

PAPER • OPEN ACCESS

## On Stochastic Reduced-Order and LES-based Models of Offshore Wind Turbine Wakes

To cite this article: Mostafa Bakhoday Paskyabi *et al* 2020 *J. Phys.: Conf. Ser.* **1669** 012018

View the [article online](#) for updates and enhancements.



**240th ECS Meeting** ORLANDO, FL

Orange County Convention Center Oct 10-14, 2021



Abstract submission due: April 9

**SUBMIT NOW**

# On Stochastic Reduced-Order and LES-based Models of Offshore Wind Turbine Wakes

**Mostafa Bakhoday Paskyabi, Maria Krutova, Finn Gunnar Nielsen, Joachim Reuder, and Omar El Guernaoui**

Geophysical Institute, University of Bergen, Bergen Offshore Wind Centre, and Bjerknes Centre for Climate Research, Bergen, Norway

E-mail: [Mostafa.Bakhoday-Paskyabi@uib.no](mailto:Mostafa.Bakhoday-Paskyabi@uib.no)

**Abstract.** In this paper, the primary objective is to investigate flow structures in the wake of wind turbines based on applying a truncated Proper Orthogonal Decomposition (POD) approach. This scheme decomposes the three-dimensional velocity fields produced by the high-fidelity PARallelized LES Model (PALM) into a number of orthogonal spatial modes and time-dependent weighting coefficients. PALM has been combined with an actuator disk model with rotation to incorporate the effects of a turbine array. The time-dependent deterministic weights from applying the POD scheme are replaced by stochastic weights, estimated from two independent stochastic techniques that aim to account for unresolved small-scale features for a number of POD modes. We then reconstruct the flow field by a small number of stochastic modes to investigate how well the applied stochastic methodologies can reproduce the flow field compared to the original LES results.

## 1. Introduction

Highly variable wakes and wake-wake interaction within the area of a wind farm play a key role in governing the performance of the wind farm (e.g. its total power output) and controlling the fatigue loads and intermittency events acting on downwind turbines. Furthermore, the asymmetrical nature of wake evolution due to its interactions with multiscale processes in the atmospheric boundary layer may lead to an average energy loss of about 5-20% [1]. Therefore, accurate modelling of wake dynamics and interaction is crucial for optimizing the layout as well as operational performance of (offshore) wind farms.

Wake effects can be studied using numerical wake models aiming to essentially capture the momentum deficit and the enhanced level of turbulence within the area of wind farm/turbines (i.e. near- and far-wake regions). How to predict key wake features, such as meandering of the wake, far-field (approximately  $\geq 4$  rotor diameter) recovery of the velocity deficit, effects of rotor geometry, wake interactions with shear and turbulence in the atmospheric boundary layer, near-field rotor-induced turbulence, and wake-induced turbine loads have led to emergence of different modelling approaches over the last decade, see for example [1]. We may classify these models into three main groups: empirical models that estimate analytically the wake velocity based on some over-simplified assumptions and ideal hypothesis [4, 8]; models that treat turbines as roughness elements applicable for wake loss prediction over a large wind farm [10, 3]; and Computational Fluid Dynamics (CFD) models with different levels of accuracy and complexity that include the dynamics of turbines and the flow fields [2]. While CFD models, such as Large Eddy Simulations



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

(LESs), account for large and small scales relevant to the wind turbines/farm wakes, they are computationally expensive for practical wind energy applications. Furthermore, major concern in the first two groups is that they are either relying on over-simplified assumptions or excluding important features of the wake dynamics (e.g. steady wakes) to capture wake instabilities. To bridge the gap between computational cost and capturing the relevant scales and dynamics, efficient Reduced Order models (ROMs) need to be developed.

In reduced-order modeling, a complex dynamical system with many degrees of freedom (DOF) is replaced by a limited number of DOFs through isolating dominant dynamical features for a given flow by an effective use of high-fidelity CFD models. Proper Orthogonal Decomposition (POD), as one frequently used modal decomposition methods, is a computationally efficient approach in the analysis of wind energy. The modal basis functions in the POD method, characterized by spatial orthogonality, can also be coupled to governing equations of flow motions through a Galerkin projection, so-called Galerkin-POD procedure, to reconstruct the entire flow field [5]. In Galerkin-based methods, the temporal dynamics of POD modes is described by a system of nonlinear ordinary differential equations. To capture the missing small-scale motions in the flow field, these temporal coefficients can be alternatively treated as a stochastic process described by, for example, a Gaussian process or a random walk model of truncated POD modes [6, 7]. However, due to the truncation of higher-order and dissipative scales in the POD-based models, the accuracy of ROM requires the use of a large number of training data, "snapshots", that can be effectively derived from CFD simulation results. The POD-based methodologies can then be used for near real-time (short-term) prediction of wind fields for the wind energy applications with an acceptable accuracy.

In the present work, LES results of wind turbine wakes under neutral atmospheric condition are analyzed through the truncated POD approaches. We focus on the temporal dynamics of modal basis functions by investigating how the representation of the missing small-scale features (i.e. turbulent structures) in the standard POD method can be improved through incorporating two stochastic procedures. We then conduct a series of statistical analyses in the wake field estimated from the POD-based ROMs against the high-fidelity LES model data.

## 2. LES model

To study flow fields and wake structure/variability, we use LES configured for neutral boundary layer conditions. LES resolves the large scale eddies by filtering out the small scales of motions. Small scales are then incorporated by the means of sub-grid scale parameterizations for scales of  $O(1)$  m. In this study, we use the Paralleized Large-Eddy Simulation Model (PALM) for solving non-hydrostatic and incompressible atmospheric boundary layer flow fields (i.e. using the Boussinesq approximation) [13]. PALM uses an Arakawa C-grid for spatial discretization and turbines are implemented in this model using an Actuator Disk Model with Rotation (ADM-R) [9]. The ADM-R parameters in this study have been set to model the NREL 5 MW reference turbine [12].

### 2.1. Simulation setup

The domain size is  $6912 \times 2304 \times 1459$  m with a grid size of  $\Delta x = \Delta y = \Delta z = 6$  m. The grid cell is stretched in  $z$  direction above 800 m with the factor of 1.04, maximum cell size is capped at  $\Delta z_{max} = 12$  m. We locate two wind turbines aligned from the west to east boundary with  $8D$  distance from each other, where  $D$  denotes the turbine diameter. At the sea surface, we use the Neumann Boundary Condition (BC) for the velocity and scalar fields. We use Dirichlet BC for the upper boundary condition. We assign periodic BC for the north and the south boundaries and non-periodic (inflow-outflow conditions) for the west-east (direction of main flow) boundaries. Using non-periodic BC for the west and east boundaries necessitates the use of a precursor run to produce initial steady and fully turbulent flow in a smaller domain

(2304×2304×1459 m) which will be used later to fill out the main run domain through a cyclic repetition procedure. In the main run, we use radiation boundary conditions at the outflow (east) boundary for velocity and Neumann boundary conditions for the scalar fields. We have discarded planetary rotation in both the precursor and the main runs. The main run has been performed for a period of 3 days. We use the first 2.97 days for the analysis and the rest of the data for a forecast in this study. The output data have been stored every 10 minutes.

### 3. Methods

Since all ROMs rely on information and features provided by the data (i.e. a data-driven procedure), the quality of LES snapshots is a key to capture important features by modal decomposition. In this section, we first give a brief overview of POD for dynamic wake modelling and then construct our POD-based ROM. Finally, temporal weight coefficients in the modal representation are modelled as independent stochastic processes to check whether the stochastic forcing can partly capture the effects of missing small scale dynamics.

#### 3.1. Proper Orthogonal Decomposition (POD)

Let us write  $\mathbf{x} = (x, y, z)$  for the position vector,  $\mathbf{u} = (u, v, w)$  for the wind velocity vector,  $\bar{\mathbf{u}}$  as the mean, and  $\mathbf{u}' = \mathbf{u} - \bar{\mathbf{u}}$  as fluctuating velocity, all estimated from the LES data. The idea behind the POD is to decompose the fluctuating velocity field into a number of deterministic spatial fields,  $\phi_i(\mathbf{x})$ , modulated by random time-dependent weighting coefficients,  $a_i(t)$ :

$$\mathbf{u}'(\mathbf{x}, t) = \sum_{i=1}^N a_i(t) \phi_i(\mathbf{x}), \quad (1)$$

where  $t$  presents time, and  $N$  is the total number of modes in the above truncated representation. The POD is optimal in the sense that the expansion in Eq. (1) maximizes the kinetic energy captured by the first  $N$  POD modes. To calculate these  $a_i(t)$  and  $\phi_i(\mathbf{x})$ , we start by creating a snapshot matrix,  $A$ , using  $n_t$  LES 3D data, each with the size of  $n_x \times n_y \times n_z$ :

$$A = \begin{bmatrix} \cdots & \cdots \\ \cdots & \cdots \\ \cdots & \cdots \end{bmatrix}_{n_t \times m}, \quad (2)$$

where  $m = 3n_x \times n_y \times n_z$  for the 3D velocity field. The Singular Value Decomposition (SVD) of the correlation matrix  $C = AA^T$  will result in eigenvalues (energy of each mode) and eigenvectors (orthogonal modes) of the POD decomposition. Due to orthogonality of these modes, we have

$$a_j(t) = \int \int \int_{\mathbf{x}} \mathbf{u}'(\mathbf{x}, t) \phi_j(\mathbf{x}) d\mathbf{x} \quad \text{and} \quad \overline{a_i(t) a_j(t)} = \lambda_j \delta_{ij}, \quad (3)$$

where the overbar denotes temporal averaging, and  $\lambda_j$  denotes the strength of the  $j$ th POD mode. Here,  $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$  are real eigenvalues (calculated from SVD) so that  $\overline{a_j(t)^2} = \lambda_j$  for  $j = 1, \dots, N$ . In the POD, modes are ranked by their energy.

#### 3.2. Stochastic models

In Eq. (1), all dynamical information associated with temporal variation of wind is encapsulated in the time-dependent weighting coefficients,  $a_j(t)$ , where  $j = 1, \dots, N$  and  $0 < t < T$  for the final analysis time  $T$ . As the wind field describes a stochastic phenomenon, it is natural to use probabilistic methods to model/predict  $a_j(t)$  by including some stochasticity. In the statistical viewpoint, the empirical probability density of these time-dependent weights fits almost properly

to the normal distribution with an auto-correlation approximated by an exponential-decaying function in time (this is according to our analysis). These characteristics allow us to apply different stochastic models such as Gaussian process and state-dependent Stochastic Differential Equations (SDEs) to construct time series of  $a_j$ . Here, we provide a very brief description of these methods.

*3.2.1. Standard Gaussian Process (GP)* Assume that  $\hat{a}_j$  is a Gaussian process used to produce the time series of  $a_j$  for any  $j$  [6]. This stochastic process at the  $j$ th POD mode is explained by its mean,  $\mu_j$ , and covariance function,  $k_j(t, t')$ , as  $\hat{a}_j(t) \sim GP(\mu_j, k_j(t, t'))$  so that

$$\mu_j(t) = \Xi[\hat{a}_j(t)], \quad (4)$$

$$k_j(t, t') = \Xi[(\hat{a}_j(t) - \mu_j)(\hat{a}_j(t') - \mu_j)]. \quad (5)$$

Here,  $\Xi$  denotes the mathematical expectation value. We can consider the *GP* as a classical regression of  $M$  observation pairs  $(t_j^i, y_j^i)$  (or training set at  $j$ th POD mode) so that

$$y_j^i = \hat{a}_j(t_i) + \epsilon_j, \quad (6)$$

where  $i = 1, \dots, M$ , and  $\epsilon_j$  denote an additive Gaussian noise with a zero mean and variance of  $\sigma_\epsilon^2$ , i.e.  $\epsilon_j \sim N(0, \sigma_\epsilon^2)$ . Given the training set, we aim to predict the target values  $y_j^*$  for new target input  $t^*$  (i.e.  $t \leq t^*$  as interpolation and  $t > t^*$  as forecast). Since  $p(y_j|t) = N(0, k_j + \sigma_\epsilon^2 I)$ , the distribution for the new input in connection with the training set is given by:

$$\begin{bmatrix} y_j \\ y_j^* \end{bmatrix} \sim N \left( 0, \begin{bmatrix} k_j(t, t) + \sigma_\epsilon^2 I & k_j(t, t^*) \\ k_j(t^*, t) & k_j(t^*, t^*) \end{bmatrix} \right), \quad (7)$$

where  $k_j(t, t^*) = k_j(t^*, t)$ . According to the property of the Gaussian distribution, the prediction at each  $j$  is obtained as:

$$\bar{y}_j^* = k_{*j}^T (k_j + \sigma_\epsilon^2 I)^{-1} y_j, \quad (8)$$

where  $k_{*j} = k_j(t^*, t^*)$  and  $k_j = k_j(t, t)$ .

*3.2.2. SDE-based model* We aim to reconstruct the POD weighting coefficients,  $a_j$ , by another SDE-based process  $\hat{a}_j$  at the  $j$ th POD mode. In Ito-sense, the one-dimensional stochastic differential equation for the evolution of  $\hat{a}_j$  obeys the following general form [15]:

$$d\hat{a}_j(t) = f(\hat{a}_j(t), t) \cdot dt + g(\hat{a}_j(t), t) \cdot dW(t), \quad (9)$$

where  $f(\hat{a}_j(t), t)$  and  $g(\hat{a}_j(t), t)$  are drift and diffusion terms that define the statistical properties of  $\hat{a}_j(t)$ .  $W(t)$  denotes a standard Wiener process which satisfies three conditions: (i)  $W(0) = 0$ ; (ii) for random variables  $0 < s < t < T$ , the increment  $W(t) - W(s)$  has normal distribution with zero mean and variance of  $t - s$ ; and (iii) for  $0 < s < t < \alpha_1 < \alpha_2 < T$ , the increments  $W(t) - W(s)$  and  $W(\alpha_1) - W(\alpha_2)$  are statistically independent. Here,  $T$  denotes the final analysis time.

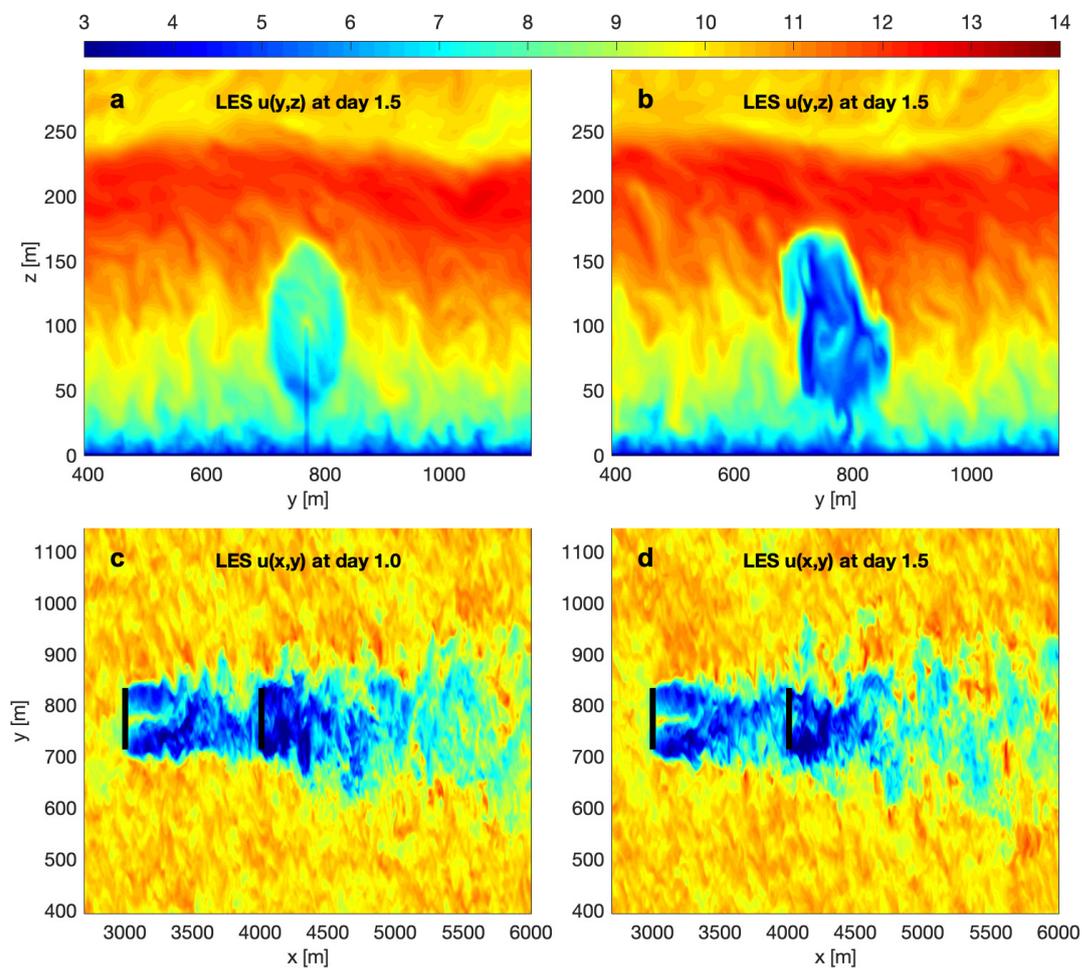
In Eq. (9),  $\hat{a}_j(t)$  represents the time series calculated from projecting the LES velocity field on the  $j$ th POD mode. In this method, we should define functions  $f$  and  $g$  such that the autocorrelation of  $\hat{a}_j(t)$  could be fitted to an exponential-decaying curve. In this paper, we assume that  $a_j(t)$  at each POD mode follows the Ornstein-Uhlenbeck process [17]:

$$d\hat{a}_j(t) = -\alpha_j(\mu_j - \hat{a}_j(t)) \cdot dt + \overbrace{\sigma_j \sqrt{2\alpha_j}}^{d\gamma(t)} \cdot dW(t), \quad (10)$$

where  $\gamma$  is a driftless subordinator,  $\mu_j$  and  $\sigma_j$  are the mean and the standard deviation of  $\hat{a}_j(t)$  at the  $j^{\text{th}}$  POD mode, respectively. The autocorrelation is governed by an exponential-decaying function with decay rate of  $\alpha_j$  for the  $j^{\text{th}}$  POD mode as follows:

$$\rho(\tau) = \overline{\hat{a}_j(t)\hat{a}_j(t+\tau)} = e^{-\alpha_j \cdot \tau}, \quad (11)$$

where  $\tau$  denotes the time lag, and the over-bar represents the time averaging operator. Generally, Eq. (10) needs to be solved numerically for two types of convergence, i.e. strong and weak. Strong convergence focuses on the stochastic trajectories by the means of describing the goodness of the approximation (i.e. mean of error), while the weak convergence captures the average behaviour of the stochastic process (i.e. error of the mean). Studying the convergence properties of SDEs is beyond the scope of the present work and readers are referred to [14] for further information.



**Figure 1.** Snapshots of the LES data for the stream-wise  $u$  velocity field: (a) in the  $yz$ -plane at the location of the first turbine; (b) in the  $yz$ -plane at the location of the second turbine; (c) in the  $xy$ -plane at hub height and  $t = 1$  day; and (d) in the  $xy$ -plane at hub height and  $t = 1.5$  day. To highlight effects of atmospheric boundary layer height on wake characteristics, results in this figure are performed under stable boundary conditions.

## 4. Results

In this section, we first show LES model results and then apply the POD approach to the LES  $xy$ -data at hub height (i.e. reducing the three-dimensional problem into two-dimension at a constant height of  $z = 104$  m). The two stochastic schemes explained in Section 3.2 are then applied in order to construct probabilistic models for  $a_j$  over all mode numbers. Finally, we show tentative results on how to use the SDE-based approach for, a short-term, prediction of the stream-wise velocity field  $u$ .

### 4.1. LES model results

Figure 1 shows instantaneous wake structures acting on turbines in the vertical  $yz$  (at the locations of the first and the second turbines) and the horizontal  $xy$  (at hub height) sections under stable condition. Figures 1-c and d show non-symmetrical behaviours of the flow field in the wake area, particularly for the second turbine. We use LES-ADM-R implementations to obtain the ROMs that are able to retain the important statistical features of the main flow and the wake fields. The shown results in this figure are associated with the simulations under stable boundary condition in order to show the effects of atmospheric boundary layer height variability on developments of wake. We conduct hereafter all simulations under neutral condition.

### 4.2. POD projection

Figure 2 shows histograms calculated directly from  $a_j$  data at the  $j$ th POD mode. Applying the best fit functions to the data (continuous red lines) suggests that the time-dependent weight coefficients are normally distributed for these modes. The normality test of  $a_j$  for all modes, based on values of skewness and kurtosis, confirms further the fact that the weighting coefficients follow a normal distribution (skewness and kurtosis are approximately around 0 and 3 for all  $j$  at sampling rate of 600 s, not shown).

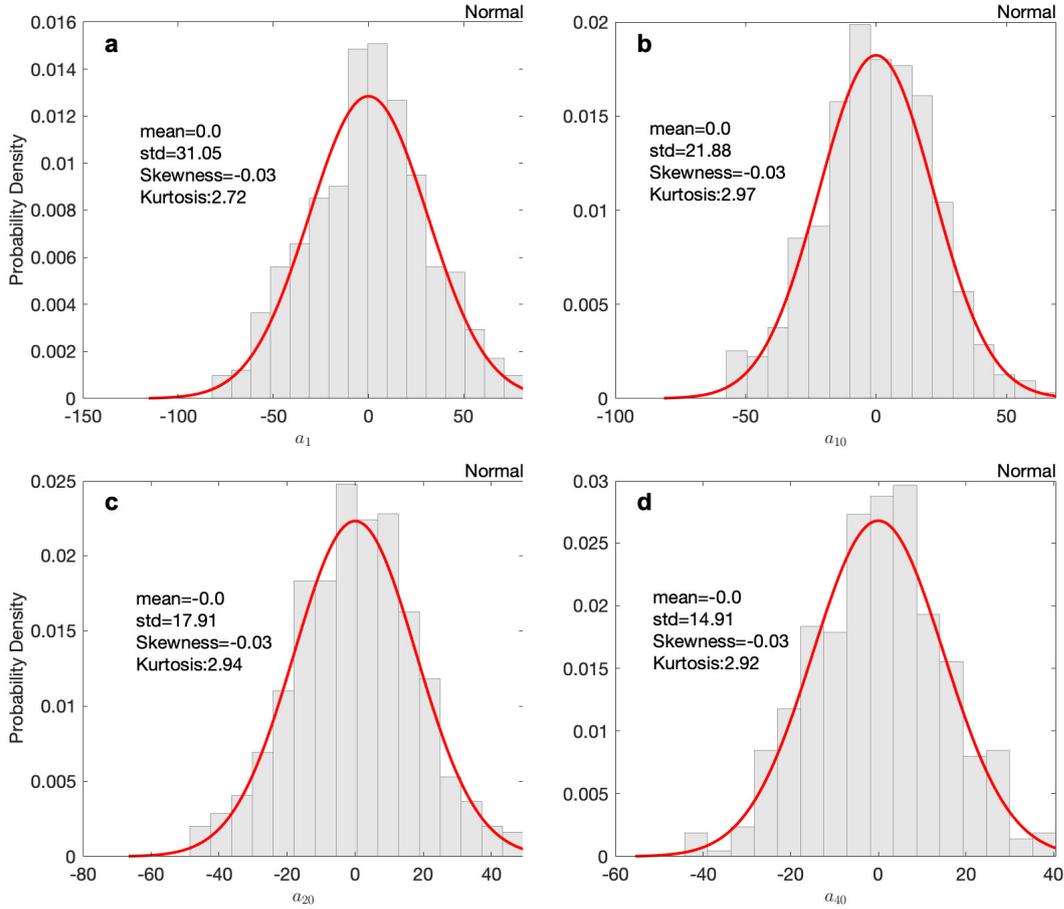
Figures 3-a and b show two modes of POD representing spatially coherent features in the wake areas. The second-ranked mode associated with the most energetic structures in the wake and the mode number 10 shows correlated features with the energetic modes in the wake. Figure 3-c illustrates energy spectrum of  $a_j(t)$  for 4 different POD modes calculated from all input LES snapshots. The first mode of decomposition captures energetic structures in  $a_j(t)$  (large-scale motion) for  $\geq 99\%$  of total energy of all LES snapshots. It is observed that the energy of fluctuations in the energy spectra declines by increasing  $j$ . Energy content of each POD mode for the turbulent flow can be further determined through the POD-based eigenvalue and its cumulative sum, Fig. 3-d. According to the truncated cumulative energy,  $\sum_{j=1}^n \lambda_j / \sum_{j=1}^{n_t} \lambda_j$ , around 60% of the kinetic energy of the velocity field is recovered by the first 50 modes while the energy of the first 200 modes represents around 80% of the energetic features. Here,  $n = 1, \dots, n_t$ .

We reconstruct the  $u$ -component of the velocity field using the POD modes calculated according to Eq. (1) with varying number of modes  $N$ , see Fig. 4. It is obvious that a more accurate description of the flow field is achieved by increasing the number of modes, the number of LES snapshots, and the model resolution that enables the method to recover small-scale structures particularly in the wake areas.

### 4.3. Probabilistic Models

Figures 5-a and b illustrate the Squared Exponential (SE) covariance kernel of GP defined by:

$$k(t, t') = \sigma_{ker} \exp \left( -\frac{(t - t')^2}{\xi^2} \right), \quad (12)$$

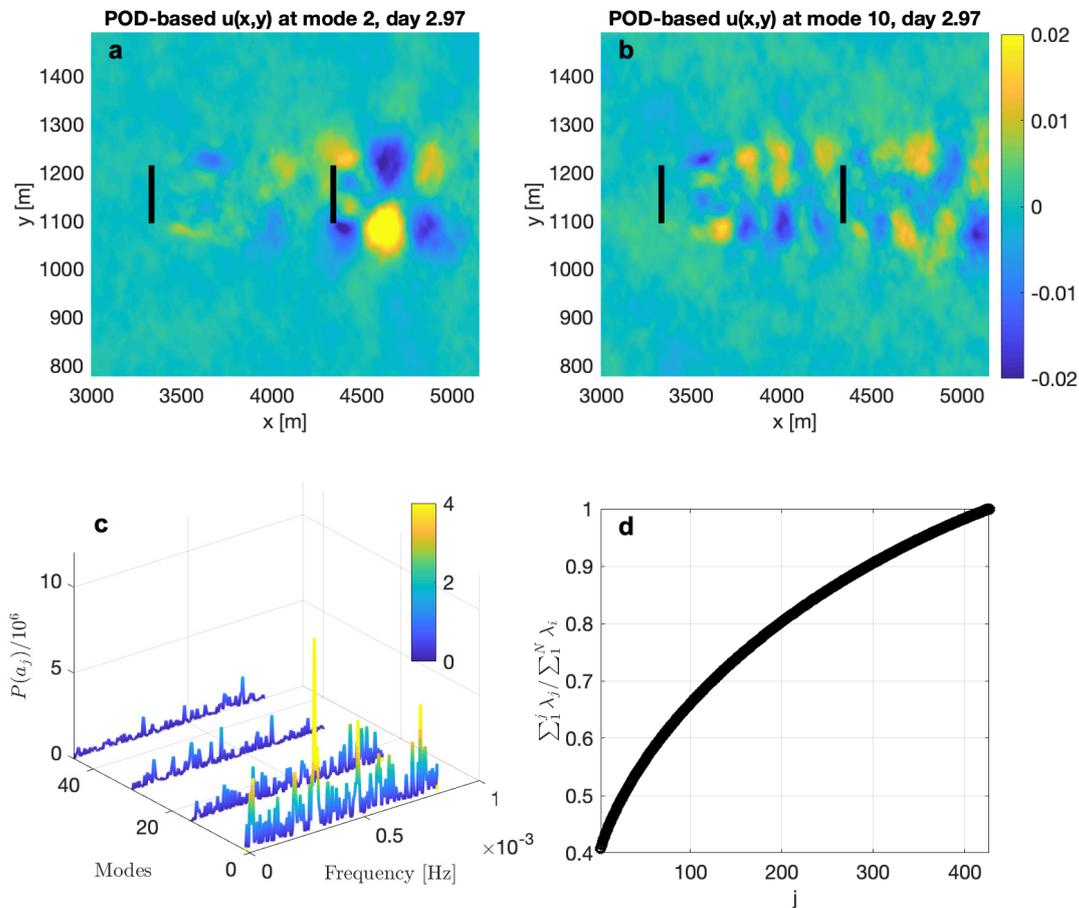


**Figure 2.** Comparison between histograms of the time-dependent coefficients  $a_j$  in Eq. (1) at (a) first mode,  $j = 1$ ; (b)  $j = 10$ ; (c)  $j = 20$ ; and (d)  $j = 40$ . Plots contain also the best-fitted distributions (red curves).

where the input pair  $(t, t')$  denotes the positions of data points and the (positive definite) kernel  $k(t, t')$  maps the input pair to a real value, representing the distance between data points. The shape of the kernel depends further on a couple of hyper-parameters, i.e.  $\sigma_{ker}$  and  $\xi$ . Here  $\sigma_{ker}$  and  $\xi$  control the normalisation and width of the kernel, respectively. In Fig. 5-a and b, the covariance matrices are shown for  $\sigma_{ker} = 1$  and two different widths. Figure 5-c shows a number of GP-based realisations (with zero mean) corresponding to each of the covariance matrices. Another example is a periodic kernel that allows the kernel function to repeat itself. It is expressed as follows:

$$k(t, t') = \sigma_{ker} \exp\left(-\frac{2 \sin^2(|t - t'|/\chi)}{\xi^2}\right), \quad (13)$$

where  $\chi$  determines the distance between sequential repetitions (see Fig. 5-d and e). Length-scale  $\xi$  and normalization factor  $\sigma_{ker}$  are defined similar to the SE kernel. This kernel is properly applicable for the data with cyclic nature. Fig. 5-f illustrates a number of GP realizations generated by a periodic kernel for given values of hyper-parameters [6].

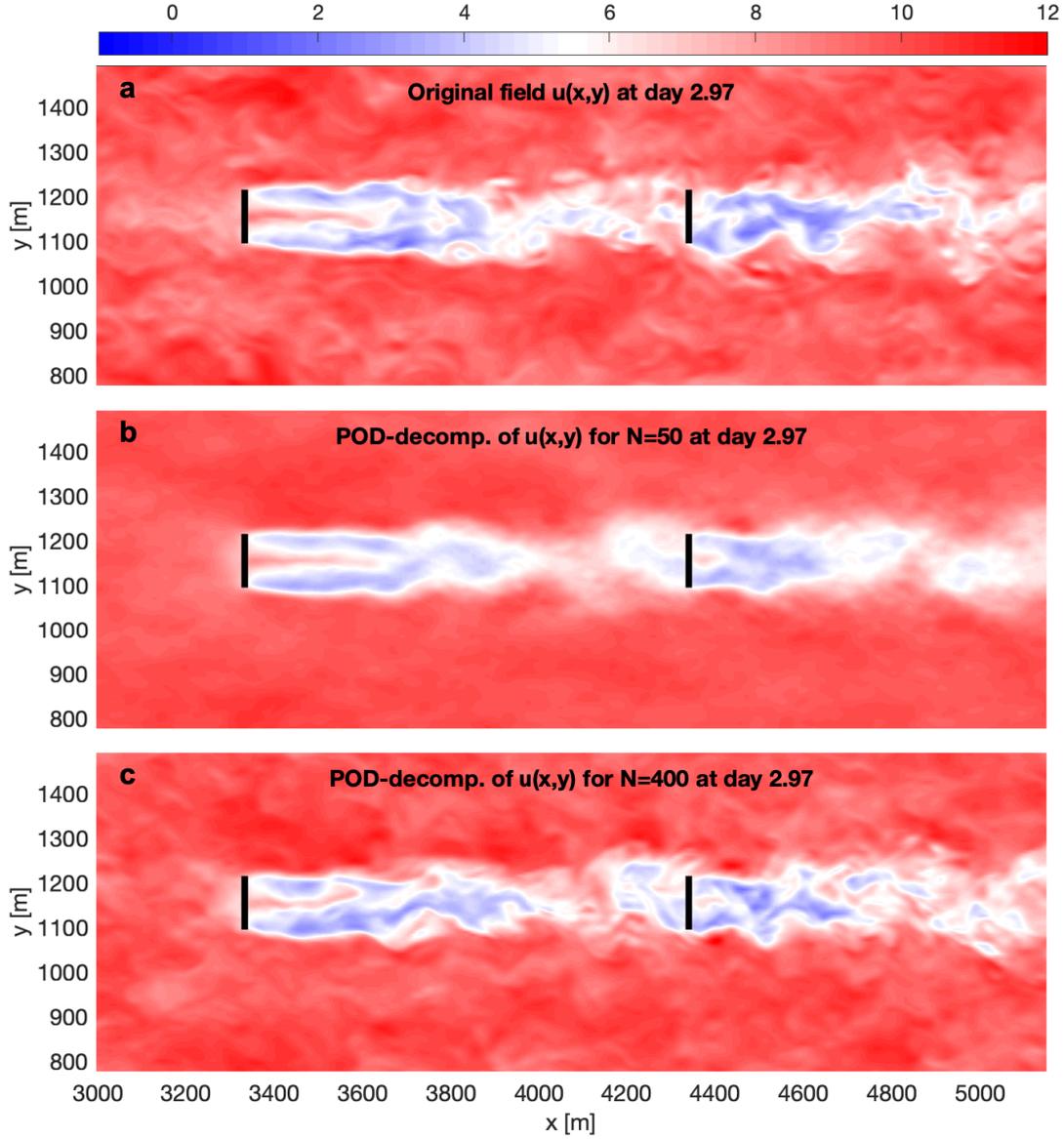


**Figure 3.** (a,b) POD spatial basis functions for mode number of 2 and 10, respectively; (c) Energy power spectra of  $a_j$  for  $j = 1, 15, 30, 45$ ; and (d) cumulative energy of POD modes.

To check the ability of GP in reproducing weighting coefficients, we compare between time series of  $a_j$  for  $j = 1$ , derived from the LES data, and the one estimated from the GP with SE kernel in Fig. 5-g. It is shown that two time series generally share similar behaviour, except slightly faster fluctuations for the GP one.

Now, we aim to construct time series of the second stochastic model by solving Eq. (10) through a numerical integration scheme. Before discussing the numerical solver, it is important to point out that Eq. (10) (i.e. Ornstein–Uhlenbeck process) is defined by the mean, standard deviation, and autocorrelation function of the time series. Figure 6-a shows a day-long time series of  $u$ -velocity for a point located in the free-stream flow. Autocorrelation of corresponding data (red markers) is shown against the exponential fit (black line) using the expression given by Eq. (11), to estimate the decay rate  $\alpha_j$ . The exponential function in Fig. 6-b provides a good fit to the observed autocorrelation data. Note that the convergence of numerical stochastic model depends directly on how well the observed autocorrelation data can be fitted by an exponential-decaying curve.

We now apply explicit stochastic Euler integration to solve this problem [16, 14]. Explicit

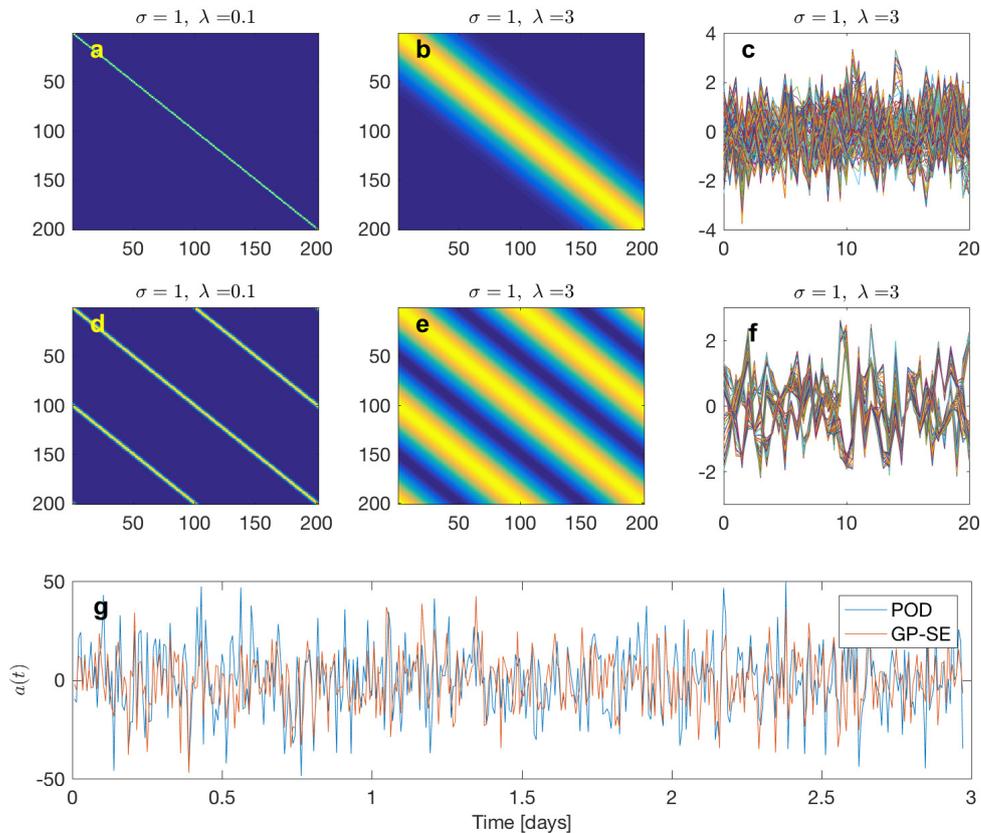


**Figure 4.** The velocity field  $u$  at day 3: (a) the original field; (b) the reconstructed flow field using  $N = 50$  modes; and (c) the reconstructed flow field using  $N = 400$  modes.

Euler scheme is written as

$$\hat{a}_j^{n+1} = \hat{a}_j^n + f(t_j, \hat{a}_j^n)dt + g(t_j, \hat{a}_j^n)\Delta W_j^n, \quad (14)$$

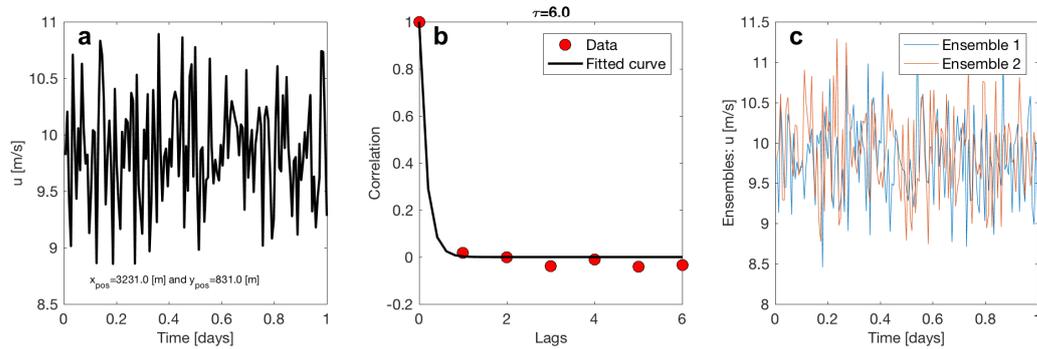
where  $\hat{a}_j^n$  denotes the numerical solution for  $a_j(t_n)$  at  $j$ th POD mode at discrete time  $t_n$ ,  $dt = t_{n+1} - t_n$  is the length of time discretization.  $\Delta W_j = W_j^{n+1} - W_j^n \sim N(0, \sqrt{dt})$  denotes the random increments of a Brownian process [16]. Wiener process is approximated using the MATLAB pseudo-random generator tool. For the initial values of  $\hat{a}_j^n$ , we use the POD-based  $a_j(t_0)$ , this will increase the rate of convergence. Figure 6-c shows two realisations of stochastic



**Figure 5.** (a,b) Numerical covariance matrices for  $\sigma_{ker} = 1$ , and  $\xi = 0.1, 3$  respectively; (c) a number of GP-based realisations for  $\sigma_{ker} = 1$  and  $\xi = 3$ ; (d,e) the same as (a,b) except using periodic kernel; (f) a number of GP-based realizations with a periodic kernel for  $\sigma_{ker} = 1$  and  $\xi = 3$ . For all periodic kernel realizations, we set  $\chi = 1$ ; and (g) comparison of POD based time series of  $a_j$  and the one generated from the SDE process. Colour maps in (a,b,d,e) indicate higher correlations by warmer colors.

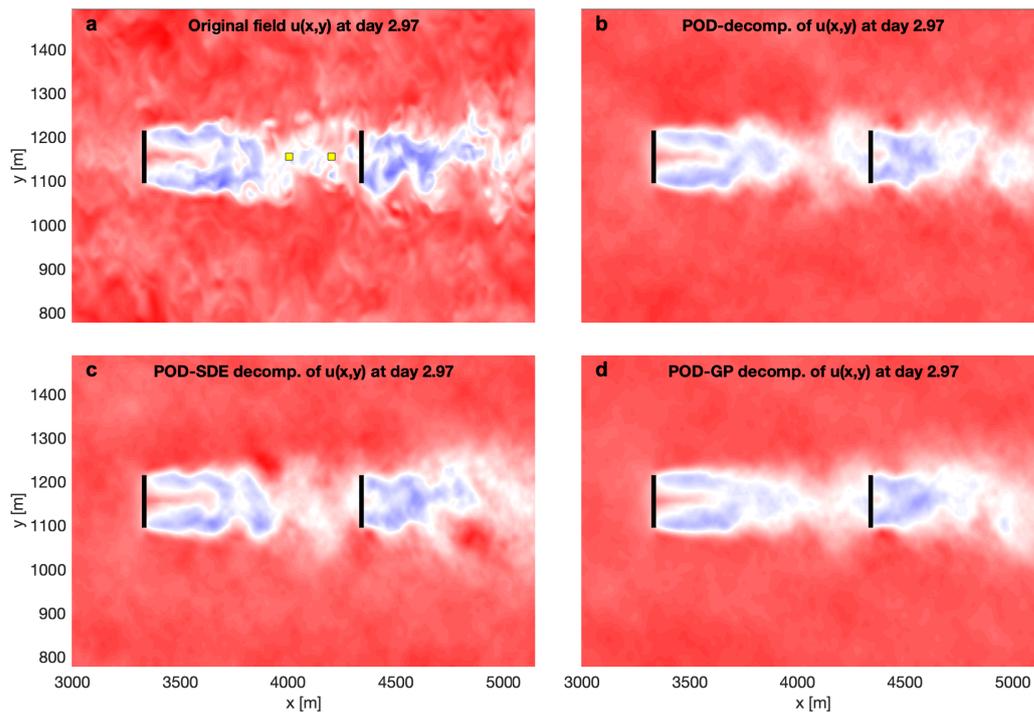
trajectories generated by solving Eq. (14) with the same initial values for the velocity time series as shown in Fig. 6-a.

Figure 7 shows the stream-wise velocity snapshot from LES (as a reference) to check the performance of different truncated POD-based velocity reconstructions for the fixed value of  $N = 50$  at day 2.97. The flow fields just behind each turbine have been successfully recovered by all methods, highlighting the ability of decomposition in capturing large eddies particularly in the near-wake areas (i.e. more energetic modes). However, in far-wake zones, the interactions between wakes and ambient flow cause more discrepancies between the full-scale model and the ROMs results. Figures 7-b and c show that two stochastic models are not significantly different, the SDE-based model provides, however, a slightly similar spatial distribution of energy over variety of scales compared with the reference distribution. Visual inspections suggest that small scale features are recovered better in the SDE-based model than the GP-based models, in both the near- and the far-wake areas. The discrepancies of the two stochastic models with respect to the original field can be explained partly by the fact that the temporal stochastic models cannot capture the nonlinear dynamical mode-mode interactions, and we have to account these



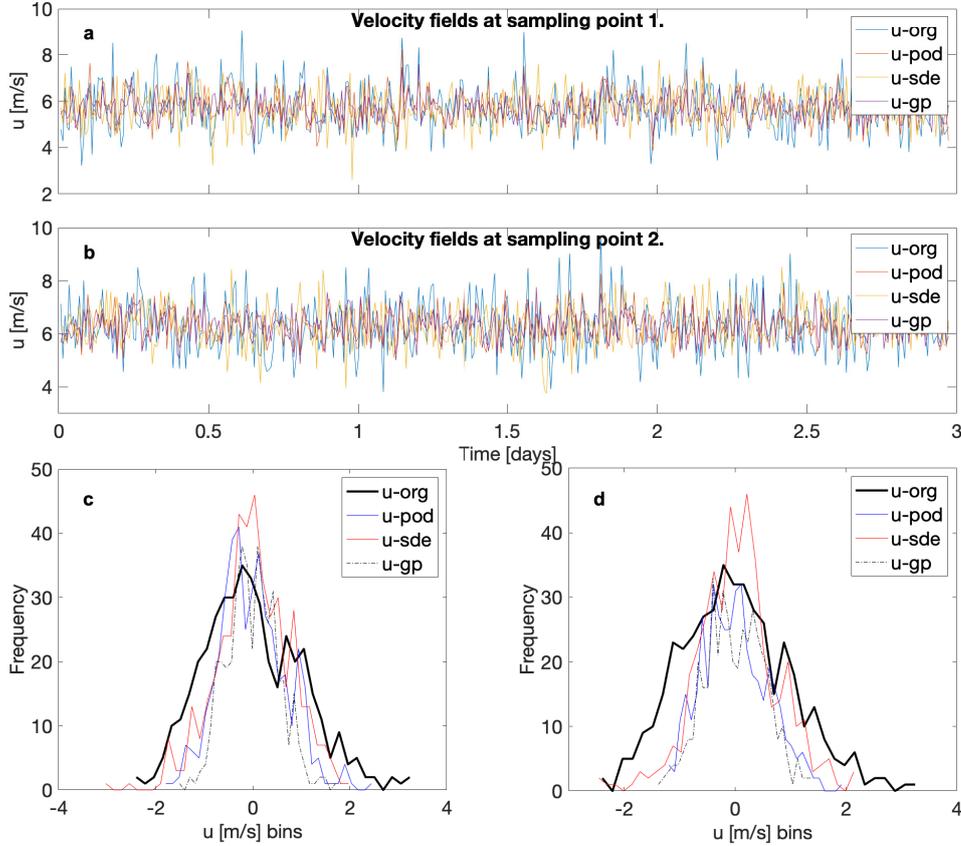
**Figure 6.** (a) Wind time series at point  $(x_{pos}, y_{pos}) = (3231, 831)$  m from LES; (b) autocorrelation of the corresponding data (red markers) and fitted exponential curve (black line), i.e. Eq. (11); and (c) two SDE-based realization.

interactions through applying, for example, Galerkin-POD scheme [5].



**Figure 7.** (a) Original stream-wise velocity  $u$  directly from LES at day  $t = 2.97$ . Two yellow markers indicate two locations for extracting time series used in the coming examples; (b) reconstructed  $u$  by the standard POD scheme; (c) reconstructed  $u$  from the SDE-based model; and (d) reconstructed  $u$  from the GP-based model.

Figures 8-a and b show time series of stream-wise wind sampled at two different, spatially separated points, as shown by markers in Fig. 7-a. In this figure, we aim to show whether the stochastic models preserve the statistical properties of the main time series by the means



**Figure 8.** (a,b) time series of stream-wise velocity fields estimated from different models at two positions shown by markers in Fig. 7-a; (c) histograms of corresponding time series at position 1; and (d) histograms of corresponding time series at position 2.

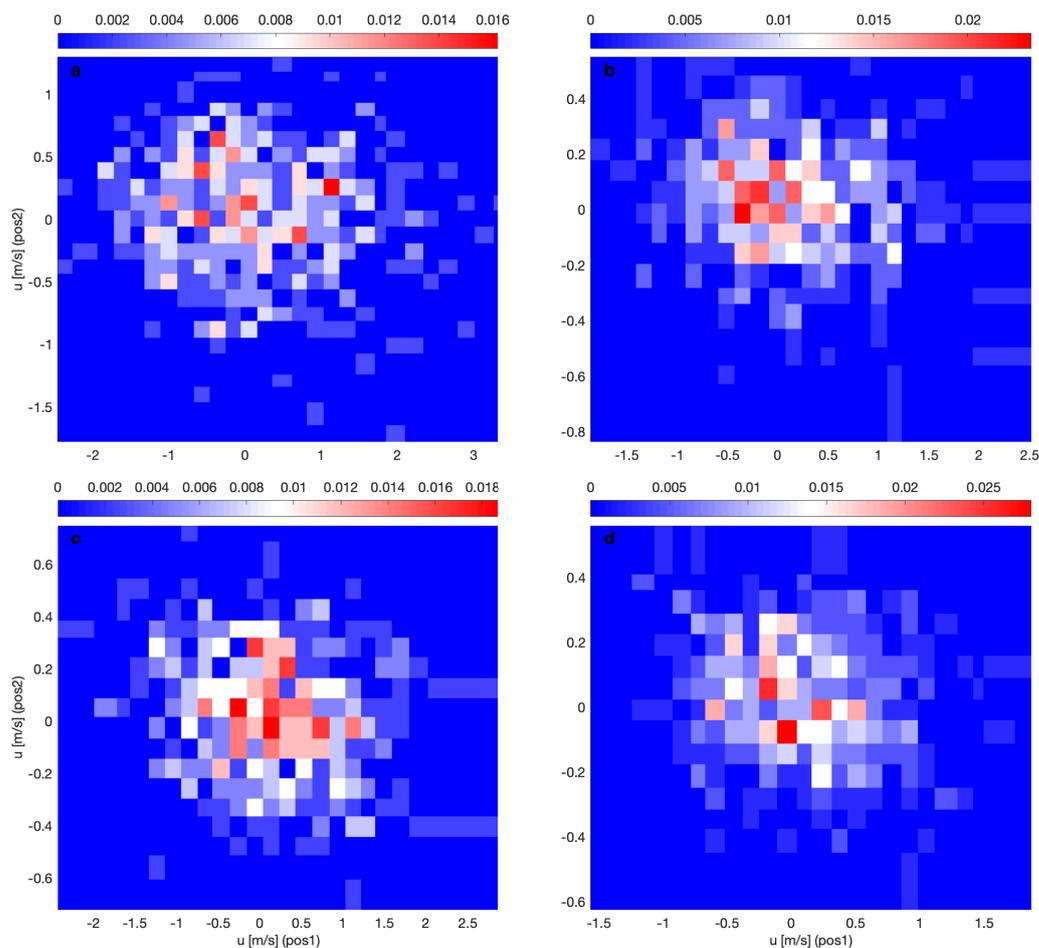
**Table 1.** Results of the PDF-fitting for two time series. For data from both locations, all distributions are positively skewed, since  $\tilde{b} < 3.6$  except for the position 2 where  $\tilde{b} > 3.6$  for POD-based  $u$ . In this case, the Weibull distribution becomes more Gaussian.

Model	Position 1		Position 2	
	$\tilde{a}$	$\tilde{b}$	$\tilde{a}$	$\tilde{b}$
LES	13.6	1.0	13.1	0.9
POD	12.0	0.8	13.4	10.1
POD-SDE	10.5	0.7	12.0	0.8
POD-GP	11.9	0.8	12.1	0.9

of statistical distributions. The histograms of the corresponding time series in Fig. 8-a and b suggest that Weibull functions have been properly fitted to the observed histograms at these stations. PDF of the Weibull distribution for the  $u$  component is given by

$$W(u; \tilde{b}; \tilde{a}) = \frac{\tilde{a}}{\tilde{b}} \left( \frac{u}{\tilde{a}} \right)^{\tilde{b}-1} e^{-\left(\frac{u}{\tilde{a}}\right)}, \quad (15)$$

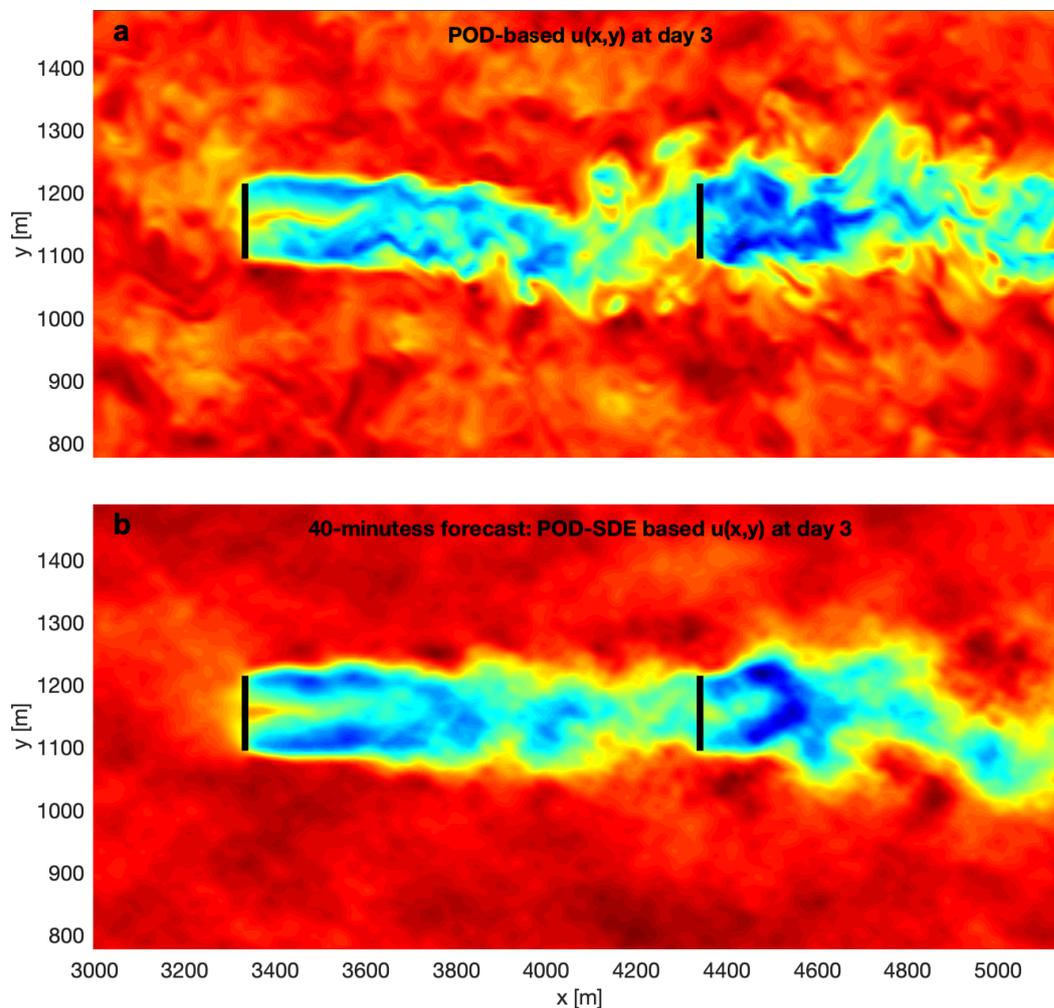
where  $\tilde{a}$  and  $\tilde{b}$  indicate the scaling and the shape parameters of the Weibull distribution. The Weibull curves capture both the peak and the sloping edge of the observed histograms (not shown). However, the fitted Weibull functions in Figs. 8-c and d offer different statistics, e.g. various skewnesses and kurtosis, compared to the original LES stream-wise velocity distribution (see Table 1). The discrepancies are strongly enhanced in the inflow areas of the waked turbine (i.e. position 2). In this case, two stochastic models can capture peaks while overestimating the sloping edge compared to the reference histogram (black line). We observed also larger departure between the shape of the POD-based histogram and the reference histogram in this case.



**Figure 9.** Illustration of JPDFs calculated from two points as marked in Fig. 7-a for: (a) original LES data; (b) standard POD model; (c) SDE-based POD; and (d) GP-based POD. Number of bins for all samples is set to 32.

In Fig. 9, the stream-wise wind velocity at hub height is used to form a Joint Probability Density Function (JPDF) for wind time series at two selected positions. The reason for examining the JPDF is to focus: (i) on the performance of models in recovering two-point statistics of stream-wise velocity; and (ii) checking how two separated points in space are correlated across the scales of interest in the far-field area of the first turbine. Figure 9-a and

b show the two-point JPDF of the original high-fidelity data (as reference) and the POD-based results, respectively. In Fig. 9-a, the smaller variations in  $u$  values at position 1 correspond somewhat to the larger variation of  $u$  values at position 2. It is observed graphically that the reference JPDF reveals slightly negative (but scattered) correlation for these two-point  $u - u$  relationship. All POD-based models retain the same characteristics. More specifically, the negative correlation is clearly pronounced in the GP-based plot (i.e. Fig. 9-d) and the SDE-based JPDF (Fig. 9-c) compares better, in terms of spatial distribution, with the ground truth JPDF (see the velocity axes in this figure). While the slopes of fitting lines to JPDF for all models are in good agreement, they are steeper than the one related to the reference JPDF (not shown). It is important to mention that the two-point-correlation comparisons as carried out here depend fully on the location of the two points in the study domain.



**Figure 10.** (a) Original stream-wise velocity  $u$  from LES at day 3; and (b) the corresponding predicted field at day 3 produced by the SDE-based scheme has been run between day 2.97 and day 3, i.e. 40 minutes forecast.

#### 4.4. Short-term forecast

So far we have interpolated the stream-wise velocity field for a period of  $[0, 2.79]$  days. In this section, we provide a tentative result associated with the short-term (i.e. 40 minutes) forecast ability of the SDE-based model. Here, the stochastic trajectories are generated for the period between  $[2.97, 3]$  days from the initial values at  $t = 2.97$  day. Figure 10 compares between the original LES  $u$  velocity at day 3 and the one predicted by applying the SDE-based model, i.e. 40 minutes forecast between day 2.97 and day 3. It is observed that the model is reproducing effectively large scales of motions, particularly in the near-field areas. Uncertainty of the forecast increases by transition from near-wake to far-wake areas. We postpone further quantitative investigations on predictive characteristics of developed stochastic models in another independent work elsewhere.

## 5. Conclusion

In this study, we used LES snapshots under neutral atmospheric boundary condition, that incorporate the effects of wind turbines (using ADM-R), to develop low-order models based on a truncated POD approach. The standard POD model decomposes the LES snapshots in a series of orthogonal spatial basis functions modulated by time-dependent (deterministic) weighting coefficients. Two different probabilistic models were utilized to replace the deterministic coefficients in the standard POD model by stochastic (time-dependent) weighting coefficients. The reconstructed wake fields by applying deterministic and stochastic coefficients showed the ability of stochastic models to capture some of the missing small-scale features that could not be captured by the use of the standard POD method, for a given number of modes. We further checked tentatively the predictive ability of one of the stochastic approaches.

Future work includes to account for the effects of mode-mode nonlinear interactions and applying the method for the wind farm control. In addition, the results of developed ROMs will be validated carefully against the high-fidelity simulations and available field data.

#### 5.1. Acknowledgments

This study has awarded funding from the academic agreement between UiB and Equinor. The simulations were performed on resources provided by UNINETT Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway.

## References

- [1] Barthelmie R J, Rathmann O, Frandsen S T, Hansen K, Politis E, Prospathopoulos J, Rados K, Cabezón D, Schlez W, Phillips J, et al., 2009 Modelling and measuring flow and wind turbine wakes in large wind farms offshore. *Wind Energy*, **12**, 431-444.
- [2] Ferrer E, Browne O M F, Valero E, 2017 Sensitivity Analysis to Control the Far-Wake Unsteadiness behind Turbines". *Energies*, **10**, 1599.
- [3] Peña, A, Rathmann O, 2014 Atmospheric stability-dependent infinite wind-farm models and the wake-decay coefficient. *Wind Energy*, **17**, 1269–1285.
- [4] Hasager C B, Rasmussen L, Peña A, Jensen L E, Réthoré P E, 2013 Wind farm wake: The Horns Rev photo case. *Energies*, **6**, 696–716.
- [5] Hamilton N, Viggiano B, Calaf M, Tutkun M, Cal R B, 2017 A generalized framework for reduced-order modeling of a wind turbine wake. *Wind Energy*, DOI: 10.1002/we.2167.
- [6] Rasmussen C, Williams C, 2005 Gaussian Processes for Machine Learning. *MIT Press: Publisher*, ISBN 026218253X.
- [7] Bastine D, Vollmer L, Wächter M, and Peinke J, 2018 Stochastic Wake Modelling Based on POD Analysis. *Energies*, **11**, 612, DOI:10.3390/en11030612.
- [8] Katic I, Højstrup J, Jensen N O, 1986 A simple model for cluster efficiency. In Proceedings of the *European Wind Energy Association Conference & Exhibition*, Rome, Italy, 7–9 October, pp. 407–410.
- [9] Dörenkämper M, Witha B, Steinfeld G, Heinemann D, Kühn M, 2015 The impact of stable atmospheric boundary layers on wind-turbine wakes within offshore wind farms. *Journal of Wind Engineering and Industrial Aerodynamics*, **144**, 146-153.

- [10] Mortensen N G, Heathfield D N, Myllerup L, Landberg L, Rathmann O, 2007 Getting Started with WASP 9, Technical Report Risø-I-2571(EN). *Risø National Laboratory*, Roskilde, Denmark.
- [11] Tennekes H and Lumley J L, 1972 A First Course in Turbulence. *The MIT Press*.
- [12] Jonkman J M, Butterfield S, Musial W, and Scott G, 2009 Definition of a 5-MW reference wind turbine for offshore system development. Technical Report NREL/TP-500-38060, *National Renewable Energy Laboratory*, 1617 Cole Boulevard, Golden, Colorado 80401- 3393, February.
- [13] Maronga B, Gryschka M, Heinze R, Hoffmann F, Kanani-Sühring F, Keck M, Ketelsen K, Letzel M O, Sühring M, and Raasch S, 2015 The Parallelized Large-Eddy Simulation Model (PALM) version 4.0 for atmospheric and oceanic flows: model formulation, recent developments, and future perspectives. *Geoscientific Model Development*, **8(8)**, 2515– 2551.
- [14] Bakhoday-Paskyabi M, 2017 Turbulence-particle interactions under surface gravity waves. *Ocean Dynamics*, **67(3-4)**, 557-557.
- [15] Grigoriu M, 2002 Stochastic Calculus. *Applications in Science and Engineering*, Boston, MA: Birkhauser.
- [16] Kloeden E and Platen E, 1999 Numerical Solution of Stochastic Differential Equations. New York, NY, 3rd edition: *Springer*.
- [17] Barndorff-Nielsen O E and Shephard N, 2001 Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics (with discussion). *Journal of the Royal Statistical Society, Series B*, **63**, 167–241.