orthogonal decomposition and clustering; Section 3 illustrates in particular large-eddy simulations; Section 4 provides a discussion of the finding and the relevance in the field. Conclusion is presented in Section 5.

2. Theory

The main part of CROM is a k-means clustering algorithm that furnishes the basis parts for reduced-order models (ROMs) [28,29,41–43]. The k-means algorithm minimises the total variations between the cluster centroid (the average of the data within each group) and all data contained in the cluster. The cluster analysis is achieved based on the time coefficients of the proper orthogonal decomposition. The second part of the CROM is cluster probability. In each cluster, the cluster probability is equivalent to the normalised number of ensembles included in each group [28]. A dynamical model is developed from the probability of transition and the Markov chain. The transition explains the number of individual elements of the maximum likelihood estimator.

2.1. Snapshot proper orthogonal decomposition

Proper orthogonal decomposition provides a reduced order modelling approach, based on an optimised projection [44–47],

$$\min \quad \frac{1}{N+1} \sum_{l=0}^{N} \left\| u(\hat{x}, t_l) - \sum_{j=1}^{r} \left(u(\hat{x}, t_l), \phi^j(\hat{x}) \right) \phi^j \right\|^2 \tag{1}$$

$$\text{subject to} \quad (\phi^j, \phi^i) = \delta_{ij} \quad \& \quad 1 \le i, \quad j \le r.$$

In here, u is the velocity as a function of spatial coordinates, $\hat{x}$, and time, t, and ϕ is the basis function. Also, r presents the low-rank approximation. For L^2 space, the flow field can be represented as

$$u(\hat{x}, t) = \sum_{j=1}^{N} a^j(t) \, \phi^j(\hat{x}). \tag{2}$$

The largest projection can be determined using the two point correlation tensor and Fredholm integral equation. The L^2 is a measure space, where the distance between two functions is minimum. Here the two functions are the velocity field and the projection modes. The basis functions, $\phi^j(\hat{x})$, are the eigenfunctions of the correlation tensor between two points, and are handed by

$$\phi^j(\hat{x}) = \frac{1}{\sqrt{\gamma_j}} \sum_{l=0}^{N} (G_j)_l u(\hat{x}, t_l), \tag{3}$$

where G is the eigenvector and γ is the eigenvalue. The POD eigenvectors illustrate the spatial structure of the turbulent flow and the eigenvalues measure the energy associated with corresponding eigenvectors. The summation of the eigenvalues presents the total turbulent kinetic energy in the flow domain. The mode coefficients that carry the dynamic details of the flow, $a^j(t)$, can be determined by the back projection of the POD modes on the flow velocity as

$$a^j(t) = \int_\Omega u(\hat{x}, t)\phi^j(\hat{x})\, dx. \tag{4}$$

2.2. Clustering theory

The main part of CROM is a k-means clustering algorithm that furnishes the basis functions for low-order models [28,29,41,42]. The observations are connected as

$$T_j^m := \begin{cases} 1 & \text{if } u^m \in c_j, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

The number of snapshots included in each cluster n_j is given by

$$n_j := \sum_{m=1}^{M} T_j^m. \tag{6}$$

and accordingly the total number of snapshots can be found as

$$N = \sum_{j=1}^{K} n_j = \sum_{k=1}^{K} \sum_{m=1}^{N} T_k^m. \tag{7}$$

Further, the cluster centroid, C_k, is defined as

$$C_k = \frac{1}{n_j} \sum_{u^m \in C} u^m = \frac{1}{n_j} \sum_{m=1}^{M} T_j^m u^m. \tag{8}$$

The total cluster variance is defined as

$$J(c_1, \ldots, c_K) = \sum_{j=1}^{K} \sum_{m=1}^{N} T_j^m ||u^m - C_j||_\Omega^2, \tag{9}$$

and the minimisation algorithm finds the optimal locations of the centroids,

$$c_1^{opt}, \ldots, c_K^{opt} = \arg\min_{c_1,\ldots,c_K} J(c_1, \ldots, c_K). \tag{10}$$

The cluster analysis is achieved regarding the POD coefficients, and consequently, the distance between two snapshots is determined as

$$S_{mn} := ||u^m - u^n||_\Omega^2 = ||a^m - a^n||. \tag{11}$$

The initial cluster centroid position is specified via k-means^{++} [48]. The cost function J introduced in Equation (9) converges monotonically [49].

In each cluster, the cluster probability q is identical to the normalised number of snapshots included in the cluster as [28]

$$q_j = \frac{n_j}{N} = \frac{1}{N} \sum_{m=1}^{N} T_j^m.$$
(12)

A dynamical model is generated from the cluster transition matrix (CTM), and the Markov chain as

$$P_{jk} = \frac{n_{jk}}{n_k},$$
(13)

where P_{jk} is the probability of moving from cluster k to j. Also, the n_{jk} presents the number of snapshots that are moved to the new clusters (from C_k to C_j), and is determined by

$$n_{jk} = \sum_{m=1}^{N-1} T_k^m T_j^{m+1}.$$
(14)

The transition presents the number of unique elements (cardinality) of maximum likelihood estimator. The evolution of the probability vector is defined by the iteration formula as

$$p^{l+1} = P p^l, \quad l = 0, 1, 2, \ldots$$
(15)

The initial probability distribution is allocated as p^0, and the successive distributions are defined compactly as

$$p^l = P^l p^0,$$
(16)

where P is the stochastic matrix (propagator). Asymptotically, the probability distribution can be determined as

$$p^\infty = \lim_{l \to \infty} P^l p^0.$$
(17)

Based on the convergence of P^l to stationary matrix (p^∞), the flow system will be probabilistically reproducible if the flow system is ergodic.

3. Numerical simulations

The incompressible form of Navier–Stokes equations, along with the divergence-free vector field, are used to simulate the flow,

$$\frac{\partial \tilde{u}_i}{\partial t} + \tilde{u}_j \left(\frac{\partial \tilde{u}_i}{\partial x_j} - \frac{\partial \tilde{u}_j}{\partial x_i} \right) = -\frac{1}{\rho} \frac{\partial \hat{p}^*}{\partial x_i} - \frac{\partial \tilde{\tau}_{ij}}{\partial x_j} + f_i,$$
(18)

$$\frac{\partial \tilde{u}_i}{\partial x_i} = 0,$$
(19)

where a tilde presents resolved instantaneous quantity, and $\tilde{\tau}_{ij}$ is the deviatoric part of the momentum subgrid stress term. The $\hat{p}^*$ is the filtered pressure term. Fourier discretisation with periodic boundary condition is applied in the horizontal directions. The LES analysed herein models a pseudospectral strategy where a staggered-grid is used. The top

boundary has zero-flux and zero-shear boundary conditions, The vertical velocity and the gradients of the streamwise and spanwise velocities are equal to zero. The lower surface has a no-slip condition, and a comparable surface shear stress is applied at the first grid point. The horizontal velocities are parameterised from the traditional log-law. A forcing term f_i exhibits the momentum sink caused by the wind turbine that includes both axial and tangential components. In the simulation, f_i is modelled using the actuator-disc with rotation following [50,51]. The LES analysed here imposes a single wind turbine working in the atmospheric boundary layer. Boundary effects, resulting from a finite computational domain and periodic boundary conditions, are alleviated by securing that the domain is adequately large. This provides for the wake of the wind turbine to fully recover before arriving at the outlet. The uniform numerical grid of the full-domain is $512 \times 128 \times 384$. A sub-domain of $64 \times 32 \times 32$ points is selected to achieve the current analysis. The spatial resolution between the grid is $\Delta x = 24.5$ m, $\Delta y = 24.5$ m, and $\Delta z = 7.8$ m. The $L_x = 12.3$, $L_y = 1$, and $L_z = 1.5$ km. The turbine diameter and hub height are identical and equal to 100 m. More details about the simulation can be found in Hamilton et al. [30]. The inflow mean velocity at the hub height is approximately 14.4 m/s. The simulation time covers 40 sweeping time of a fluid particle to travel a distance of 15 D. The surface roughness and hub-height turbulent intensity are 0.0003 m, and 6%, respectively. The number of snapshots used in the current analysis is 2000. The time resolution of the analysed data is $\Delta t = 2$ s.

4. Results

Figure 1 presents the first six of the POD eigenvectors, where they show the spatial coherent feature of the wake, and depict the low-rank miniature that contains the energetic features in the wake. Figure 2 presents the time coefficient associated with the first six POD modes. These coefficients are the main element to derive a time-dependent reduced order model based on dynamical systems theory.

The CROM algorithm is implemented to the POD coefficients classified into specific combinations to ensure the dissimilarity among the time coefficients inside each cluster is minimised and the dissimilarity among the groups is maximised. The dissimilarity between clusters is determined from Equation (11). The clusters are linked using an agglomerative hierarchical clustering algorithm [52]. Figure 3 presents the dissimilarity dendrogram tree with a granularity of 30 clusters. The y-axis (S) presents the distance between the clusters (dissimilarity). Also, the x-axis presents the number of clusters. The algorithm repeats the computations with the new number of clusters until the POD coefficients become hierarchically merged into a single cluster [53]. According to a certain threshold, the tree can be cut at distinctive heights, and the cluster number is recognised. The seven clusters display the most dissimilarity in the dendrogram tree.

The optimal number of cluster is critical in all unsupervised clustering algorithm. The optimal number should ensure that all states of interest are captured in the discretised phase space. Thus the low order mechanism analysis searches for the lower bound that properly defines for this flow the wake and recovered states. Figure 4 shows the comparison in the transition matrix among 3, 7 and 12 clusters. The transition matrix shows clear groups with three clusters, although, the flow is more complex than assigning by the one group of three clusters. The transition with $K = 3$ is also shown in the assignment of $K = 7$,
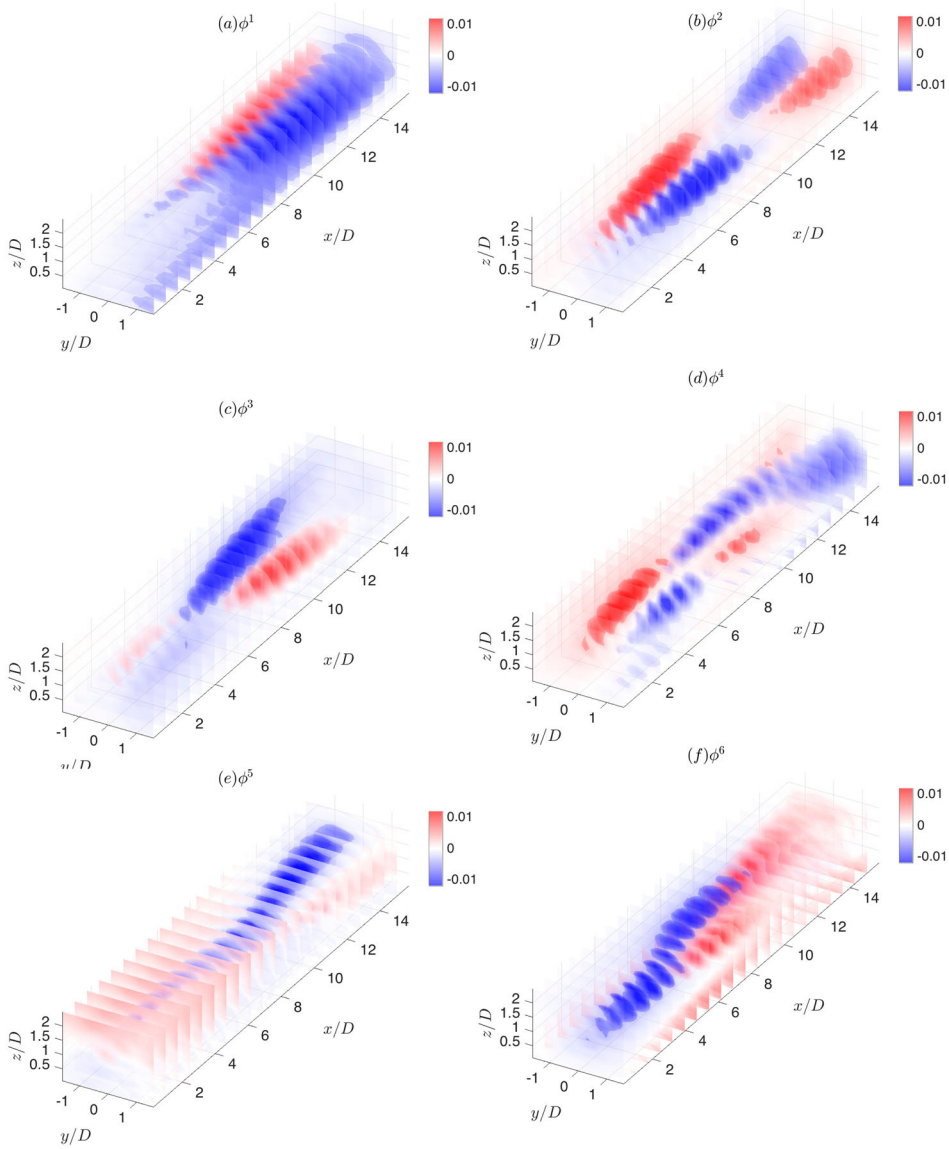
Figure 1. The eigenvectors of the proper orthogonal decomposition results.

with considering the index of the classification is different with varying the number of clusters. By using this criterion, the smallest cluster number needs to display the states of interest with a lower uncertainty in the structure of the transition matrix. With increasing the number of clusters, the transition becomes loss the clear group structure.

After appointing the number of clusters, the figurative trajectories of each cluster are shown. Figure 5 presents the cluster assignment of POD time coefficient vectors. The attractors are a set of states (features) that describes the evolution of the dynamical system. The changes in the distribution and transition from cluster to cluster are driven by
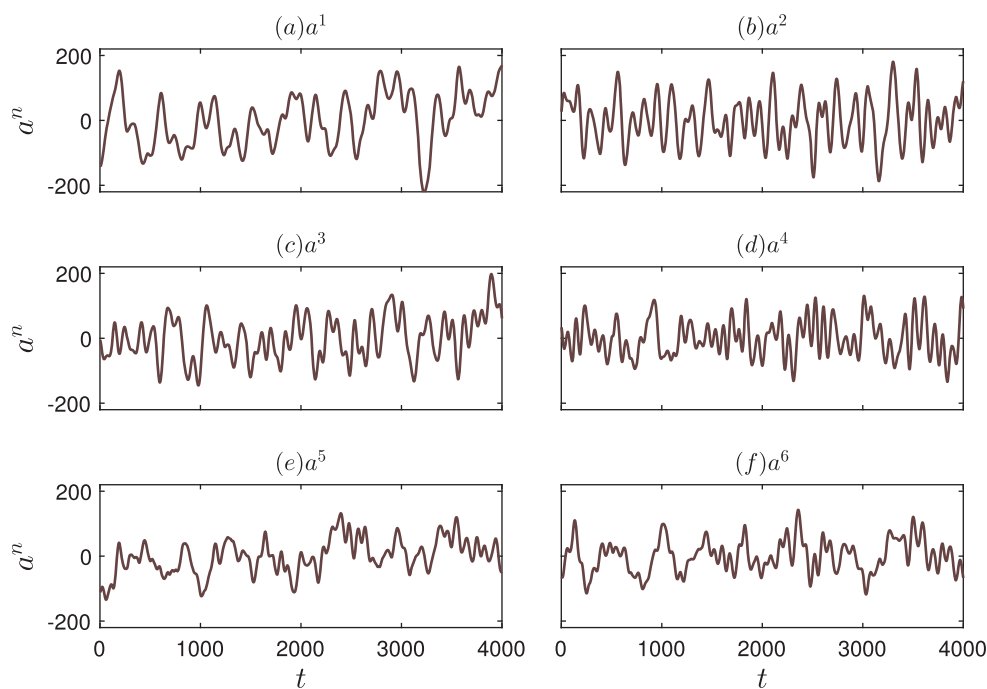
**Figure 2.** The time evolution of the POD coefficients.
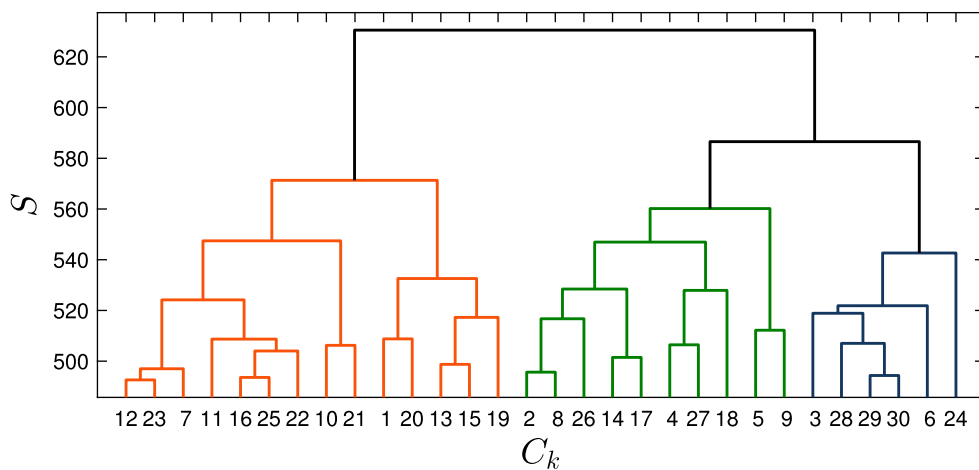
Figure 3. Dendrogram obtained through a hierarchical clustering algorithm applied on the time coefficient of POD projection.

the dynamics of the flow structure and they are controlled by the nature of the flow features of the wake and the exchange with the outer flow downstream the rotor. For example, the transition from cluster C_1 is directly moved to the groups of clusters that include C_4, C_5 and C_6 at specific times. Meaning that these clusters are connected and that the features of each cluster are in direct relation and interaction with the features of other clusters. Periodic transition is connected with cluster C_6 and cluster C_1. Cluster C_5 is minor in the time
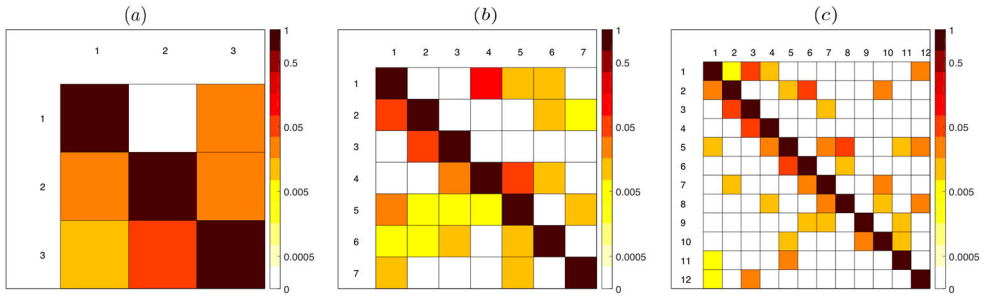
Figure 4. Transition matrices for $K = 3, 7$ and 12.
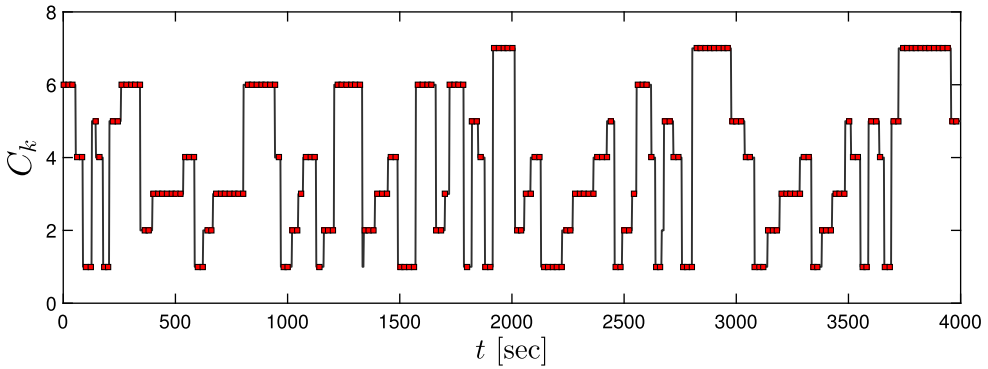
Figure 5. Time series of the cluster assignment of data.

assignment. Cluster C_7 becomes dominant at the end of the simulation time. Indicating that these clusters are more representative of the rare feature in the flow. The cluster assignment is the basis to introduce the flow kinematics through the fluctuating velocities.

The clustering labels of POD time coefficients are matched with the fluctuating velocity indices to find the centroids, see Figure 6. These centroids are obtained from Equation (8). The large-rotating structures in which an asymmetric wake are observed in these centroids. The rolling structures draw large-scale features down via the entrainment process [54–56]. That implies flow mixing creates an enhanced coherence in the wake region. The dipole-like structures that are assertive are shown, for example, in centroids C_1, C_2 and C_3. Fourier-like behaviour is dominant in the flow and is consistent with the rotation of the rotor. The structure depicted connects to the turbulent scales in the flow, such as turning features, which may extend to the far wake region or the structures that are distinct in the near-wake area. The structures at hub height are printed near the ground, see C_1, C_2 and C_4 in Figure 7. The clusters also highlight streaks (low-momentum structure near the surface [57]) at the surface and deviate towards the streamwise direction. In the POD, mode pairing is commonly shown in the time-resolved data, however, clustering approaches ensure under rotations that the statistical result is invariant. Therefore, no mode pairing is shown here in the current analysis. Very large structures align with the streamwise direction and cover the length of the physical domain. Due to the flow rotation, the sign of the structures is altered from one cluster to another. The seven centroids show wake accumulation
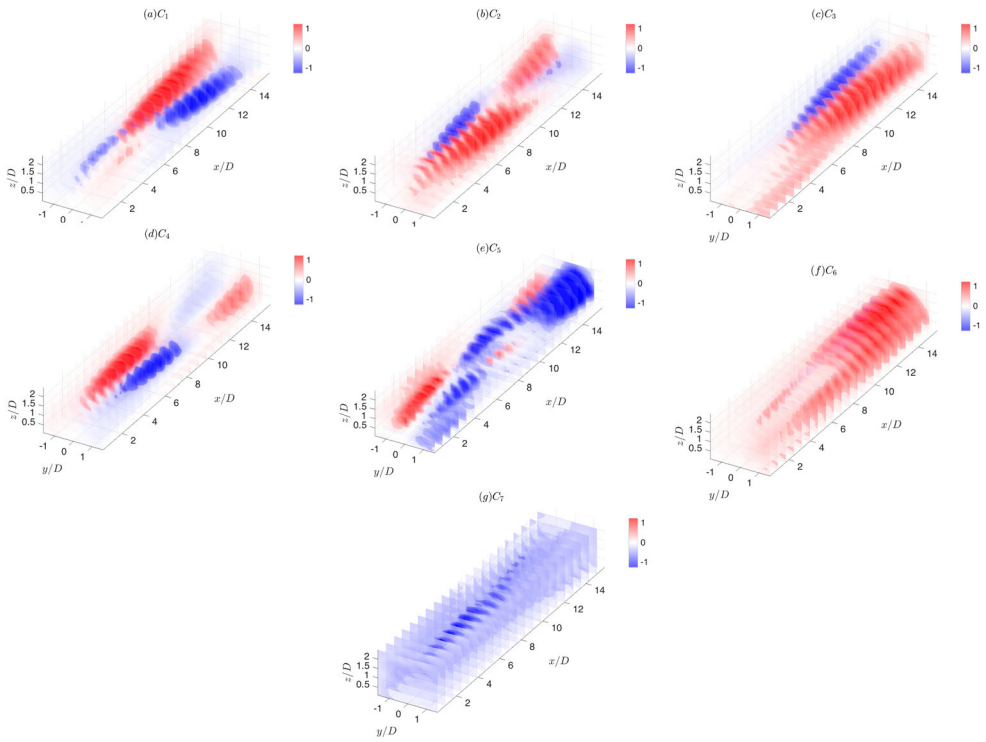
Figure 6. The kinematic description of the flow filed via the centroid of clusters downstream the turbine rotor disk. The positive and negative colourmap present positive and negative streamwise fluctuating velocities, respectively.

in the streamwise and wall-normal directions. Also, the change in the sign of the structure through the domain in each cluster highlights the flow complexity, unsteady vortices created in indifferent orientations and sizes. The kinematic description of the wake also reveals that there is a small rotating structure that grows in size downstream the rotor as shown by cluster C_1 and C_6 due to the flow recovery. The region between $9 \leq x/D \leq 12$ is very sensitive location, where it changes the orientation and advection of the flow structures, as shown in clusters C_1, C_2, and C_4. At this location, the flow is in the transition period and thereafter the flow recovers. The centroids C_3 and C_4 are similar to the first and second POD modes. This is due to that the POD modes or centroids are linear combinations of the original snapshots.

To quantify the dynamical system, the cluster transition matrix is shown in Figure 8. The horizontal plane of centroid features at the hub height is shown in the figure to provide a view of the transition. Also, a direct graph of the Markov model is pictured in subfigure(b). The flow is specified via the CTM, and the flow activity is emphasised from the probability outlook. The colours mark the transition matrix to differentiate the probability of transition from one cluster to other. The colour bar of subfigure(a) gives to the probability magnitude of the transition. Also, the colour bar of the contours presents the fluctuating velocity of the centroid. The CTM has non-zero elements in the principal and sub-diagonals, highlighting that the observation points can persist in the same cluster or reposition to the directed route
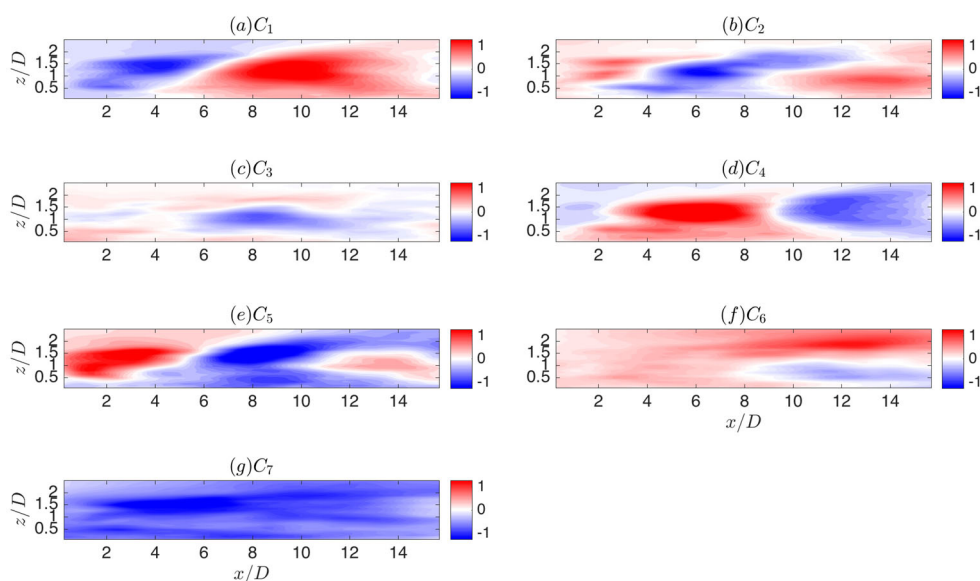
Figure 7. The kinematic description of the flow filed via the centroid of clusters downstream the turbine rotor disk. The positive and negative colourmap present positive and negative streamwise fluctuating velocities, respectively.

of the next cluster. More, the higher likelihoods in the principal diagonal than those in the sub-diagonal indicate that lingering is more recurring than transitions. This means that the probability of transitioning into the same cluster is dominant in this flow. The summation of the transition probability from one cluster to other is unity. A Markov model classifies observations in distinctive features. It is also quite evident that these observations are created under the impact of the rotor rotation and advected by the wake downstream the rotor. The impact of the rotor rotation is shown in the structure, where for example cluster one the structures are altered in the sign in the cluster 2. The transition states show that there is natural grouping between the clusters. For example, there is a group that shows the transition between C_1, C_4, C_3 and C_2. These cluster subsets display a large number of transitions within them than with the other subsets. Examining the CTM helps in understanding the flow behaviour and feature interactions. For example, the flow can call cluster C_4 exclusively from clusters C_1 and C_5, indicating that the flow can transition to the long feature at the transition period and beyond through these clusters. The prediction regarding structures can be advected or chopped into small features are observed. A large chance for the large structure shown in cluster C_3 to be split than advected as shown in the probability of the transition between C_3 and C_2. The variety of the clusters shows different types of states. For example, clusters C_4 and C_5 are the communicating states, as $C_4 \longleftrightarrow C_5$. The same kind of communicating state is observed between $C_7 \longleftrightarrow C_5$, and $C_1 \longleftrightarrow C_5$. The cluster transits as $C_3 \longrightarrow C_2 \longrightarrow C_7$. Additionally, C_3 only transitions to C_2, but C_5 is more likely to transition to C_1 (and C_6) than to C_7. There is a walk from cluster C_3 to C_7 passing through cluster C_2. More, all clusters can be attained from others, implying that no cluster regards in a class of its own. Also, there is the possibility of moving to a cluster from which there can be return. However, cluster C_3 shows less probability to return and
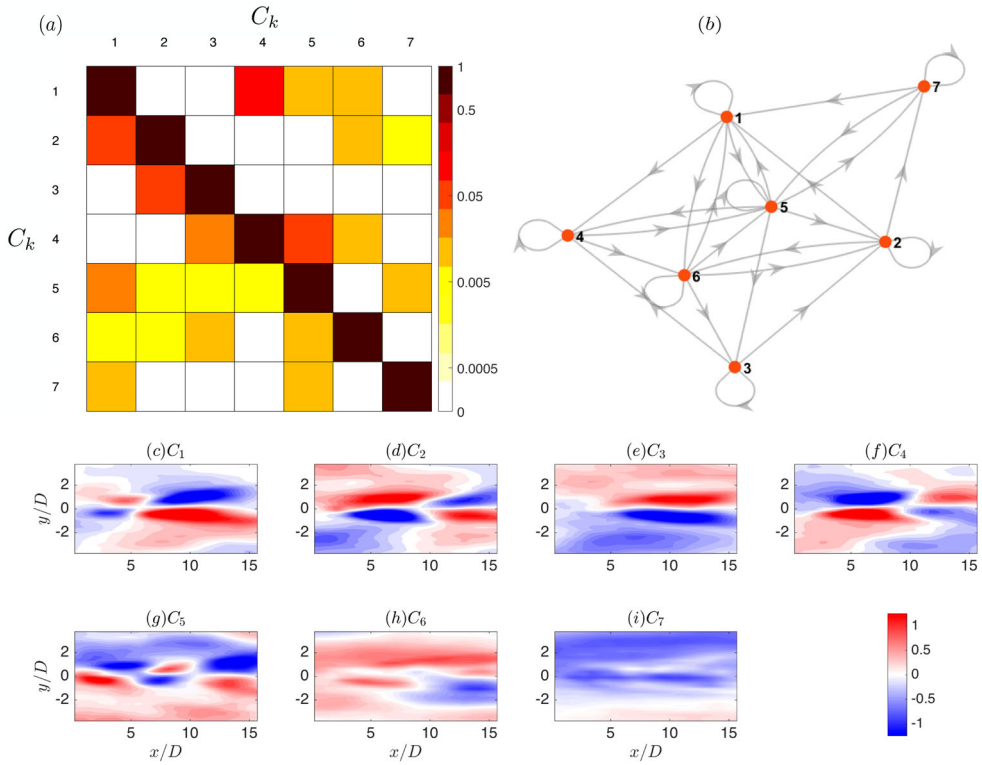
Figure 8. The cluster parameters: (a) CTM, (b) Markov and (c)-(i) centroids. Figure 7(b) is generated using directed graph function of the Markov chain. Each of the nodes in the graph represent a cluster. The arrows indicate the direction of movement from one state to another. The contour plots present the centroids for different clusters at the hub height. The colour bar presents the fluctuating velocity.

continues as a class of its own. All clusters considered as aperiodic because they have a self-loop. Similarly, the aperiodic class is shown in the other clusters.

To give a physical interpretation for each cluster and its relation with other clusters through transition schematics, the Reynolds shear stress is computed per cluster. This step is obtained by determining the label (the index of each cluster that quantifies each snapshot belonging to a particular cluster) of each snapshot and combining them based on a similar label. The total of the Reynolds shear stress per cluster is equal to that determined from the total snapshot. The weighted means are calculated as

$$A = \frac{\sum_{i=1}^{N} A_i B_i}{\sum_{i=1}^{N} B_i},$$ (20)

where A is the weighted mean and B is the weights. Figure 9 presents the probability of each cluster as determined based on the number of snapshot per cluster and the Reynolds shear stress of each cluster. The probability obtained using Equation (12) reveals that most probabilistic features in this flow are those shown in clusters C_1 and C_3, with $q = 18\%$ and $q = 16.5\%$, respectively. From the centroids shown in Figures 6 and 8, one can see the features with higher probability. The minor probabilistic features are those shown in clusters
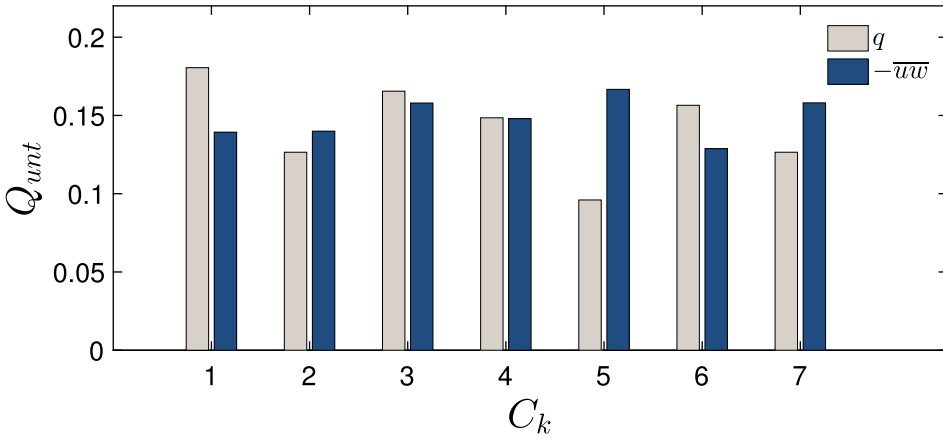
Figure 9. Quantities corresponding the clusters analysis. The probability and Reynolds shear stress.

C_2 and C_5. These clusters are related to the feature of small structures of the wake that also advected in the far-wake region. Although the probability shows a diverse distribution between clusters, the Reynolds shear stress is more uniformly distributed per cluster. Also, with only 9.6% of the probability, cluster C_5 carries about 16% of the Reynolds shear stress. In contrast, cluster C_1 displays the probability of 18% and carries about 13% of the stress. Cluster C_5 displays the structure near the rotor that is also responsible about the turbulence kinetic energy production and flux event in the near-wake region.

To define the region of the wake participating in the energy exchange per cluster, the conditional sampling method introduced by Lu and Willmarth [58] provides a framework to interpret, detect coherent motion signatures, and quantify the contributing nature of the Reynolds stress. In wind turbine wake flows, several studies have highlighted wake structures, entrainment mechanisms and recovery processes of energy downstream of the turbines [59]. Conditional averaging breaks up the signal of the fluctuating velocity components into four quadrants based on the signs of the components as [60–62]

$$\langle uw \rangle_{C_i} = \frac{1}{N_{C_i}} \sum_{t=1}^{N_{C_i}} u_t(x, y, z) w_t(x, y, z) \, \mathbb{S}_m \left[u_t(x, y, z); w_t(x, y, z) \right], \tag{21}$$

where u and w represent the flux components, m presents the quadrant index (i.e. $m = 1$, 2, 3, and 4), t is the time, and N_{C_i} is the total number of the snapshots belong to a specific cluster. Further, $\mathbb{S}_m$ is a conditional function defined as

$$\mathbb{S}_m \left[u_t(x, y, z); w_t(x, y, z) \right] = \begin{cases} 1 & \text{if } (u, w) \in m, \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

The corresponding quadrant combination for the momentum is presented in Table 1. Quadrant analysis with respect to momentum flux is presented here to interpret and detect the signature of the coherent motions per cluster. Only the sweep and ejection events are shown here due to the dominance of these quadrants in the flow. Figure 10 presents the distribution for the second and fourth quadrants per cluster. The sweep and ejections are

Table 1. Four quadrants based on the signs of the velocity components.

Flux	Quadrant 1 (Q1)	Quadrant 2 (Q2)	Quadrant 3 (Q3)	Quadrant 4 (Q4)
uw	u^+w^+	u^-w^+	u^-w^-	u^+w^-
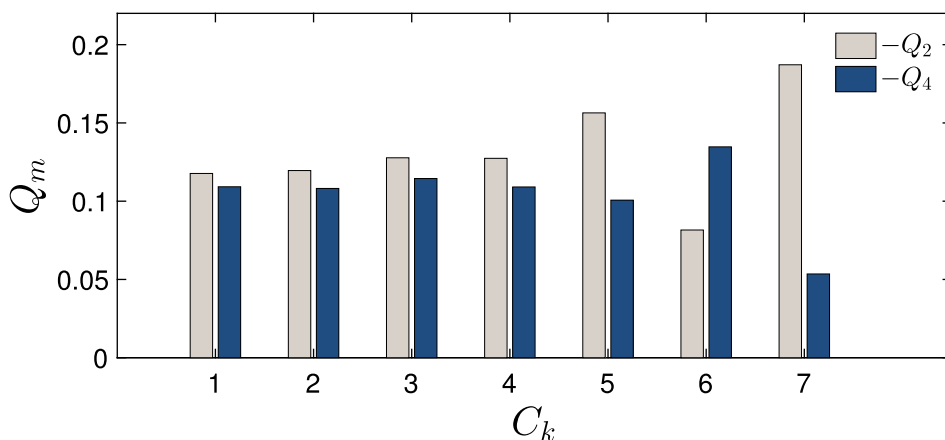
Figure 10. Quadrant analysis corresponding with Reynolds shear stress of each cluster.

the names for the negative shear stress. The sweep is defined when ($u' > 0$ & $w' < 0$) and the ejection is defined when ($w' > 0$ & $u' < 0$). Results show that a maximum for ejections is found just at cluster C_5 and C_7, while a maximum for sweeps is observed at clusters C_3 and C_6. Both of these maxima for ejections and sweeps contribute to the vertical kinetic energy flux. The patterns found in the Reynolds shear stresses in regards to ejection events show a preference for vertical expansion and interaction of the structure near the rotor, see Figure 6 and Figure 8. Accordingly, the region of the near-wake shows a substantial contribution to the cluster. Indicating that the coherent structures arise from the large vertical Reynolds shear stress near the wake boundary. In contrast, the features observed in the Reynolds shear stress in regards to sweep display that a preference for downwards entrainment expansion comes from the flow outside the wake region.

The spectrum of the transition is determined as $\Gamma = \log(\lambda_c)/\Delta t$, where λ_c are the eigenvalues of the CTM, and Δt is the time step. The analysis furnishes a view of the stability of the transition matrix via the growth rates and frequencies. The eigenvalues of the CTM are pictured in the complex plane, see Figure 11. The real part σ illustrates the growth rate and the imaginary part ω demonstrates the frequency. The Strouhal number (St) depicts the oscillating flow mechanisms and is defined as $St = \omega/2\pi$. As shown in Figure 11, the uniform distribution at the frontier stable eigenvalue $\lambda_c = 1$ has zero growth rate and frequency. Thus there are two oscillatory modes: the smallest damping is located at frequency $St = \omega/2\pi = 0$, and the largest one is displayed at $St \approx 0.0461$. An intuitive sight of the underlying dynamics is obtained via the visualisation of trajectories in two-dimensional space, i.e. the Voronoi diagram. Figure 11 shows the visualisation based on the magnitude of the first two POD coefficients (a^1 and a^2). The edges of the Voronoi cells are surrounded by black lines, and the uni colour observation points are located inside a specific cell. The big black dots are the centroids of the cluster. The discretisation displays the internal and
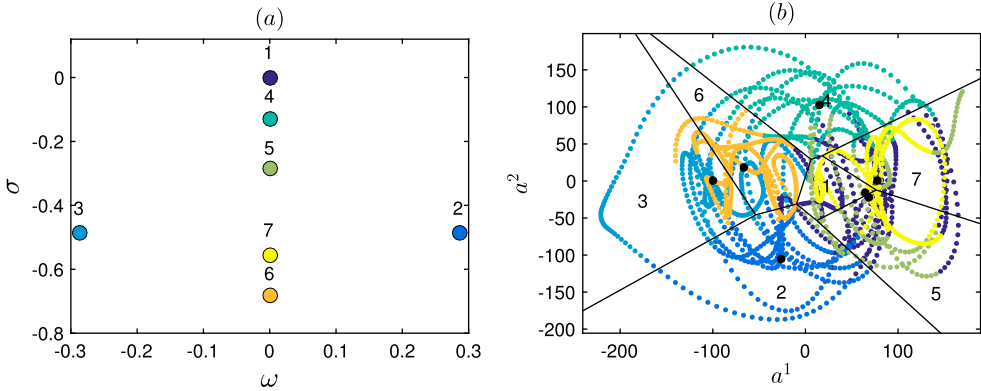
Figure 11. (a) The spectrum of the CTM and (b) the visualisation of the Voronoi diagram of the clusters based on the space of the first two POD coefficients. Each colour presents a specific cluster.

external ring of clusters. This result shows again the efficiency of the CROM to detect the flow regimes.

After identifying the flow features via cluster centroids and quantifying the interconnections among the clusters by the transition matrix, it is the time to assess the performance of the probabilistic dynamical model by corresponding the predicted asymptotic likelihood distribution at a large number of iterations. The Markov model has compared with the relative probabilities obtained from the data q_j, see Figure 9. The transition matrix highlights the temporal trend of a Markov chain. With a considerable number of time steps, the Markov chain converges to a special and stationary probability distribution. In other words, the system is probabilistically reproducible. The asymptotic state presents the proportion of time that each trajectory consumes in each cluster. More, the stationary distribution corresponds to the eigenvector associated with the largest eigenvalues of the CTM as a result of the Perron–Frobenius theorem. The cluster probabilities shown in Figure 12 are not
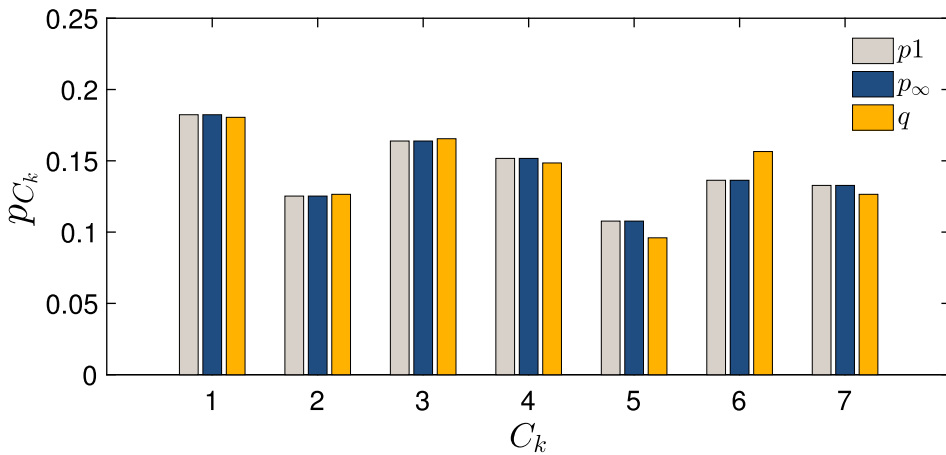
Figure 12. The cluster probability distributions from the data, converged iteration of the transition dynamics, and eigenvector associated with the dominant eigenvalue.
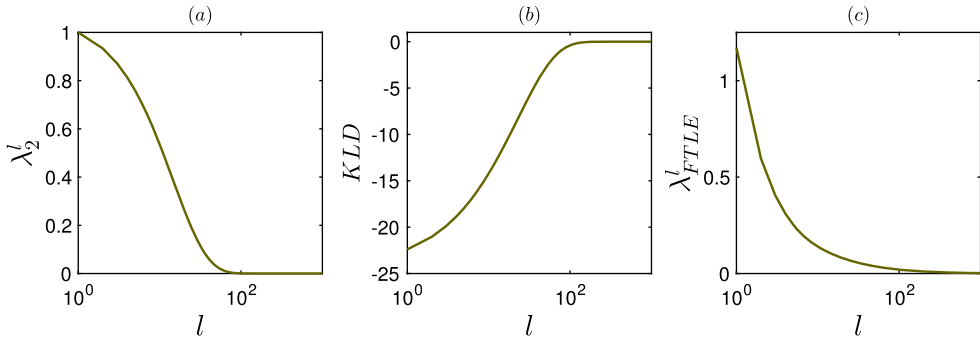
Figure 13. The convergence of the CROM with respect to the number of iterations *l*. The convergence is tested based on (a) second eigenvalue modulus, (b) Kullback–Leibler entropy and (c) finite-time Lyapunov exponent.

strongly unreasonable to any state. The variations among probabilities are a result of the time evolution of the attractors that are assessed by the CTM.

To check the validity of the Markov model presented in Figure 8, further analysis is applied to inspect the evolution of the attractors. Finite-time Lyapunov exponents (FTLE) can quantify the mixing property of the propagator P and divergence of the distance as

$$\lambda_{FTLE}^{l} = \frac{1}{l} \ln \left\| \frac{P^l \delta p^0}{\delta p^0} \right\|. \tag{23}$$

FTLE is particularly dependent on initial condition and the dependency decreases rapidly with increasing l,

$$\lambda_{FTLE} = \lim_{l \to \infty} \lambda_{FTLE}^{l} = 0. \tag{24}$$

Thus, after evaluating the transition matrix, the propagator can be approximated with a transition matrix as

$$Q_{jk} = q_j, \quad j, k = 1, \ldots, K. \tag{25}$$

Kullback–Leibler entropy [28,63–67] is used to determine the variation between the transition matrix as,

$$KLD(P, Q) = -\sum_{j=1}^{K} \sum_{k=1}^{K} P_{jk} \ln \frac{P_{jk}}{Q_{jk}}. \tag{26}$$

Figure 13 displays the convergence of the FTLE, Kullback–Leibler entropy, and second eigenvalue modulus. The convergence to zero is noticed at approximately $l = 10^2$. The results confirm the convergence behaviour of the transition matrix. The evolution of the Kullback–Leibler entropy quantifies important view about the Markov model, where the divergence of iterated CTM means that the information is lost via irreversible diffusion [28].

5. Conclusion

The techniques used to analyse the fluid dynamics in the wake of a single wind turbine revealed important features that provide significant insight into the physical processes occurring in the wakes. Cluster-based reduced-order modelling is a novel probabilistic model that is applied to expose the characteristic features of the flow. The time coefficients of the proper orthogonal decomposition are grouped based on the clustering technique and classified via Markov chain to describe the evolution of the probability distribution of the trajectories. The validity of the Markov model is tested through the FTLE and Kullback–Leibler entropy. The cluster analysis decomposes the data based on the similarity of snapshots and compresses them into a low number of representative states. The centroids of these states show the spatial structures of the flow and highlight the wake of the turbine, rotating structure, and their transition downstream the rotor. The flow states are sorted by analysis of the transition matrix that is dynamically modelled in an unsupervised manner. Characteristic frequencies obtained from the spectrum analysis display the oscillation modes of the velocity. The results show that the Markov chain converges to a unique and stationary distribution and that the system is reproducible. The spectral analysis of the clusters provides further information on characteristic frequencies (oscillation damping) of the ensemble trajectories. Quadrant analysis is presented to detect the signature of the coherent motions per cluster. The sweep and ejection events are mapped through the transition from a cluster to another. The aspects that are obtained from the current data-driven method show the great potential for generating a probabilistic predicting model for the large-scale features in large wind farms.

Accurate and timely measurements are the optimal keys to control turbines. The convenient locations for installation and maintenance are usually used to place the sensors, although these places might not be the optimal places to obtain optimal measurements. The transition of the Markov chain can be used to highlight the optimal locations for measurement to improve the dynamic feedback control through estimating the full state from the model. This point is a practical view, especially in the real wind farm the number of the sensor is much less than the number of observations. From this perspective, the future work will combine the cluster-based reduced-order model with the sparse sensor placement optimisation for classification to obtain real-time prediction and control. The probabilistic dynamics from few optimised measurements can be identified from the sparsity to facility faster computations, prediction and control high-dimensional systems. Sparse sensor placement optimisation for classification framework is based on dimensionality reduction and discrimination techniques as well to learn sparse sensor locations that allow classification of full-state measurements for the wind plant from part-state measurements.

6. Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Naseem Ali http://orcid.org/0000-0002-8560-4406
Raúl Bayoán Cal http://orcid.org/0000-0003-1642-787X

References

[1] Scott R, Bossuyt J, Cal RB. Characterizing tilt effects on wind plants. J Renew Sust Energ. 2020;12(4):043302.
[2] Hamilton N, Tutkun M, Cal RB. Anisotropic character of low-order turbulent flow descriptions through the proper orthogonal decomposition. Phys Rev Fluids. 2017;2(1):014601.
[3] Ali N, Hamilton N, Calaf M, et al. Turbulence kinetic energy budget and conditional sampling of momentum, scalar, and intermittency fluxes in thermally stratified wind farms. J Turbulence. 2019;20(1):32–63.
[4] Hu J, Tang J, Lin Y. A novel wind power probabilistic forecasting approach based on joint quantile regression and multi-objective optimization. Renew Energ. 2020;149:141–164.
[5] Kadum H, Cal RB, Quigley M, et al. Compounded energy gains in collocated wind plants: energy balance quantification and wake morphology description. Renew Energ. 2020;150:868–877.
[6] Scott R, Viggiano B, Dib T, et al. Wind turbine partial wake merging description and quantification. Wind Energ. 2020;9:1610.
[7] Bossuyt J, Scott R, Ali N, et al. Quantification of wake shape modulation and deflection for tilt and yaw misaligned wind turbines. J Fluid Mech. 2021;917:A3. doi:10.1017/jfm.2021.237.
[8] Keane A, Aguirre PEO, Ferchland H, et al. An analytical model for a full wind turbine wake. Journal of Physics: Conference Series; Vol. 753, IOP Publishing; 2016. p. 032039.
[9] Mortensen NG, Landberg L, Troen I, et al. Wind atlas analysis and application program (WAsP). Vol. 1, Getting started 1.
[10] Larsen GC, Aagaard HM, Bingöl F, et al. Dynamic wake meandering modeling.
[11] Boersma S, Doekemeijer BM, Gebraad PM. A tutorial on control-oriented modeling and control of wind farms. 2017 American Control Conference (ACC); IEEE; 2017. p. 1–18.
[12] Annoni J, Fleming P, Scholbrock A, et al. Analysis of control-oriented wake modeling tools using Lidar field results. typeTech. Rep., institution National Renewable Energy Lab. (NREL); Golden, CO: 2018.
[13] Calaf M, Meneveau C, Meyers J. Large eddy simulation study of fully developed wind-turbine array boundary layers. Phys Fluids. 2010;22(1):015110.
[14] Wu Y-T, Porté-Agel F. Large-eddy simulation of wind-turbine wakes: evaluation of turbine parametrisations. Boundary Layer Meteorol. 2011;138(3):345–366.
[15] Melius M, Tutkun M, Cal RB. Identification of Markov process within a wind turbine array boundary layer. J Renew Sust Energ. 2014;6(2):023121.
[16] Melius MS, Tutkun M, Cal RB. Solution of the Fokker–Planck equation in a wind turbine array boundary layer. Phys D: Nonlinear Phenomena. 2014;280:14–21.
[17] Camp EH, Cal RB. Mean kinetic energy transport and event classification in a model wind turbine array versus an array of porous disks: energy budget and octant analysis. Phys Rev Fluids. 2016;1(4):044404.
[18] Rockel S, Peinke J, Hölling M, et al. Dynamic wake development of a floating wind turbine in free pitch motion subjected to turbulent inflow generated with an active grid. Renew Energ. 2017;112:1–16.
[19] Ali N, Cortina G, Hamilton N, et al. Turbulence characteristics of a thermally stratified wind turbine array boundary layer via proper orthogonal decomposition. J Fluid Mech. 2017;828:175–195.
[20] Ali N, Hamilton N, DeLucia D, et al. Assessing spacing impact on coherent features in a wind turbine array boundary layer. Wind Energy Sci. 2018;3(1):43–56.
[21] Ali N, Cal RB. Data-driven modeling of the wake behind a wind turbine array. J Renew Sust Energ. 2020;12(3):033304.

[22] Antoulas AC. Approximation of large-scale dynamical systems. Vol. 6. Philadelphia: SIAM; 2005.

[23] Rowley CW. Model reduction for fluids, using balanced proper orthogonal decomposition. Modeling And Computations In Dynamical Systems: In Commemoration of the 100th Anniversary of the Birth of John von Neumann; World Scientific; 2006. p. 301–317.

[24] Bistrian DA, Susan-Resiga RF. Weighted proper orthogonal decomposition of the swirling flow exiting the hydraulic turbine runner. Appl Math Model. 2016;40(5–6):4057–4078.

[25] Jin Y, Lu K, Hou L, et al. An adaptive proper orthogonal decomposition method for model order reduction of multi-disc rotor system. J Sound Vib. 2017;411:210–231.

[26] Noack BR, Afanasiev K, Morzyński M, et al. A hierarchy of low-dimensional models for the transient and post-transient cylinder wake. J Fluid Mech. 2003;497:335–363.

[27] Noack BR, Morzynski M, Tadmor G. Reduced-order modelling for flow control. Vol. 528. Italy: Springer Science & Business Media; 2011.

[28] Kaiser E, Noack BR, Cordier L, et al. Cluster-based reduced-order modelling of a mixing layer. J Fluid Mech. 2014;754:365–414.

[29] Östh J, Kaiser E, Krajnović S, et al. Cluster-based reduced-order modelling of the flow in the wake of a high speed train. J Wind Eng Ind Aerod. 2015;145:327–338.

[30] Hamilton N, Viggiano B, Calaf M, et al. A generalized framework for reduced-order modeling of a wind turbine wake. Wind Energ. 2018;21(6):373–390.

[31] Cao Y, Kaiser E, Borée J, et al. Cluster-based analysis of cycle-to-cycle variations: application to internal combustion engines. Exp Fluids. 2014;55(11):1837.

[32] Kaiser E, Morzyński M, Daviller G, et al. Sparsity enabled cluster reduced-order models for control. J Comput Phys. 2018;352:388–409.

[33] Wei Z, Yang Z, Xia C, et al. Cluster-based reduced-order modelling of the wake stabilization mechanism behind a twisted cylinder. J Wind Eng Indust Aerod. 2017;171:288–303.

[34] Barwey S, Hassanaly M, An Q, et al. Experimental data-based reduced-order model for analysis and prediction of flame transition in gas turbine combustors. Combust Theor Model. 2019;23:1–27.

[35] Ali N, Viggiano B, Tutkun M, et al. Cluster-Based reduced-Order descriptions of two phase flows. Chem Eng Sci. 2020;222:115660.

[36] Liu Y, Gao X, Yan J, et al. Clustering methods of wind turbines and its application in short-term wind power forecasts. J Renew Sust Energ. 2014;6(5):053119.

[37] Wang K, Qi X, Liu H, et al. Deep belief network based k-means cluster approach for short-term wind power forecasting. Energ. 2018;165:840–852.

[38] Wu W, Peng M. A data mining approach combining k-means clustering with bagging neural network for short-term wind power forecasting. IEEE Int Things J. 2017;4(4):979–986.

[39] Ali N. Thermally (un-) stratified wind plants: stochastic and data-driven reduced order descriptions/modeling [Ph.D. thesis] Portland State University; 2018.

[40] Ali N, Calaf M, Cal RB. Clustering sparse sensor placement identification and deep learning based forecasting for wind turbine wakes. J Renew Sust Energ. 2021;13(2):023307.

[41] Burkardt J, Gunzburger M, Lee H-C. POD and CVT-based reduced-order modeling of Navier–Stokes flows. Comput Methods Appl Mech Eng. 2006;196(1–3):337–355.

[42] Ali N, Calaf M, Cal RB. Reduced-order modeling of the wake behind a single wind turbine. iTi Conference on Turbulence; Springer; 2018. p. 285–290.

[43] Hamilton N, Tutkun M, Cal RB. Low-order representations of the canonical wind turbine array boundary layer via double proper orthogonal decomposition. Phys Fluids. 2016;28(2):025103.

[44] Holmes P, Lumley JL, Berkooz G. Turbulence, coherent structures, dynamical systems and symmetry. UK: Cambridge University Press; 1998.

[45] Hamilton N, Tutkun M, Cal RB. Wind turbine boundary layer arrays for Cartesian and staggered configurations: part II. low-dimensional representations via the proper orthogonal decomposition. Wind Energ. 2015;18(2):297–315.

[46] Ali N, Kadum HF, Cal RB. Focused-based multifractal analysis of the wake in a wind turbine array utilizing proper orthogonal decomposition. J Renew Sust Energ. 2016;8(6):063306.

[47] Camp EH, Cal RB. Low-dimensional representations and anisotropy of model rotor versus porous disk wind turbine arrays. Phys Rev Fluids. 2019;4(2):024610.

[48] Arthur D, Vassilvitskii S. k-means++: The advantages of careful seeding. Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms; SIAM; 2007. p. 1027–1035.

[49] Ali N, Hamilton N, Calaf M, et al. Classification of the Reynolds stress anisotropy tensor in very large thermally stratified wind farms using colormap image segmentation. J Renew Sust Energ. 2019;11(6):063305.

[50] Sharma V, Calaf M, Lehning M, et al. Time-adaptive wind turbine model for an LES framework. Wind Energ. 2015;18(5):0.

[51] Ali N, Hamilton N, Cortina G, et al. Anisotropy stress invariants of thermally stratified wind turbine array boundary layers using large eddy simulations. J Renew Sust Energ. 2018;10(1):013301.

[52] Dutt A, Subramani DN, Kulkarni CS, et al. Clustering of Massive Ensemble of Vehicle Trajectories in Strong, Dynamic and Uncertain Ocean Flows. OCEANS 2018 MTS/IEEE Charleston; IEEE; 2018. p. 1–7.

[53] Brunton SL, Kutz JN. Data-driven science and engineering: machine learning, dynamical systems, and control. UK: Cambridge University Press; 2019.

[54] Newman AJ, Drew DA, Castillo L. Pseudo spectral analysis of the energy entrainment in a scaled down wind farm. Renew Energ. 2014;70:129–141.

[55] VerHulst C, Meneveau C. Large eddy simulation study of the kinetic energy entrainment by energetic turbulent flow structures in large wind farms. Phys Fluids. 2014;26(2):025113.

[56] Andersen SJ, Sørensen JN, Mikkelsen RF. Turbulence and entrainment length scales in large wind farms. Philos Trans Royal Soc A Math Phys Eng Sci. 2017;375(2091):20160107.

[57] Quan LH, Longhetto A, Ferrero E. The characteristics of low-speed streaks in near-neutral and unstable atmospheric boundary layer. Nuovo Cimento Della Soc Italiana Di Fisica. B, Gen Phys, Relativity Astron Math Phys Methods. 2009;124(3):325–340.

[58] Lu SS, Willmarth WW. Measurements of the structure of the Reynolds stress in a turbulent boundary layer. J Fluid Mech. 1973;60(3):481–511.

[59] Cal RB, Lebrón J, Castillo L, et al. Experimental study of the horizontally averaged flow structure in a model wind-turbine array boundary layer. J Renew Sust Energ. 2010;2(1):013106.

[60] Kadum HF, Knowles D, Cal RB. Quantification of preferential contribution of Reynolds shear stresses and flux of mean kinetic energy via conditional sampling in a wind turbine array. J Fluids Eng. 2019;141(2):021201.

[61] Viggiano B, Dib TS, Ali N, et al. Turbulence, entrainment and low-order description of a transitional variable-density jet. J Fluid Mech. 2018;836:1009–1049.

[62] Viggiano B, Sakradse S, Smith G, et al. Intermittent event evaluation through a multifractal approach for variable density jets. Chaos Solitons & Fractals. 2021;146:110799.

[63] Kullback S, Leibler RA. On information and sufficiency. Ann Math Stat. 1951;22(1):79–86.

[64] Kullback S, Information theory and statistics. New York: Courier Corporation; 1997.

[65] Ali N, Tutkun M, Cal RB. Turbulent boundary layer features via lagrangian coherent structures, proper orthogonal decomposition and dynamic mode decomposition, preprint arXiv:1704.04090

[66] Ali N, Fuchs A, Neunaber I, et al. Multi-scale/fractal processes in the wake of a wind turbine array boundary layer. J Turbulence. 2019;20(2):93–120.

[67] Ali N, Cal RB. Scale evolution, intermittency and fluctuation relations in the near-wake of a wind turbine array. Chaos, Solitons & Fractals. 2019;119:215–229.