Energy 238 (2022) 121723

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Dynamic wind turbine wake reconstruction: A Koopman-linear flow estimator

Zhenyu Chen [a, b], Zhongwei Lin [a, *], Xiaoya Zhai [c, d], Jizhen Liu [a]

[a] State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, School of Control and Computer Engineering, North China Electric Power University, Beijing, PR China
[b] Delft Center for Systems and Control (DCSC), Faculty of Mechanical, Maritime and Materials Engineering (3mE), Delft University of Technology, the Netherlands
[c] School of Mathematical Sciences, University of Science and Technology of China, Hefei, PR China
[d] Department of Design Engineering, Delft University of Technology, the Netherlands

ARTICLE INFO

Article history:
Received 30 January 2021
Received in revised form
21 July 2021
Accepted 5 August 2021
Available online 16 August 2021

Keywords:
Wake effect
Wind turbine
Flow reconstruction
Koopman operator
Extended dynamic mode decomposition

ABSTRACT

A challenging topic arising in dynamic wind turbine wake is modeling, especially the low-order approximation. The central problem is the fact that it has high-dimensional and nonlinear wake characteristics. In this paper, a Koopman-linear flow estimator is designed according to the Koopman operator theory. Different from the conventional flow reconstruction with the linear stochastic estimation method, a dynamic state-space model with physical states is constructed. The wake dynamics are approximated using a limited number of measurable physical parameters by the dynamic part; then, the full wake flow is reconstructed from the low-order states by the estimation part. The flow estimator is designed into three different forms following Extended Dynamic Mode Decomposition (EDMD) method. Each form has its unique advantages. Precisely, probe sensors are placed in the studied space and provide direct information of the wake, and a few in-directly physical parameters are also included. Nonlinear integer programming is further adopted using a heuristic optimization algorithm, by which the sensor configurations are optimized. Comparisons with the standard Dynamic Mode Decomposition (DMD)-based wake model are adopted in time domain and frequency domain to verify the effectiveness of the proposed flow estimators. The results show acceptable accuracy in typical modeling cases and maintain good estimation accuracy when the measurement noises are involved. Finally, the proposed Koopman-linear flow estimator is compared with related stochastic estimation methods, in which the connections of the proposed estimator with stochastic ones are also discussed.

1. Introduction

As interests towards wind farm research grow, there has been an increasing focus on wake effects on wind farm power capture or structural loads [1]. The steady-state wake model, like *FLORIS* [2] and Gaussian-based model [3,4], was proposed and considered during wind farm design and operations to achieve better power utilization and conversion [5]. In its continuation, the dynamic characteristics of wind turbine wake are gradually becoming the focus of current research [6]. The dynamic wind is usually modeled by the Navier-Stokes equations (NS-equations), which characterize the fluids dynamics from the mechanism level and reach high modeling fidelity [7]. However, the computation cost of the NS-equation-based wake model is huge. Therefore, it is usually regarded as a simulation tool rather than a modeling technics [8]. The dynamic wake model herein is expected to characterize the wake's dynamic behavior while still retaining low computational requirements. In this area, several models are introduced from different aspects. Typical examples are steady-state wake model-based FLORIDyn model [9], Dynamic Wake Meandering (DWM) model [10], and reduced-order model (ROM) [11]. Among these reviewed dynamic wake models, the data-driven linear reduced-order model is the primary consideration in this paper.

Proper Orthogonal Decomposition (POD) and Dynamic Mode Decomposition (DMD) methods are usually adopted to construct the dynamic wake ROMs. In Ref. [12], the wind turbine wake is decomposed using finite numbers of POD modes and then

* Corresponding author.
E-mail address: lzw@ncepu.edu.cn (Z. Lin).

https://doi.org/10.1016/j.energy.2021.121723
0360-5442/

reconstructed with good reconstruction accuracy. Furthermore, in Ref. [13], a reduced-order wake model is proposed using POD. After the dominant POD modes are identified, a dynamic model is achieved through a polynomial expansion of the POD coefficient permutations. Though a reduced-order model could be constructed using POD, the establishment of dynamic models to describe fluid dynamics is more like the focus of DMD than POD [14]. A DMD-derived dynamic wake ROM is introduced in Ref. [11] and several typical DMD modes are visualized. Using a few stable DMD modes, the DMD-derived ROM linearly approximates the wake dynamics and predicts the wake flow in the upcoming time. Usually, the states in POD/DMD-derived wake model are calculated from the original measurement that an order-reducing linear matrix maps the measured state to a low-dimensional subspace [11]. After prediction, an order-ascending mapping reconstruct the full-wake flow from the reduced-order states. The POD/DMD-derived wake ROMs have inspired us for a new question: is it possible that the wind turbine wake can be estimated and predicted through a linear model with physical states? The states are expected to be measurable-physical parameters in the wake flow, which benefits for industrial applications and further improvements like the Kalman filter. Based on the measured physical states, the wind turbine wake can be reconstructed and predicted.

In the physical measurement-based flow reconstruction field, achievements have been accomplished. In 1988, the Linear Stochastic Estimation (LSE) method was first introduced in Ref. [15], in which velocity and the deformation tensor of conditional homogeneous-shear flow are approximated by two-point spatial-correlation tensor using linear mean-square stochastic estimation. In this work, the LSE method is first proposed to approximate the conditional eddies with given local kinematics, and then further applied under different events to show good estimation results. Like the mapping from reduced-order states to full-flow field in DMD, LSE constructs a pre-trained linear mapping from the direct-measured flow-related information to the full-flow field. Furthermore, in Ref. [16], the flow around the airfoil is also studied using LSE: sensors are placed on or around the airfoil to provide direct-measured information as probes, and LSE reconstructs the steady flow based on the probes-measured data. The main advantages of LSE method are that: the reconstruction is based on physical measurements. The to-be-reconstructed state space is assumed to have a linear relationship with the measurement space, and LSE calculates a linear approximation of the unknown linear relationship. However, the LSE-based flow reconstruction is hardly regarded as a *model* since the fluid dynamics are not considered. The dynamic wake ROMs are expected to capture the flow field dynamic performance, at least part of the dynamics.

Although the reviewed methods achieve good accuracy and have been widely used, whether a linear approximation model may carry advantages of both LSE and DMD-derived models has rarely been examined directly. A desired dynamic wake ROM is expected to model the wake dynamics while remaining measurable-physical states. This paper introduces a flow estimator following the Koopman operator theory and Extended Dynamic Mode Decomposition (EDMD) method [17]. Koopman operator theory is an alternative formalism for dynamical system theory that provides a linear operator to characterize nonlinear and high-dimensional systems. It has been applied in the flow modeling and control area recently [18]. DMD and EDMD solve finite-dimensional approximations of the infinite-dimensional Koopman operator [19]. In this way, the nonlinear dynamics can be easily analyzed and may be controlled [20,35]. The Koopman operator and EDMD method allow a scalable reconstruction of the underlying dynamical system from measurement data while remaining in linear form. Such property is partly used in Ref. [21] to help design the MPC controller, which is their main consideration, but the reconstruction using Koopman operator theory has not been fully verified.

This paper orients to fill the gap between DMD and LSE while broadening the Koopman operator theory-based flow reconstruction in Ref. [21]. The flow estimator is designed to linearly approximate the wind turbine wake dynamics using physical and measurable states while reconstructing the full wake flow. A total of three forms of flow estimators are designed, in which all states are physical. Each form has unique advantages. Furthermore, to optimize the reconstruction accuracy of developed estimator, sensor configuration optimization is also carried out using a heuristic algorithm. Finally, the flow estimator is analyzed for both time- and frequency-domain characteristics and compared with DMD and LSE-related methods to verify the effectiveness.

The rest of this paper is organized as follows. In Section 2, the Koopman operator theory and EDMD procedure are briefly introduced, then the flow estimator is designed into three different forms following the reviewed approaches. The studied wind turbine wake data is acquired using large eddy simulation, which is presented in Section 3. The probe sensors are placed in the wake zone using an nonlinear integer-programming method, which is also introduced. Three form flow estimators are compared with the DMD-derived wake model in both time and frequency domain in Section 4. In Section 5, the prediction ability of the proposed flow estimator is verified, as well as the sensors-based consistency. Section 6 discusses the flow estimator with LSE-related methods through series comparisons. Section 7 concludes this paper.

2. The flow estimator

In this section, the theoretical basis is briefly introduced. The flow estimator is designed. Three different forms are introduced, and each form will be introduced in an independent subsection.

2.1. Theoretical basis

For a nonlinear dynamic system, the Koopman operator is a linear infinite-dimensional transformation defined for the system observable, fully capturing all the underlying system properties. It theoretically guarantees the feasibility of linear approximation to nonlinear systems. The Koopman operator is first defined for uncontrolled systems and then generalized to controlled systems. Detailed theoretical proof are referred to Refs. [17,18,22]. In practice, the DMD method is used to solve a finite-dimensional approximation to the Koopman modes, *i.e.*, the DMD modes. In fluid modeling, the standard DMD method provides a practical solution to approximate the high-dimensional fluid dynamics using lower-order state-space models. The DMD-derived model requires a full measurement of the fluid. The connections between DMD and the Koopman operator are referred to Refs. [23,24].

Although DMD has been widely used in the flow field, it is still limited under fairly restrictive conditions: direct measurement of the flow field is required. As an improved method, EDMD broadens the extensiveness of DMD that different observables are included. It is introduced by Williams in Ref. [19], which shows an alternative way to approximates the Koopman operator using a collection of observables. In this way, the Koopman operator $\mathcal{K}$ could be approximated with direct measurement of the full system and observables, such as polynomials, radial basis functions, or spectral elements. Both DMD and EDMD are finite-dimensional approximation methods to the infinite-dimensional Koopman operator. The DMD procedure may be viewed as a spectral collection method of approximating $\mathcal{K}$. In contrast, the EDMD approximated $\mathcal{K}$ is defined on a subspace spanned by a collection of observables with wider extensiveness. The EDMD procedure can be referred to

Refs. [18,21]. The EDMD method is a comprehensive solution to $\mathcal{K}$ than the DMD method, which could be viewed as a particular case of EDMD. The feasibility of both DMD and EDMD are guaranteed by the Koopman operator theory, which proves the nonlinear dynamics can be linearly approximated, see Ref. [25]. Detailed Koopman Operator theory, DMD, and EDMD procedures are given in the Appendix section. Connections between the Koopman operator and EDMD are analyzed in Refs. [17,22].

The Koopman operator theory, DMD, and EDMD methods provide a solid theoretical foundation for the desired flow estimator. Firstly, The designed flow estimator is not an exact Koopman operator. It could be regarded as a Koopman-like or Koopman operator-based finite-dimensional linear system, which approximates the dynamic flow using physically measured states and then reconstructed the full flow. The Koopman operator theory proves the feasibility, while the design procedure follows the EDMD method. Furthermore, as a linear estimator, the flow estimator's basic function is that the full-wake flow is *estimated* from low-order physical states. Thus it is referred to as a Koopman-linear flow estimator. The designed flow estimator does not require specific reconstruction objects; a typical choice could be the velocity or pressure in common flow studies. This work demonstrates the velocity field of wind turbine wake. In the following of this paper, the flow estimator aims to reconstruct the velocity field in a wind turbine wake region. Considering the measured or simulated flow data are usually given in girds, the *flow* here means the gridded velocity data in a given space, which could be two- or three-dimensional ones.

2.2. Basic form

In this basic form, several probe measurement sensors are assumed to exist in the studied space, by which the flow velocity could be measured at the placed locations. The flow estimator reconstructs the full flow based on the measured data. The basic form takes the state-space model as:

$$\begin{aligned} z^{+} &= Az \quad z\in\mathbb{R}^{n}, \\ \hat{x} &= Cz \quad x\in\mathbb{R}^{m}, \end{aligned} \tag{1}$$

in which z is the probes measured states, z^{+} is the time-shifting states with one time-step in advance, and $\hat{x}$ is the estimated x, the studied flow field. The desired state matrix A is with $n \times n$ dimension and the estimation matrix C is with $m \times n$ dimension. Each state represents the measured velocity of one probe. The state-space model explains the working mechanism of this basic form flow estimator: similarly as the EDMD flow model in Ref. [22], the first half in (1) describes a linear approximation of the flow dynamic using measurable states, while the second half estimates the full-flow field from states. Thus, the first half could be regarded as the predictor while the second half is the estimator. Since the states are physically measured from the flow field, the estimator matches with the LSE method that the full-flow field is reconstructed from samplings.

Assuming a total of n cells at k time step are selected from all measured cells as can-be-obtained measurements. The flow data in the studied space have m measured cells and are marked as X. The measured dataset is marked as X_{p}, and works as the lifted matrix similarly as the EDMD procedure in (C.3). Dimensions of datasets are $m \times k$ of X and $n \times k$ of X_{p}. Define

$$X_{\mathrm{lift}} = X_{\mathrm{p}} = g_1(x) = \mathrm{select}(X, n). \tag{2}$$

The select(X, n) operator represents a given or random selection from the dataset X, that n rows data are used as probes. The X^{+}_{lift} data is the time-shifting data of the lifted data X_{lift}, and have one time-step in advance (2 s in this study). The designed procedure matches the EDMD method that the selection operation is regarded as an observable of the studied flow field: a non-uniform random sampling, which captures some of the linear performance of the true flow from measured data. Instead of a collected observables defined in the EDMD procedure (C.2) and (C.3) in Appendix C, the selection/sampling is one observable and marked as $g_1(x)$.

A n-th SVD is performed as:

$$X_{\mathrm{p}} = U_{\mathrm{p}}\Sigma_{\mathrm{p}}V_{\mathrm{p}}^{T}, \tag{3}$$

in which the subscript p is used to identify the SVD result of X_{p} from SVD operations to different dataset. Since the SVD operation takes the same order as the state rank, it is rather a full-order decomposition than a reducing-order decomposition. U_{p} and Σ_{p} are $n \times n$ and V_{p} is $k \times n$. The A matrix is calculated following the standard DMD method, while the C matrix can be obtained by solving the following linear least-squares problem:

$$\min_{C}\left\|X - CX_{\mathrm{lift}}\right\|_{F}. \tag{4}$$

The corresponding solution is

$$\begin{aligned} A &= U_{\mathrm{p}}^{T}X^{+}_{\mathrm{lift}}V_{\mathrm{p}}\Sigma_{\mathrm{p}}^{-1}, \\ C &= XX^{\dagger}_{\mathrm{lift}}. \end{aligned} \tag{5}$$

where † indicates the Moore-Penrose pseudoinverse.

As can be seen, the dynamic part takes the standard DMD method to decompose the dynamic performance based on measured data. More details are referred to B. The estimation part in the basic form flow estimator has the same form as the linear statistic estimation.

2.3. Extended form

As an extended form flow estimator, more observables are included. Unlike the measured probe data as observable in the basic form, the added observable is identified as deterministic states. Usually, in the flow studies, some measurable parameters are also noticed. Other than the direct flow information like measured data, deterministic parameters could have different but important physical meanings. Take the wind turbine wake as an example. Multi-wind turbines are usually placed nearby and organized as a wind farm. The wake of the front-row wind turbines affects the downstream wind turbines, resulting in the loss of downstream wind speed, an increase in turbulence intensity, and a decrease in the captured power of the downstream wind turbines. In this case, the affected parameters like rotor speed or output power of the downstream wind turbines could be adopted as deterministic states.

With the extension of deterministic states, the lifted dataset in the extended form is defined as:

$$X_{\mathrm{lift}} = [g_1(x), g_2(x)] = [\mathrm{select}(X, n), X_{\mathrm{d}}], \tag{6}$$

in which X_{d} is the lift of deterministic parameters dataset. The X^{+}_{lift} set is also the time-shifting data of X_{lift}. Set the added dataset X_{d} with p parameters and $p \times k$ order, then X_{lift} and X^{+}_{lift} are with $(n + p) \times k$. Both sampling measurements and deterministic states, which are flow-related parameters and showing part of the flow performance in direct or indirect manners, are regarded as observables and included in states. The state-space model of the extended form is given as:

$$\begin{aligned} [z^+, z_\mathrm{d}^+]' &= A[z, z_\mathrm{d}]' \quad z \in \mathbb{R}^n,\ z_\mathrm{d} \in \mathbb{R}^p, \\ \hat{x} &= C[z, z_\mathrm{d}]' \quad \hat{x} \in \mathbb{R}^m, \end{aligned} \tag{7}$$

in which z_d is the added deterministic states. Strictly, the extended form can take the same state-space model as the basic form. Both probe measured states and deterministic states are observables to the flow and could be collected as (6), similarly as the EDMD procedure in (C.2). But the separate expression is beneficial to a clear understanding and sustains the further introduced ext-DMD form, therefore distinguished.

Compared with the linear stochastic estimation, the extended form introduces the deterministic states and the possibility that they can involve more observables. After the lifted dataset is obtained, state matrices are calculated following the linear least-squares solution in (C.6) or the alternative way in (C.7). The estimate matrix C is solved following the same solution (4), and the state matrix A is solved by:

$$\min_A \left\| X_\mathrm{lift}^+ - A X_\mathrm{lift} \right\|_F, \tag{8}$$

and the result is given as:

$$A = X_\mathrm{lift}^+ X_\mathrm{lift}^\dagger. \tag{9}$$

In terms of dimensions, A is with $(n+p) \times (n+p)$ and C is with $m \times (n+p)$.

2.4. Ext-DMD form

As introduced in Ref. [26], POD could be used to decompose the flow before reconstruction. Functionally, POD works as a filter that the dominant components are reserved, while the measurement noise and high-frequency nonlinear components are filtered. Before reconstruction, the POD/SVD also helps to reduce the computational cost and achieve a reduced-order model. Although the added deterministic states might share similar physical meaning as the probe-measured states, *e.g.*, the regionally averaged velocity, the SVD operation is still introduced herein to ensure the universality.

Firstly, both the full flow dataset and the probe-measured dataset are decomposed using a same order of SVD:

$$\begin{aligned} X &= U_\mathrm{x} \Sigma_\mathrm{x} V_\mathrm{x}^T, \\ X_\mathrm{p} &= U_\mathrm{p} \Sigma_\mathrm{p} V_\mathrm{p}^T. \end{aligned} \tag{10}$$

Usually, the SVD operation order is less than the flow order in the introduced DMD-related methods. It dues to the fact that a reduced-order model is expected. In the designed flow estimator, SVD servers more as full-order mapping than reduced-order mapping, by which the dominant linear components are reserved for further reconstruction. In this way, the nonlinear, especially high-frequency nonlinear component, could be filtered before reconstruction. To maintain consistency with the previous sections, the SVD operations are set to have the same order as the rank of X_p, resulting in the matrices with dimensions of: U_x is with $m \times n$, Σ_x is with $n \times n$, V_x is with $k \times n$. Matrices with subscript p take the same dimensions as introduced in Section 2.2. Instead of a reduced-order model, the model that fits the SVD states is regarded as a reconstructed model:

$$\begin{aligned} [z_\mathrm{rc}^+, z_\mathrm{d}^+]' &= F[z_\mathrm{rc}, z_\mathrm{d}]' \quad z_\mathrm{rc} \in \mathbb{R}^n,\ z_\mathrm{d} \in \mathbb{R}^p, \\ \hat{x_\mathrm{d}} &= H[z_\mathrm{rc}, z_\mathrm{d}]' \quad \hat{x_\mathrm{d}} \in \mathbb{R}^n. \end{aligned} \tag{11}$$

The reconstructed states have the following linear relation with true states through using the corresponding left-singular vectors:

$$\begin{aligned} z_\mathrm{rc} &= U_\mathrm{p}^T z \quad z \in \mathbb{R}^n, \\ x_\mathrm{d} &= U_\mathrm{x}^T x \quad x \in \mathbb{R}^m. \end{aligned} \tag{12}$$

State matrices could be obtained by solving the following linear least-squares problems:

$$\begin{aligned} &\min_F \left\| \left[U_\mathrm{p}^T X_\mathrm{p}^+, X_\mathrm{d}^+ \right] - F[U_\mathrm{p}^T X_\mathrm{p}, X_\mathrm{d}] \right\|_F, \\ &\min_H \left\| U_\mathrm{x}^T X - H[U_\mathrm{p}^T X_\mathrm{p}, X_\mathrm{d}] \right\|_F. \end{aligned} \tag{13}$$

And the solutions are

$$\begin{aligned} F &= [U_\mathrm{p}^T X_\mathrm{p}^+, X_\mathrm{d}^+][\Sigma_\mathrm{p} V_\mathrm{p}^T, X_\mathrm{d}]^\dagger, \\ H &= \Sigma_\mathrm{x} V_\mathrm{x}^T [\Sigma_\mathrm{p} V_\mathrm{p}^T, X_\mathrm{d}]^\dagger. \end{aligned} \tag{14}$$

In terms of dimensions, F is with $(n+p) \times (n+p)$ and H is with $n \times (n+p)$. When the reconstructed model obtained, the estimated flow $\hat{x}$ is reconstructed from the output $\hat{x_\mathrm{d}}$:

$$\hat{x} = U_\mathrm{x} \hat{x_\mathrm{d}}. \tag{15}$$

As a comparison, diagrams of the proposed three forms flow estimator are drawn in Fig. 1. After the estimator parameters are identified, the measured states and deterministic states are input into the Koopman linear flow estimator, which processes the inputs to the estimated flow output as the shown diagram.

3. Wind turbine wake: simulation and sensors configuration

In this study, the flow estimator is designed to reconstruct the flow in three-dimensional space, including a wind turbine wake region. This section introduces the studied case configurations and data acquisition. The sensor configurations are optimized using a genetic algorithm-based optimization procedure. Based on a series of quantitative analyses, the number of sensors is determined as five, by which functional flow estimators with acceptable accuracy are achieved.

3.1. Simulation configuration and data acquisition

Firstly, the three-dimensional wind turbine wake data is

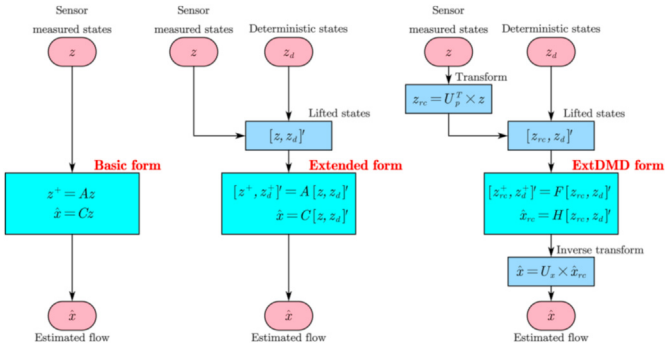

Fig. 1. Diagram of the proposed Koopman linear flow estimator.

obtained from high-fidelity simulations. Simulations are carried out using a large eddy simulation tool of Simulator fOr Wind Farm Applications (SOWFA). SOWFA is a set of computational fluid dynamics (CFD) solvers, boundary conditions, and turbine models based on the OpenFOAM CFD toolbox developed at the National Renewable Energy Laboratory (NREL). It includes a version of the turbine model coupled with FAST. Wind turbine and wind farm performance can be investigated with a full range of atmospheric conditions and terrain. Detailed SOWFA introduction and configuration are referred to Ref. [7].

In this study, a DTU 10 MW wind turbine is simulated in SOWFA. The simulated region covers a large area, including both upstream and downstream regions of the wind turbine, in order to simulate the wake appropriately. Wind speed in the wind turbine downstream area, *i.e.*, the wake region, is measured. The simulated space and the measured area are shown in Fig. 2. During simulations, the simulated space is divided into fixed-size cells with equal length in three directions, and wind speed on each cell vertices is measured. This paper only uses the axial velocity, considering it is the main component and has the most significant energy in the true flow.

The simulation runs a total of 3000 s. After 1500 s of simulation, the wake has been fully prorogation and expanding. A measured space is selected from the simulation space and measured every 2 s. The measured data is divided into identification dataset and validation dataset. The simulation configurations are listed as tables in Appendix. D, which include the simulation and measurement configurations in Table 3, wind turbine properties and ALM configurations in Table 4. Detailed instructions about the DTU 10 MW wind turbine and its ALM configurations are referred to Ref. [27].

3.2. Sensor configuration optimization

As shown in the introduced forms, the measured velocity at setting locations provides direct information of the flow field and plays an essential part in the flow estimator. Especially for the basic form, probe measurements are the only known information, while the added deterministic states in extended forms improve the accuracy helpfully. The measurement sensor's configuration in the flow field should be optimized to pursue higher accuracy than an intuitive configuration. In this subsection, the integer-programming optimization method is introduced to optimize the measurement sensor configurations. The optimization is adopted on the basic form to ensure an optimal sensor configuration without deterministic states.

In the industrial field, the sensors could be placed at any location in space as long as the studied space is reasonable. Since the pre-obtained simulated three-dimensional flow data is divided into cells, the selection in probes is better on the existing cells to avoid extra errors due to data completion and interpolation. Not just simulations, with an industrial wind turbine wake measurement using 3D scanning lidar, the flow is also measured at the number of data points on the laser beams and forms the three-dimensional flow after post-processing. Therefore, the sensor configuration optimization problem can be constructed and solved as an nonlinear integer-programming (NILP) problem. All measurement cells in the simulated space domain have a unique integer number starting from 1. The optimized parameters are maintaining integer numbers, which represent physical cells. During optimization, the Root Square Error (RSE) of the estimated and true flow is used as cost function J and defined as follows:

$$J = \left\| X - \hat{X} \right\|_F = \sqrt{\sum_{t=1}^{k} \sum_{i=1}^{m} \left| X_{i,t} - \hat{X}_{i,t} \right|^2}, \tag{16}$$

in which $\hat{X}_{i,t}$ is the i-th element of the estimated flow at time step t. $\|\cdot\|_F$ represents the Frobenius norm. m is the number of measurement points. k is total measurement time steps.

The proposed NILP problem is solved using a Genetic Algorithm (GA) called from Matlab. Considering the wide applicability of GA for different types of cost functions, most linear and nonlinear optimization problems could be satisfactorily solved. Detailed methods and applications are referred to Refs. [28,29]. During optimizations, the lower and upper boundaries ensure the variables remaining in a feasible region with desired physical meanings. An optimal solution is achieved, by which five optimized cells are selected from the studied space and play the role of sensors. The optimized sensor configuration is used for the designed three forms of flow estimators, and the performance will be compared later. Based on the error evaluation criterion RSE, a fitness index is introduced to analyze the fitness:

$$\begin{aligned} \text{Fit} - \text{Dyn} &= \left(1 - \frac{\left\| X - \hat{X} \right\|_F}{\left\| X - \bar{X} \right\|_F} \right) \times 100\% \\ &= \left(1 - \sqrt{\frac{\sum_{t=1}^{k} \sum_{i=1}^{m} \left| X_{i,t} - \hat{X}_{i,t} \right|^2}{\left| X_{i,t} - \bar{X}_i \right|^2}} \right) \times 100\%. \end{aligned} \tag{17}$$

The suffix '-Dyn' indicates that this index calculates the fitness for the dynamic wake without the mean value. Using the sensor optimization procedure, an optimal sensors configuration plan is achieved, and the fitness with the change in sensors number is drawn in Fig. 3. The sensor number starts from 5 and increases to 17 in steps of 2, and only the flow estimator in basic form is analyzed.

As shown in Fig. 3, the 5-order basic form flow estimator achieves an acceptable reconstruction accuracy higher than 55 %. Once the mean flow is also considered, the reconstruction accuracy will be higher than 99.3 %. With the number of measuring sensors increasing, the flow field reconstruction accuracy is improving. The 17-order flow estimator has good Fit-Dyn higher than 67 %. This trend consists of the DMD-based wind turbine modeling in

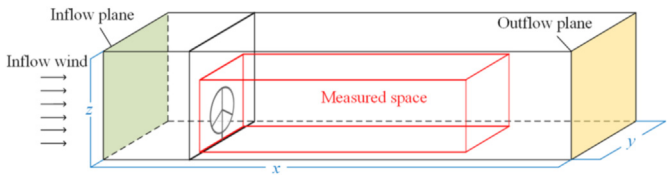

Fig. 2. Simulation space and the measured space. The Measured space is a three-dimensional space which starts from the wind turbine location and covers the wind turbine wake zone.

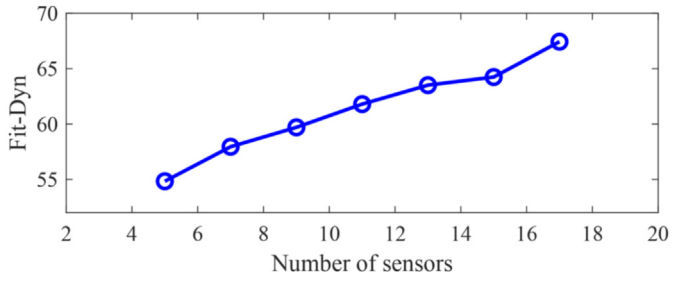

Fig. 3. Optimal sensors configuration fitness with the change in sensors number.

Ref. [30]. Since only the basic form flow estimator is analyzed here, the added deterministic states will further improve the reconstruction accuracy. Out of the industrial application possibility and investment-friendly consideration, the 5-order probe-measured states are used in later sections.

4. Estimation accuracy evaluation

In this section, the proposed flow estimator under different forms are compared to evaluate the estimation accuracy. Since the reconstruction is the main considered ability of the flow estimator, only the estimation part in the designed flow estimator is used, while the prediction ability will be verified later.

4.1. Fitness comparison

The designed flow estimators take a similar form as the reduced-order model obtained by the standard DMD method. More details could be found in B. States in the DMD-derived model are constructed from an SVD-based reducing-order mapping, and reconstructed following an inverse mapping in (B.6). So the 5-order DMD-derived flow reconstruction is compared with the proposed flow estimators. However, better reconstruction accuracy than the DMD-derived model is not the expectation of the designed flow estimator. Considering that the DMD states are defined on a subspace spanned by the dominant POD modes, the reduced-order DMD states could be regarded as virtual sensors, although they have no physical meanings. Such virtual sensors keep dominant linear information of the true flow and better reconstruct accuracy than the used physical states in this study. The proposed flow estimator is expected to have similar but maybe lower reconstruction accuracy as the DMD-derived model, which plays the role of optimal benchmark in comparisons.

The states between the basic form and other improved forms are the added observables. While in this section, the rotor-size-disk averaged wind speeds are used. Firstly, a virtual wind turbine (VWT) is placed in the wind turbine downstream wake flow. The VWT has the same y and z coordinates as the real wind turbine, while the coordinate x is adjustable. The swept area of the VWT rotor is a rotor size disk. Averaged wind speed (AWS) on the VWT disk is calculated as the inflow wind for the VWT. In this work, AWS on VWTs placed in the wake from 1D to 8D are defined as the deterministic states. Comparisons are divided into without-noise scenario and with-noise scenario. White Gaussian noises with fixed power are added to each measured point in the studied flow field. The proposed flow estimator is expected to reconstruct the true flow from the measurements with noise.

Other than the Fit-Dyn index introduced in Section 3, another fitness index is also introduced based on the error evaluation criterion of RSE:

$$\begin{aligned} \text{Fit}-\text{Full} &= \left(1-\frac{\left\|X-\hat{X}\right\|_F}{\|X\|_F}\right)\times 100\% \\ &= \left(1-\sqrt{\sum_{t=1}^{k}\sum_{i=1}^{m}\frac{\left|X_{i,t}-\hat{X}_{i,t}\right|^2}{\left|X_{i,t}\right|^2}}\right)\times 100\%. \end{aligned} \tag{18}$$

Unlike Fit-Dyn, Fit-Full retains the mean flow; thus, it represents the full-wake flow modeling accuracy. Strictly speaking, the Fit-Dyn index suits better as a quantitative index than Fit-Full, considering the mean flow is subtracted, and only the dynamic part is calculated. However, it requires an acceptable modeling/identification accuracy to ensure the fitness within an acceptable range. Once large deviations exist, the dynamic flow fitness might be negative, which shows no meaning. Thus, both are introduced. The fitness indexes are adopted to the validation case after the flow estimator is trained using the identification dataset. Results are shown in Table 1.

It can be found that, in the without-noise scenario, good accuracy is achieved by the basic form flow estimator, and the added deterministic states still improve the reconstruction accuracy: the extended form achieves similar accuracy as the DMD-derived model. On the other hand, once the SVD is included in the flow estimator, fitness in the without-noise scenario reduces: the ext-DMD form fitness is lower than the extended form. This phenomenon could be easily understood: a full-order SVD can be regarded as a linear approximation of the underlying nonlinear dynamics. In this way, some of the dynamics, especially the nonlinear part, are filtered.

Although the loss of information is inevitable, it may not be wrong. Once the measurement noise is added, the situation is different. As shown in Table 1, the ext-DMD form flow estimator takes the best performance than the basic form and the extended form. Compared with the loss information in the without-noise scenario, the SVD operation works as a filter under the with-noise scenario that the irregular nonlinear high-frequency noise will be filtered before flow reconstruction. Thus the true flow can be approximated by the flow estimator in the ext-DMD form. Another thing to note is that, the basic form reconstruction accuracy is too bad that a negative Fit-Dyn has been resulted. The fitness comparison has to be carried out on the Fit-Full index. The added deterministic states help significantly that the estimation error is highly reduced: Fit-Full of the extended form is higher than the result of the basic form. Despite this, the Fit-Dyn of the extended form is still negative, which means that the estimation accuracy is terrible. The ext-DMD form has the highest Fit-Dyn among all three estimators, approximately half Fit-Dyn compared to the DMD-derived model.

4.2. Reconstruction error visualization

The absolute average errors between the reconstructed flow and the true flow on the hub-height section are shown in Figs. 4 and 5. It shows the results of without-noise scenario and with-noise scenario separately. Sub-figures are set with the same color bar in the same parameter region to ensure fair comparisons. The extended form and ext-DMD form achieve similar effectiveness as the DMD model when ignoring the measurement noise. Both are better than the basic form. Consistency is shown during comparisons: the evidence modeling error is mainly distributed at the far-wake region and the wake outer boundary. The rotation of wind turbine blades leads to a periodic change of velocity at the wind turbine location, which has not been appropriately modeled.

In Fig. 5, the difference in reconstruction error under the with-noise scenario is shown more clearly than the without-noise scenario. The reconstruction bias of the basic form is particular: the

Table 1
Comparison of three flow estimators with DMD-derived model.

Fitness (%)	Without-Noise		With-Noise	
	Fit-Dyn	Fit-Full	Fit-Dyn	Fit-Full
Basic	53.63446336	99.29331061	−220.3070049	95.11797816
Extended	56.02613984	99.32976381	−61.59428722	97.53702908
Ext-DMD	54.97115286	99.31368402	9.671978454	98.62324780
DMD	56.68645340	99.33982811	21.76660281	98.80759038

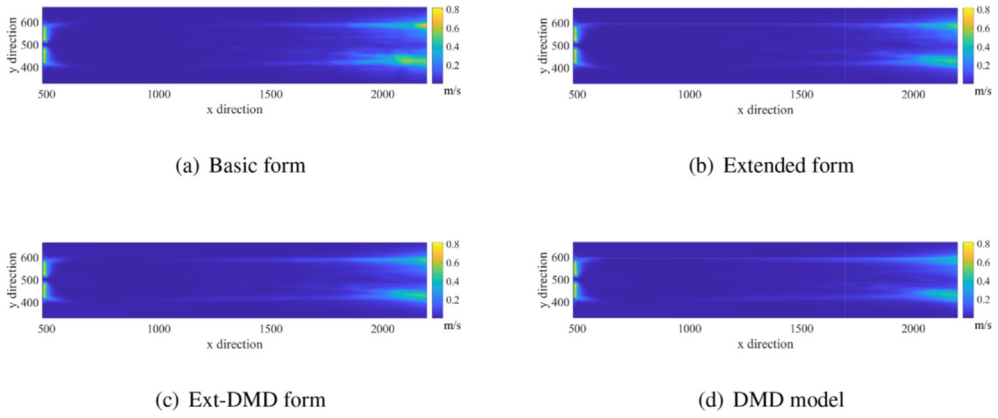

Fig. 4. Absolute average error on the hub-height section (without-noise scenario).

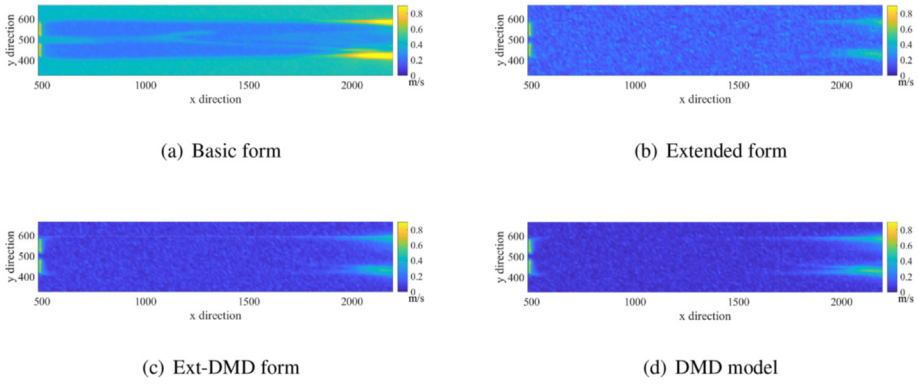

Fig. 5. Absolute average error on the hub-height section (with-noise scenario).

free-flow field out of the wake region is shown to have a constant error. In the far-wake zone, the modeling error around the wake outer boundary is also clearly marked. The bias of the free flow is so evident that the wake shape is clearly addressed. Once the deterministic states are added, the bias is reduced significantly. As to the extended form result, the far-wake region and the region close to the wind turbine have a similar model bias as the without-noise scenario, while random noises are also captured. The noise is clear enough and visible. Furthermore, the ext-DMD form filters the measurement noise effectively that the whole figure becomes cleaner. Nevertheless, the DMD-derived model maintains the best effectiveness. Although the reconstruction bias is shown to be similar to the DMD model and ext-DMD flow estimator, measurement noise under the DMD model has been filtered more effectively than the ext-DMD form. Such a feature could be explained. Although the same SVD to the probe-measured dataset is adopted as the full-flow field in the ext-DMD form in (10), five sensors are still not enough to provide sufficient dynamic information of the underlying full flow that the measurement noise can be perfect filtered. One can believe that higher reconstruction accuracy and better noise filtering result could be achieved if measurement sensors are increased.

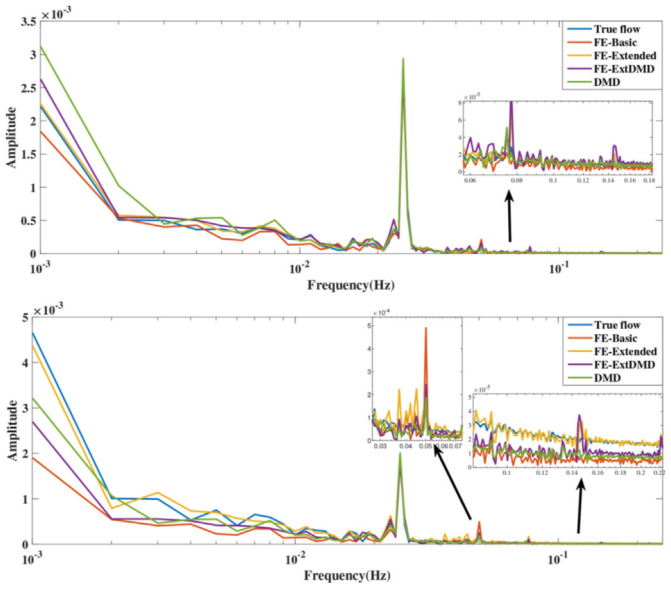

Fig. 6. Frequency-domain characteristic comparisons (Top: 5D downstream. Bottom: 8D downstream).

4.3. The frequency domain performance

Apart from the time-domain comparisons, the reconstruction accuracy in the frequency-domain is also important. Therefore, two VWTs are placed at 5D and 8D downstream, and the VWTs' AWS are analyzed using Fast Fourier Transform (FFT). Results are given in Fig. 6.

The first thing we noticed is that the wake performance at 5D and 8D downstream are different, but both the proposed estimators and DMD achieve good approximation in the frequency domain. In

5D downstream analysis, the compared four methods approximate the true flow frequency characteristics properly at the wake meandering frequency spike and higher-frequency region. The proposed flow estimator in the extended form matches the true flow among all four methods in the lower-frequency region. In the bottom of Fig. 6, the 8D-downstream location frequency characteristics are different. On the one hand, the extended form-estimated flow still has the closest frequency-domain performance as the true flow. The approximation in both high-frequency and low-frequency regions is excellent. On the other hand, apart from the extended form, the results of three methods, including DMD-estimated flow, have explicit biases from the true flow in the frequency domain. The basic form has the most significant error.

In [30], the frequency-domain analysis demonstrates the clear estimation error, especially in high-frequency region by DMD-derived dynamic wake models. Considering the system order and linear-representation form, losing information in frequency domain, especially high-frequency regions, is inevitable during DMD-derived dynamic wake modeling. It is also a typical feature during dynamic flow approximation using finite-dimensional linear-approximation methods in current research. Unlike the DMD model, the proposed flow estimators take physically measured parameters as states, capturing part of the wake performance in the high-frequency region. Thus, the extended form could achieve good approximation in the frequency domain, even better than the DMD model. But the loss comes with gain. The extended form shows less robustness than the ext-DMD form and DMD model. In general, both the three proposed flow estimators and the DMD model achieves good frequency-domain approximation, but the information loss is also evident. Better frequency accuracy requires further studies.

5. Prediction ability verification

As mentioned above, only five probes are used to provide direct information in the flow field. The added deterministic states in the extended and ext-DMD forms are also with limited scales. This raises another question: whether the dynamic part in the flow estimator has the desired prediction function? To answer this question, in this section, the prediction ability of the flow estimators is verified. Firstly, poles of the proposed flow estimators are calculated and visualized in Fig. 7. The unit circle is drawn as a stability reference.

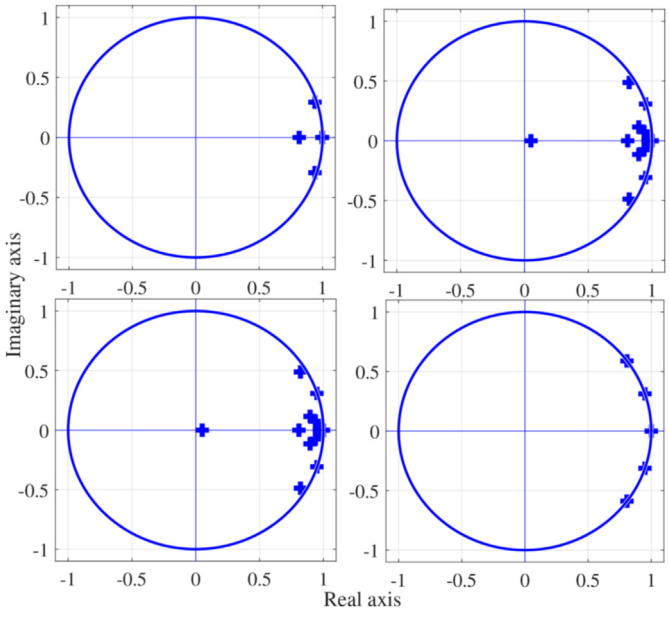

Fig. 7. Poles of the flow estimators and the DMD model. (Top left: basic form; Top right: extended form; Bottom left: ext-DMD form; Bottom right: DMD model.)

As expected, not all poles are located on the real axis. Complex poles with imaginary parts are observed in all three forms estimators. This evidence indicates that the dynamic part does describe a dynamic system. At least one oscillating mode got identified. The basic form has a pair of complex poles and three real poles, of which two poles are too close to be identified. Once the deterministic states are evolved, poles of the extended form and the ext-DMD form are almost identical but with a little difference: five poles locate on the real axis while others are complex poles. Poles of the DMD model show the true dynamics of the full flow field, where two complex pole pairs and one real pole are identified. Some consistency can also be noticed in this figure: the real pole and a pair of complex poles of the DMD model are observed in the poles of all three flow estimators, of which the poles are close but not identical. The close poles verify the effectiveness of the dynamic part for prediction ability: the flow estimators in three forms capture some true dynamics of the wake. It is worth mentioning that a complex poles pair in the DMD model is actually out of the unit circle, which should be regarded as unsteady poles pair. Nevertheless, it is very close to the unit circle even can not be identified. This is normal in dynamic system identification and can be eliminated by regularization if necessary. Meanwhile, all poles of the flow estimators remain in the unit circle and ensure stability.

Since the sensors can be selected in the studied space, poles of the flow estimator with different sensor configurations should have some consistency: randomly placed sensors should achieve similar characteristics since they all approximate part of the full-wake dynamics, which could be similar complex pole pairs. Out of this consideration, the sensors are placed randomly in the studied space, and the resulted poles of the basic form estimator are plot in Fig. 8. A total of 100,000 groups are analyzed, all sensors are limited in the wake region to ensure data validity, and only the complex poles are drawn.

In Fig. 8, DMD poles are marked in red, and the characteristics are obvious. Poles are not distributed in the unit circle arbitrarily. On the contrary, most of the poles are located near the right side of the unit circle and within a sector enclosed by the origin and the DMD poles. More specifically, poles could be divided into two clouds: most of the poles locating on the right side are marked as the central-pole cloud, some other pole near the imaginary axis and forms an arc shape is marked as the arc-pole cloud. In the central-

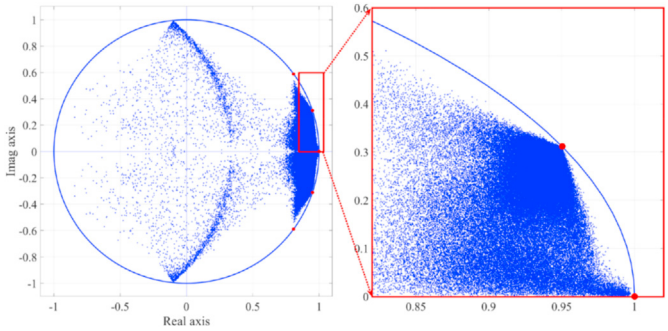

Fig. 8. Consistency check. Poles of the basic form flow estimator once the sensors are placed in the wake randomly. Pole of the 5-order DMD model are drawn in red, and unit circle is drawn in blue as the stability boundary. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

pole cloud, most poles are close to the complex pole pairs of the DMD model, which is the same as the analyzed pair above. Besides, the DMD-model real pole on the right side is sometimes identified as close pole pairs with a large real but small imaginary part. The central-pole cloud matches the consistency as expected. In addition to the central-pole cloud, the arc-pole cloud shows unique characteristics. Some poles intersect the unit circle at a particular position, similar to the intersect at the DMD poles in the central-pole cloud. A pair of DMD poles is expected to be located at the intersection point. Such a feature is beyond the captured dynamics of the five-order DMD model. Therefore, it has not been further analyzed in this paper.

The fitness of the flow reconstruction under the prediction modes is also compared. Firstly, the measured states are obtained at a given time. States in the following period are calculated using the prediction part in the introduced flow estimator. After future states are predicted, the flow is estimated using predicted states and estimator part. The flow estimators keep the same sensors configuration plan as the above section and are compared into different forms. In the following analysis, validation results under the without-noise scenario are compared from 5 to 250 times prediction in five steps. Considering the measured data are obtained in 2 s steps, the predicted period is 10−500 s in steps of 10 s. The modeling accuracy is analyzed through the quantitative criteria Fit-Dyn.

As shown in Fig. 9, the DMD model has the highest modeling accuracy during the whole prediction period. Although the general trend is the same: with increased prediction horizons, the fitness decreases continuously. Some differences are still significant and addressed as follows. Firstly, the extended form has the highest modeling accuracy but is still lower than the DMD model. The ext-DMD form has a similar but slightly lower Fit-Dyn than the extended form, and the basic form has the most insufficient prediction accuracy. Secondly, the compared methods have a similar trend with the increase in prediction steps but different decrease ratios. The basic form Fit-Dyn is decreasing quickly, while the DMD model has the slowest decrease ratio. With the further increase in prediction steps, Fit-Dyn of the basic form gradually stabilizes to a particular value, so are the extended and ext-DMD form estimators. Such characteristic is not observed on the DMD-model Fit-Dyn, at least in the current prediction process. This may due to unstable poles of the DMD model, while all poles of three flow estimators are stable. Further study is required to explain this phenomenon, which is not presented in this work. Nevertheless, acceptable accuracy is guaranteed even after a 500 s prediction: the Fit-Dyn of three proposed flow estimators is higher than 43 %. If the mean flow is retaining, the Fit-Full is higher than 99.1 %.

6. Connections with stochastic estimation related methods

The estimation part in the proposed flow estimators have the same ability as the stochastic estimation (SE) methods that, the full-flow field is reconstructed from a few physical-measured parameters. Besides, the basic form estimation part in Section 2.2 takes the same form and solution as LSE in Ref. [15]. In this section, the proposed flow estimators are compared with SE-related flow reconstruction methods, by which the reconstruction ability is focused. Specifically, four flow reconstruction methods are compared with the proposed flow estimator as follows:

- LSE-POD: In Ref. [31], the LSE-POD method is introduced as a complementary technique like POD to identify the fluid structure. It first constructs an estimated flow from probes measurement using LSE, then decomposes the estimated flow using POD and reconstructs it on a few POD bases. Both LSE and POD operations are linear. It has been applied to wind turbine inflow reconstruction in Ref. [32].
- POD-LSE: The POD-LSE method is introduced in Ref. [26]. It is an alternative way to combine POD and LSE that the POD is applied to the true-flow field and measured flow, and LSE constructs a mapping from the true-flow POD-temporal coefficient to the measured one. Once a measured flow is achieved, POD is first adopted, then the temporal coefficient of the true flow is estimated. Finally, the estimated true flow is reconstructed from the estimated POD-temporal coefficient.
- QSE: The Quadratic Stochastic Estimation (QSE) is introduced in Ref. [33]. Unlike the linear item in LSE, a quadratic term of the measured states is also included in QSE, which is expected to capture more dynamics. The linear and quadratic coefficients are connected and solved together to minimize the mean square error.
- QSE-POD: In Ref. [34], the QSE-POD is introduced to estimate the cavity flow fields. Similarly as the LSE-POD method in Ref. [31], QSE-POD also constructs the estimated flow field using QSE, then finishes the decomposition and reconstruction using dominant POD modes. As can be seen, the quadratic form is used during estimation, but only the dominant linear modes are kept after POD.

The introduced four methods are applied to the studied wind turbine wake, while the sensor configuration consists of the proposed flow estimators. The comparison results are listed in Table 2.

Among the linear methods, the first noticed thing is that the LSE-POD reaches almost the same results as the basic form, only with very little difference in Fit-Dyn under the without-noise scenario. It seems like a wrong result, but we confirm that the POD-reconstructed flow differs from the LSE estimated one. Such condition dues that the estimation basic comes from the measured flow information, *i.e.*, five probes measured velocity in this study. Considering the number of probes and the linear form in LSE, the captured flow field is limited to several linear modes. Also, the estimation part in the basic form is the same as LSE. Therefore, the

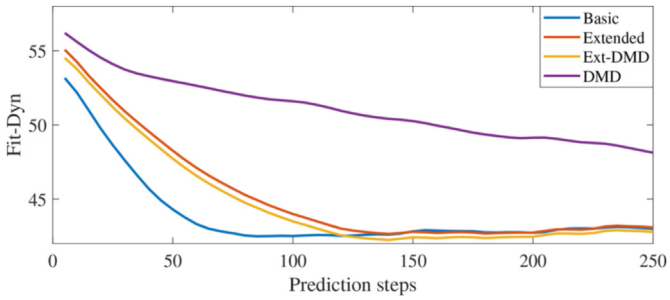

Fig. 9. Predictive fitness.

Table 2
Comparison of three flow estimators with SE-derived model.

Fitness (%)	Without-Noise		With-Noise	
	Fit-Dyn	Fit-Full	Fit-Dyn	Fit-Full
Basic	53.63446336	99.29331061	−220.3070049	95.11797816
Extended	56.02613984	99.32976381	−61.59428722	97.53702908
Ext-DMD	54.97115286	99.31368402	9.671978454	98.62324780
LSE-POD	53.63446334	99.29331061	−220.3070049	95.11797816
POD-LSE	−44.14563767	97.80297608	−82.09706950	97.22453191
QSE	55.01053546	99.31428428	−107.1421499	96.84280242
QSE-POD	53.64917103	99.29353477	−52.04458326	97.68258275

only information loss between basic form and LSE-POD are the POD projection and reconstruction, which is few. Nevertheless, POD after LSE still provides an excellent way to study the flow structure. The results of POD-LSE are different. As shown, the POD-LSE results are very different from the other methods: reconstruction accuracy under both without- and with-noise scenarios are horrible due to the limited sensors. In the original POD-LSE study of [26], the POD-LSE input data is a two-dimensional flow with a fine grid, while the flow on a coarse grid is reconstructed. The key step during POD-LSE is the POD-coefficient estimation of the to-be-reconstructed flow. For the reconstruction from fine gird to coarse gird in Ref. [26], this step is established that the measured data capture dominant flow behavior. However, measured data of only five probes are used to reconstruct the full wake flow in this work. Although the probe configurations are optimized, one can believe that the captured flow behavior of the measured data is not sufficient that the POD coefficients of aimed flow could be estimated appropriately.

The POD-LSE and LSE-POD methods share similar characteristics with the designed flow estimator. Functionally, the POD operation to the estimated/true flow works as a filter, by which the complex fluid is first decomposed into linear POD modes. The reconstruction part only takes dominant POD modes that the high-frequency components and nonlinear parts are filtered. For the POD-LSE method, the POD operation can be regarded as a *priori* filter, while it is a *posteriori* filter for the LSE-POD method. Results under the with-noise scenario also show the advantage of the 'filter'. The POD-LSE results under the without-noise scenario are the lowest among all seven compared methods. Once the measurement noise is included, the reconstruction accuracy of POD-LSE is remarkably higher than the basic form. But in LSE-POD, the estimated flow after LSE losses too much information that the POD can not help much. That is the reason why the POD-based projection is also used in the proposed flow estimator. In (11), the reconstructed states z_{rc} is obtained from a linear projection, in which the projection space is given by the n-th POD modes of X_p. We can ignore the deterministic states in the ext-DMD form, and the resulting estimator shares a similar form with the POD-LSE method. The difference is that LSE constructs linear mapping between POD coefficient in POD-LSE in Ref. [26], while in this paper, LSE maps the reduced-order states and outputs. The procedures are similar. However, during the solution in POD-LSE, eigenvalue decomposition is used to make sure the linear mapping is constructed between two unitary matrices. In contrast, it is not used in the designed estimator since the deterministic states are included.

Once we include the quadratic term, the reconstruction accuracy is improved. The QSE method has better accuracy than standard LSE (the basic form flow estimator) and LSE-POD in both with- and without-noise scenarios. The added quadratic term in QSE fills the missing modeling gap between true flow and the LSE-modeled flow. However, after QSE decreases the reconstruction bias under the without-noise scenario, the POD operation increases under the with-noise scenario. This interesting phenomenon seems contradictory but can be understood easily. QSE models the full flow from probe measurement using the linear term together with the quadratic term in the without-noise scenario. But the POD is a linear projection in essential that the quadratic term modeled part is partly filtered. This could be verified by the close accuracy between QSE-POD and the basic form under the without-noise scenario. In the with-noise scenario, some unnecessary information like noise is also partly approximated by QSE. These mistakenly approximated noises are filtered by POD so that the reconstruction accuracy of QSE-POD is better than QSE. Despite this, we are glad to see that the proposed flow estimators still achieve the best reconstruction accuracy among all seven compared methods. Under the without-noise scenario, the extended form has the highest Fit-Dyn and Fit-Full; both are higher than QSE, the second rank. The added measurement noise changes the accuracy that the ext-DMD form rank first, while the QSE-POD and extended form following behind. Such results further illustrate the necessity of deterministic states that the extended form reach acceptable accuracy under with-noise scenario, even the 'filter' function is not included.

Although we address that the estimator part in the basic form is the same as LSE, something important still should be noticed. In typical LSE study like [31,33], the state parameters are different from the estimator parameters. Take [31] as an example, both u and v are measured at the set probes. In this way, the Reynolds stress and two-point correlation tensors are used as states instead of directly measured velocity. Furthermore, measurements in an estimating event are time series, and several estimating events (in Ref. [31], K events) are considered. In this work, the LSE/basic form flow estimator is adopted simply: only one estimating event and only the streamwise velocity is studied. The states and outputs are the same kinds of variables. A dynamic part is also constructed using the standard DMD method in the basic form other than LSE. The dynamic equation partly captures or approximates the full-flow dynamics to predict the flow field in future steps. Although Fig. 7 verifies the dynamic ability that complex pole pairs approximate part of the flow dynamics, loss in higher-order dynamic characteristics in Fig. 6 and the fitness decreasing during long-term prediction in Fig. 9 are still worrying. The number of sensors limits the captured dynamics, but methodological improvements can still be adopted to improve prediction accuracy.

7. Conclusion

This paper proposed a Koopman-linear flow estimator based on the Koopman operator theory and extended dynamic mode decomposition method. The wind turbine wake flow was reconstructed using the designed flow estimator with a few probe-measured velocities and limited deterministic states. Distinguished from the states, the flow estimator was designed into three forms: the basic form, the extended form, and the ext-DMD form. Similarly, the measured probe data is always required, which provides direct information on the flow. The measured probe data in both basic and ext-DMD forms were mapped to the POD subspace that only dominant linear components were reserved. Using a heuristic integer-programming method, five probe sensors were placed optimally in the studied wake flow. Comparisons in time domain and frequency domain verified the effectiveness. Like the DMD-derived dynamic wake model, the proposed flow estimators maintained linear dynamic state-space forms that the wake flow in future steps could be predicted, among which the extended form had the closest frequency performance as the DMD-derived model. Meanwhile, all three estimators kept good predictive accuracy under long-term predictions. Once we ignore the dynamic prediction that only the flow reconstruction is considered, the proposed estimators would have good or even better accuracy than the stochastic estimation-related methods. Besides, comparisons in without- and with-noise scenarios verified the robustness of the proposed estimators, especially the ext-DMD form.

The proposed flow estimators provide a new way to approximate the dynamic wind turbine wake, taking advantage of reduced-order modeling and flow reconstruction. It helps further understanding of the wake effect, especially combined with physical measures. In practice, the proposed flow estimators serve as dynamic wake reduced-order models, in which measurable states also help further improvements like Kalman filter and Luenberger observer. Other than that, the estimation part is a new flow reconstruction method like linear stochastic estimation. In future work, the estimation part can take advantage of stochastic

estimation-related methods like quadratic stochastic estimation to achieve higher reconstruction accuracy, considering that it is relatively independent from the dynamic part. It is also restricted by specific application scenarios once linear form is required. The theoretical feasibility of the proposed flow estimators is guaranteed by the Koopman operator theory, guided by which the flow estimators could be easily applied to controlled scenarios in future works.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was partially supported by the Research on the distributed cooperative control strategy for the wake redirection of large-scale wind farm based on the intelligent clustering method, National Natural Science Foundation of China (No. 61973114), the Research on the cooperative control technology through the wake redirection of Guodian New Energy Technology Research Institute Co., Ltd. (No. GJNY-19-87), National Natural Science Foundation of China (No. U1766204), the Fundamental Research Funds for the Central Universities (No. 2018ZD05, 2018BJ0116).

Author Zhenyu Chen was sponsored by the China Scholarship Council for visiting study at Delft Center for Systems and Control (DCSC), Faculty of Mechanical, Maritime and Materials Engineering (3 mE), Delft University of Technology. Author Xiaoya Zhai was sponsored by the China Scholarship Council for visiting study at Department of Design Engineering, Delft University of Technology. Hereby thanks for the CSC fellowship support.

The authors thank Prof. Jan-Willem van Wingerden and Mr. Maarten J. van den Broek in Delft Center for Systems and Control (DCSC), Delft University of Technology for their suggestions and help on this work.

Appendix A. Koopman Operator Theory

Consider an uncontrolled nonlinear system which can be given as the following form:

$$x^+ = f(x)\, x \in M, \tag{A.1}$$

in which M is the state space of x. $g : M \to \mathbb{R}$ is a function of system which often referred to as an observable, $f(x)$ is a nonlinear function of x. The Koopman operator is a linear transformation defined by

$$(Kg)(x) = g(f(x)), \tag{A.2}$$

for every g belonging to $\mathcal{F}$, which is a space of functions invariant under the action of the Koopman operator [25]. That is to say, the Koopman operator updates the observable g based on the evolution of the trajectories in the state space. Crucially, the Koopman operator fully captures all properties of the underlying dynamical system and is always linear even if the underlaying system is nonlinear. Similar as linear systems, the observable φ is marked as a Koopman eigenfunction associated with Koopman eigenvalue $\lambda \in \mathbb{C}$ once the following equation holds:

$$K\varphi = \lambda\varphi. \tag{A.3}$$

And the spectral properties of the Koopman operator can be used to characterize the state-space dynamics. For further information and detailed explanation in Koopman operator theory [18,25], can be referred.

Appendix B. Dynamic Mode Decomposition Method

Consider a nonlinear uncontrolled discrete dynamic system in (A.1) as:

$$x_{k+1} = f(x_k), \tag{B.1}$$

where $x \in \mathbb{R}^n$ is the state vector, subscript k represents the time step. A collection of snapshot measurements $x_k \in \mathbb{R}^n$ for $k = 1, 2, \ldots, m$ is obtained for the nonlinear system via experiments. The collected data are organized into two datasets X_0 and X_1 in the following form:

$$\begin{aligned} X_0 &= [x_1, x_2, \ldots, x_{m-1}], \\ X_1 &= [x_2, x_3, \ldots, x_m]. \end{aligned} \tag{B.2}$$

The full-order linear operator A approximates the nonlinear system can be computed such that:

$$A = X_1 X_0^{\dagger}, \tag{B.3}$$

where $X_0^{\dagger}$ indicates the Moore-Penrose pseudoinverse of X_0. Dimensions of the dataset are $(m-1) \times n$ for both X_0 and X_1, so the full-order A matrix should be $(m-1) \times (m-1)$. DMD attempts to find a low-rank matrix A, capturing the most critical dynamics of the dataset and fitting the snapshots.

The r-th order singular value decomposition (SVD) of X_0 can be written as:

$$X_0 = U_{\mathrm{r}} \Sigma_{\mathrm{r}} V_{\mathrm{r}}^T, \tag{B.4}$$

where Σ_{r} is a diagonal $r \times r$ matrix with non-negative real numbers on the diagonal, V_{r} (the right-singular vectors) contains eigenvectors of $X_0^T X_0$, U_{r} (the left-singular vectors) contains eigenvectors of $X_0 X_0^T$. Both U_{r} and V_{r} are unitary matrices that satisfies $U_{\mathrm{r}}^{-1} = U_{\mathrm{r}}^T$ and $U_{\mathrm{r}} U_{\mathrm{r}}^T = I_r$, where I_r is a $r \times r$ identity matrix.

The truncated, reduced-order model takes the form:

$$z_{k+1} = (U_{\mathrm{r}}^T A U_{\mathrm{r}}) z_k = F z_k, \tag{B.5}$$

where the state matrix $F = U_{\mathrm{r}}^T A U_{\mathrm{r}} \in \mathbb{R}^{r \times r}$ describes the dynamics of the reduced-order subspace. The reduced-order state $z = U_{\mathrm{r}}^T x \in \mathbb{R}^r$. After the reduced-order predicted states are calculated, full states can be obtained following an inverse mapping:

$$x = U_{\mathrm{r}} z. \tag{B.6}$$

Reduced-order state matrix F can be calculated following

$$F = U_{\mathrm{r}}^T A U_{\mathrm{r}} = U_{\mathrm{r}}^T X_1 (U_{\mathrm{r}} X_0)^{\dagger} = U_{\mathrm{r}}^T X_1 V_{\mathrm{r}} \Sigma_{\mathrm{r}}^{-1}. \tag{B.7}$$

Eigenvalues and eigenvectors of F are calculated as $F\omega = \lambda\omega$, the DMD mode ϕ corresponding to the DMD eigenvalue λ is then given by

$$\phi = U_{\mathrm{r}} \omega. \tag{B.8}$$

Under such circumstances, the DMD modes closely approximate the Koopman modes, and the DMD method can be regarded as a solving or approximation method to compute the Koopman operator. Another thing to note is that, though the Koopman analogy provides a firmly mathematical foundation for applying DMD to

data generated by nonlinear systems, it is still limited by a number of assumptions. Detailed explains and analysis of the DMD and Koopman operator can be referred to as [23,24].

Appendix C. Extended Dynamic Mode Decomposition Method

In [19,22], an EDMD method was proposed and proven to achieve good performance in the construction of the Koopman-linear system. Like the DMD method, the nonlinear flow characteristic in the EDMD method can be approximated via a linear time-invariant system. Considering the set of states of the nonlinear system in the form of:

$$\begin{aligned} X &= [x_1, x_2, \ldots, x_k], \\ X^+ &= [x_1^+, x_2^+, \ldots, x_k^+], \end{aligned} \tag{C.1}$$

in which $x_j^+ = T(x_j)$ is an unknown nonlinear function. For this uncontrolled system, states' value is assumed to be unavailable, and the observable have physical meaning and can be obtained. Let

$$\boldsymbol{g}(x) = [g_1(x), g_2(2), \ldots, g_m(x)] \tag{C.2}$$

be a given vector of possibly nonlinear observables, in which $g_m(x)$ with specified subscript is an observable. These functions may represent user-specified nonlinear functions of the state as well as physical measurements. The collected snapshot measurements of the system are

$$\begin{aligned} X_{\text{lift}} &= [\boldsymbol{g}(x_1), \boldsymbol{g}(x_2), \ldots, \boldsymbol{g}(x_k)], \\ X_{\text{lift}}^+ &= [\boldsymbol{g}(x_1^+), \boldsymbol{g}(x_2^+), \ldots, \boldsymbol{g}(x_k^+)]. \end{aligned} \tag{C.3}$$

The measurement matrix is marked with subscript lift for distinguishment with the original states. Based on the collected data, an observer can be constructed to estimate the state values while approximating the underlying nonlinear dynamics of the uncontrolled system, which takes the following form:

$$\begin{aligned} z^+ &= Az \;\; z \in \mathbb{R}^n, \\ \hat{x} &= Cz \;\; x \in \mathbb{R}^m, \end{aligned} \tag{C.4}$$

where z is the observer states and $z^+ = Az$ describes a linear approximation of the nonlinear function $x^+ = T(x)$, and $\hat{x}$ is the estimated states x. State matrices A, C are obtained by the solution to the linear least-squares problems:

$$\begin{aligned} &\min_A \left\| X_{\text{lift}}^+ - AX_{\text{lift}} \right\|_F, \\ &\min_C \left\| X - CX_{\text{lift}} \right\|_F, \end{aligned} \tag{C.5}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. The analytical solution to (C.5) is:

$$\begin{bmatrix} A \\ C \end{bmatrix} = \begin{bmatrix} X_{\text{lift}}^+ \\ X \end{bmatrix} [X_{\text{lift}}]^\dagger. \tag{C.6}$$

The matrix A describe the linear dynamics of the Koopman-linear state $z = \boldsymbol{g}(x)$, and the estimated states $\hat{x}$ is obtained with equation (C.4).

To achieve satisfactory accuracy, the snapshot matrix X_{lift} is usually fat, *i.e.*, the number of columns exceeds the number of rows. Hence, the Moore-Penrose pseudoinverse of X_{lift} is calculation heavily. An alternative solution can be used to improve the solving efficiency of (C.6). Equation (C.6) can be modified as follows:

$$MG = V. \tag{C.7}$$

In the modified equation, M is the unknown matrix variable

$$M = \begin{bmatrix} A \\ C \end{bmatrix},$$

and the data matrices V and G are

$$\begin{aligned} V &= \begin{bmatrix} X_{\text{lift}}^+ \\ X \end{bmatrix} [X_{\text{lift}}]^T, \\ G &= [X_{\text{lift}}][X_{\text{lift}}]^T. \end{aligned} \tag{C.8}$$

Solution to (C.7) is the same as (C.6), but the size of the matrices V and G are independent of the number of samples k in (C.1). Thus, the increase in samples will affect the solution result but not directly related to the order of problem-solving.

Appendix D. Large-Eddy Simulation Configurations

Table 3
Simulation and measurement configurations.

Simulation Configuration		Measurement Configuration	
Simulated cell numbers	$800 \times 400 \times 400$	Measured cell numbers	$184 \times 37 \times 33$
Simulated cell size	3.125 m	Measured cell size	9.375 m
Simulated time step	0.5s	Measured time step	2s
Turbine type	DTU 10 MW	Measured time region	1500s–3000s
Turbine model	Actuator line model	Identification dataset	1500s–2300s
Free stream wind speed	8.0 m/s	Validation dataset	2000s–3000s

Table 4
DTU 10 MW wind turbine ALM model parameters.

Wind turbine properties		ALM configurations	
Rated power	10 MW	Number of blade points	40
Rotor orientation, configuration	Upwind, 3 blades	Number of Nacelle points	10
Rotor, Hub diameter	178.3 m, 5.6 m	Number of tower points	40
Hub height	119 m	Nacelle sample distance	1.0 m
Cut-in, Rated, Cut-out wind speed	4 m/s, 11.4 m/s, 25 m/s	Tower sample distance	3.5 m
Overhang, Shaft tilt, Pre-cone	7.1 m, 5deg, 2.5deg	Air density	1.23

Credit author statement

Zhenyu Chen: Conceptualization, Methodology, Software, Writing, Validation, Formal analysis. **Xiaoya Zhai**: Methodology, Software, Writing, Formal analysis. **Zhongwei Lin:** Methodology, Writing, Funding acquisition, Supervision, Project administrations. **Jizhen Liu**: Methodology, Writing-revision, Supervision, Funding acquisition.

References

[1] van Wingerden JW, Fleming PA, Göçmen T, Eguinoa I, Doekemeijer B, Dykes K, Lawson M, Simley E, King J, Astrain D, et al. Expert elicitation on wind farm control. In: Journal of physics: conference series, vol. 1618. IOP Publishing; 2020, 022025.

[2] Doekemeijer BM, Van Wingerden J-W, Fleming PA. A tutorial on the synthesis and validation of a closed-loop wind farm controller using a steady-state surrogate model. In: 2019 American control conference (ACC), IEEE; 2019. p. 2825–36.

[3] Qian G-W, Ishihara T. Wind farm power maximization through wake steering with a new multiple wake model for prediction of turbulence intensity. Energy 2021;220:119680.

[4] Sun H, Gao X, Yang H. Validations of three-dimensional wake models with the wind field measurements in complex terrain. Energy 2019;189:116213.

[5] Dou B, Qu T, Lei L, Zeng P. Optimization of wind turbine yaw angles in a wind farm using a three-dimensional yawed wake model. Energy 2020;209:118415.

[6] Ti Z, Deng XW, Yang H. Wake modeling of wind turbines using machine learning. Appl Energy 2020;257:114025.

[7] Churchfield MJ, Lee S, Michalakes J, Moriarty PJ. A numerical study of the effects of atmospheric and wake turbulence on wind turbine dynamics. J Turbul 2012;13:N14.

[8] Munters W, Meyers J. Dynamic strategies for yaw and induction control of wind farms based on large-eddy simulation and optimization. Energies 2018;11(1):177.

[9] Gebraad PM, Fleming PA, van Wingerden J-W. Wind turbine wake estimation and control using floridyn, a control-oriented dynamic wind plant model. In: 2015 American control conference (ACC), IEEE; 2015. p. 1702–8.

[10] Reinwardt I, Gerke N, Dalhoff P, Steudel D, Moser W. Validation of wind turbine wake models with focus on the dynamic wake meandering model. J Phys Conf Ser 2018;1037:072028. IOP Publishing.

[11] Debnath M, Santoni C, Leonardi S, Iungo GV. Towards reduced order modelling for predicting the dynamics of coherent vorticity structures within wind turbine wakes. Phil Trans Math Phys Eng Sci 2017;375(2091):20160108.

[12] De Cillis G, Cherubini S, Semeraro O, Leonardi S, De Palma P. Pod analysis of the recovery process in wind turbine wakes. In: Journal of physics: conference series, vol. 1618. IOP Publishing; 2020, 062016.

[13] Hamilton N, Viggiano B, Calaf M, Tutkun M, Cal RB. A generalized framework for reduced-order modeling of a wind turbine wake. Wind Energy 2018;21(6):373–90.

[14] Tu JH. Dynamic mode decomposition: theory and applications. Ph.D. thesis. Princeton University; Setptember 2013.

[15] Adrian RJ, Moin P. Stochastic estimation of organized turbulent structure: homogeneous shear flow. J Fluid Mech 1988;190(-1):531.

[16] Arnault A, Dandois J, Foucaut JM. Comparison of stochastic estimation methods with conditional events optimization for the reconstruction of the flow around a supercritical airfoil in transonic conditions. Comput Fluid 2016;136:436–55.

[17] Korda M, Mezić I. On convergence of extended dynamic mode decomposition to the koopman operator. J Nonlinear Sci 2018;28(2):687–710.

[18] Korda M, Mezić I. Linear predictors for nonlinear dynamical systems: koopman operator meets model predictive control. Automatica 2018;93:149–60.

[19] Williams MO, Kevrekidis IG, Rowley CW. A data–driven approximation of the koopman operator: extending dynamic mode decomposition. J Nonlinear Sci 2015;25(6):1307–46.

[20] Cassamo N, van Wingerden J-W. On the potential of reduced order models for wind farm control: a koopman dynamic mode decomposition approach. Energies 2020;13(24):6513.

[21] Arbabi H, Korda M, Mezić I. A data-driven koopman model predictive control framework for nonlinear partial differential equations. In: 2018 IEEE conference on decision and control (CDC), IEEE; 2018. p. 6409–14.

[22] Williams MO, Hemati MS, Dawson ST, Kevrekidis IG, Rowley CW. Extending data-driven koopman analysis to actuated systems. IFAC-PapersOnLine 2016;49(18):704–9.

[23] Kutz JN, Brunton SL, Brunton BW, Proctor JL. Dynamic mode decomposition: data-driven modeling of complex systems, vol. 149. SIAM; 2016.

[24] Tu JH, Rowley CW, Luchtenburg DM, Brunton SL, Kutz JN. On dynamic mode decomposition: theory and applications. J Comput Dynam 2014;1(2):391.

[25] Arbabi H. Introduction to koopman operator theory of dynamical systems. 2018.

[26] Podvin B, Nguimatsia S, Foucaut JM, Cuvier C, Fraigneau Y. On combining linear stochastic estimation and proper orthogonal decomposition for flow reconstruction. Exp Fluid 2018;59(3):58.

[27] J. W. v Doekemeijer WBM. Dtu10mw_sowfa https://github.com/Bartdoekemeijer/DTU10MW_SOWFA. [Accessed 25 April 2021].

[28] Zhang C, Wang L, Wu X, Gao W. A novel optimal configuration of sensor and actuator using a non-linear integer programming genetic algorithm for active vibration control. J Intell Mater Syst Struct 2017;28(15):2074–81.

[29] Deep K, Singh KP, Kansal ML, Mohan C. A real coded genetic algorithm for solving integer and mixed integer optimization problems. Appl Math Comput 2009;212(2):505–18.

[30] Chen Z, Doekemeijer BM, Lin Z, Xie Z, Wingerden JWV. Datadriven modeling & analysis of dynamic wake for wind farm control: a comparison study. In: 2020 Chinese automation congress (CAC); 2020.

[31] Bonnet JP, Cole DR, Delville J, Glauser MN, Ukeiley LS. Stochastic estimation and proper orthogonal decomposition: complementary techniques for identifying structure. Exp Fluid 1994;17(5):307–14.

[32] Spitler J, Naughton J, Lindberg W. An lse/pod estimation of the wind turbine inflow environment using sparse data. In: AIAA aerospace sciences meeting and exhibit; 2006. p. 1364.

[33] Murray N, Ukeiley L. Estimating the shear layer velocity field above an open cavity from surface pressure measurements. In: 32nd AIAA fluid dynamics conference and exhibit; 2002. p. 2866.

[34] Arunajatesan S, Kannepalli C, Ukeiley L. Three dimensional stochastic estimation applied to cavity flow fields. In: 37th AIAA fluid dynamics conference and exhibit; 2007. p. 4227.

[35] Cassamo N, van Wingerden JW. Model predictive control for wake redirection in wind farms: a koopman dynamic mode decomposition approach[C]//2021 American Control Conference (ACC). IEEE; 2021. p. 1776–82.