

PAPER • OPEN ACCESS

A Simulative Analysis of Physically-Motivated Koopman Lifted States for Wind Farm Model Predictive Control

To cite this article: Antje Dittmer and Herbert Werner 2026 *J. Phys.: Conf. Ser.* **3224** 032085

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

You may also like

- [Wind farm flow reconstruction and prediction from high frequency SCADA Data](#)
Andreas Rott, Vlaho Petrovi and Martin Kühn
- [Data-driven Reduced Order Model for prediction of wind turbine wakes](#)
G V Iungo, C Santoni-Ortiz, M Abkar et al.
- [Joint state-parameter estimation for a control-oriented LES wind farm model](#)
B.M. Doekemeijer, S. Boersma, L.Y. Pao et al.

Enabling Electrochemistry Inside and Outside the Lab PalmSens

Discover our diverse range of Potentiostats engineered to support your research wherever it takes you

- Intuitive software for Windows and Android
- Proven in over 4000 scientific publications
- Single-channel and multi-channel
- Benchtop, portable and wearable

Discover More



A Simulative Analysis of Physically-Motivated Koopman Lifted States for Wind Farm Model Predictive Control

Antje Dittmer¹ and Herbert Werner²

¹Institute of Flight Systems, German Aerospace Center (DLR), Brunswick, Germany

²Institute of Control Systems, Hamburg University of Technology (TUHH), Hamburg, Germany

E-mail: antje.dittmer@dlr.de, h.werner@tuhh.de

Abstract.

This paper investigates how the selection of data points from a simulated wind field, combined with nonlinear lifting functions, influences wind turbine power predictions. The objective is to develop accurate linear models within the Koopman operator framework suitable for real-time Model Predictive Control. Using Dynamic Mode Decomposition, we compare three data configurations from a high-fidelity simulation: a dense wind field with 50,000 points, a dataset restricted to turbine locations, and a sparse dataset including two additional streamwise positions. Results demonstrate that models using data from well-placed sensors can outperform those using dense wind field data. Furthermore, while turbine-only data initially yields lower accuracy, we show that physically motivated Koopman lifting functions effectively compensate for missing spatial information. By incorporating these additional prediction features, turbine-only models achieve prediction accuracies comparable to those utilizing the complete wind field data, which relies on information density unavailable for real-world turbine control.

1. Introduction

Active flow control in wind farms is challenging because upstream wakes reduce the power available to downstream turbines. Controlling the entire farm is therefore essential to increase energy capture and reduce maintenance costs. To solve this, researchers have proposed data-driven methods to create farm models [1, 2] that generate predictions within milliseconds and can be used as a basis for Model Predictive Control. Such high-speed execution is a prerequisite for this control algorithm, which requires a model that is both sufficiently accurate for good performance and computationally efficient for real-time optimization.

Recent studies applied Dynamic Mode Decomposition (DMD) to control turbine induction and wake redirection [3, 4] based on high-fidelity data generated with the software SOWFA (Simulator for Wind Farm Applications) [5]. This approach uses the Koopman operator framework to represent nonlinear wind flows as linear models. This linearization allows for the use of efficient linear control algorithms while still modeling the nonlinear physics of the wind farm. The quality of the linear approximation depends on the choice of lifting functions, which are nonlinear transformations used to generate new prediction features from state variables. For wind farm models, these functions transform wind speed data into higher-order terms, such as



Content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

squared or cubed values, to help a linear model capture nonlinear turbine and wake physics. The models in [3] are identified on a three-dimensional grid of 50,000 points to predict turbine power, reaching Variance Accounted For (VAF) values above 90%. Since VAF value measures the percentage of the signal's variation captured by the model, these scores mean that less than 10% of the signal remains as unexplained noise or error. However, such dense grids are unrealistic for real-time applications: lidar systems cannot refresh 50,000 points fast enough to track dynamic wake movements in real time, creating a lag that prevents active control. As this study shows, such high-resolution data is also unnecessary for accurate power prediction.

Other studies have focused on reconstructing wakes using only a few data points and generating additional prediction features with lifting functions. Our previous work with the mid-fidelity WFSim (WindFarmSimulator) tool [6] showed that these models can be designed using only effective wind speeds [7–9]. However, since WFSim uses simplified equations, it was unclear if these results would remain accurate in high-fidelity environments. The openly available SOWFA dataset was selected for this study because it was used in [3] and [4], providing a benchmark for direct comparison. In these specific simulation cases, the absence of vertical shear allowed the identification process to use only hub-height data. While a more complex vertical profile would require additional vertical data points, these results demonstrate that for a uniform inflow, strategically placed horizontal data points are sufficient to capture the turbine dynamics.

It is natural to expect that using data only from turbine locations leads to a loss of important wind field information. However, this study shows that a few additional, well-placed virtual sensors can actually be used to build models with prediction accuracy exceeding those based on a dense wind field. Furthermore, we explore whether physically-motivated features can recover accuracy when measurements are limited to the turbines themselves. In real-world applications, achieving high accuracy with a minimal sensor set reduces hardware costs while maximizing potential profit from the wind farm's power output.

This work investigates the accuracy of DMD models generated from a SOWFA dataset when restricted to sparse measurements. We analyze how different lifting functions affect prediction quality and demonstrate that strategically placed sparse measurements can outperform high-resolution three-dimensional grids in power prediction. Concretely, we examine the effect of including quadratic and cubic terms of wind speeds in our system identification. Finally, we examine the impact of lifting functions on rotor speeds by incorporating not only direct measurements, their squares and cubes, but also moving averages, time derivatives, and cross-multiplication terms between current and delayed turbine states. These additions allow the model to better capture the characteristic time delays introduced by wake propagation.

2. Dataset Description

The wind farm dataset used in [3] consists of an identification and a verification set containing time-series signals from two DTU 10 MW turbines and three-dimensional wind field data. The data and code for [3] is provided in the Git repository [10]. The turbines have a rotor diameter of $D = 178$ m and are spaced $5D$ (890 m) apart. Turbine signals, including power and rotor speed, were recorded every 0.2 s, while wind field data were sampled every 2 s. Both datasets have a total simulation duration of 2000 s, but the first 182 s are discarded to ensure the wake from the upstream turbine is fully developed before analysis begins. The original wind field uses a three-dimensional grid of 96,600 points, which is restricted by the available code in [10] to a reduced grid of 51,543 points centered on the turbines. To excite the system dynamics for identification, the upstream turbine's yaw signal is varied in discrete steps between 0° and 20° . While both datasets use these step changes, the start times and durations differ between the

identification and verification sets to ensure the model is tested on unseen transitions.

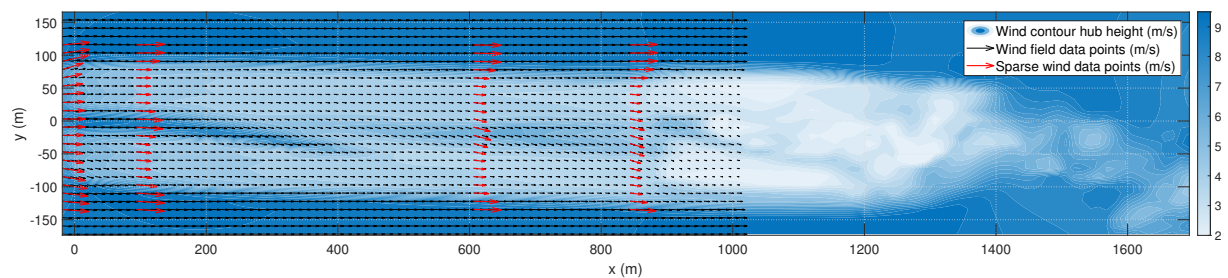
The code to reproduce the results presented here is available in the Git repository [11]. Table 1 summarizes the dataset dimensions used to evaluate prediction accuracy under realistic sensing limits. This study compares the full wind field, wind at turbines, and sparse wind data.

Table 1. Comparison of Dataset Dimensions and Sparse Subsets

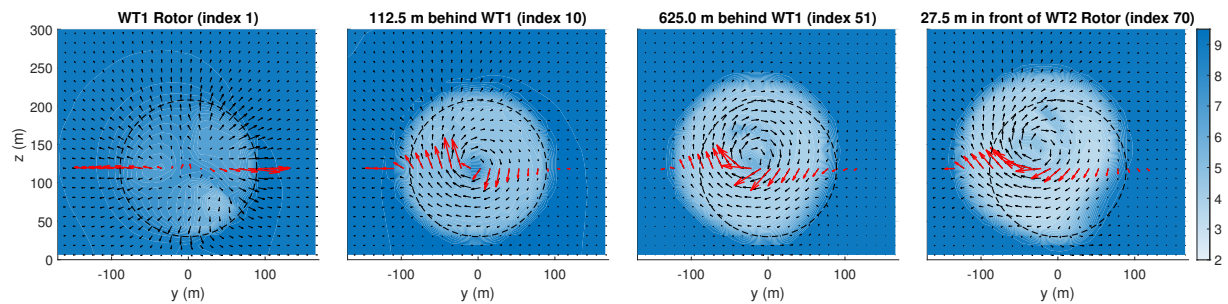
Parameter	Wind field	Wind at turbines	Sparse wind data
Total Grid Points	51,543	42	84
Hub-Height Points	2,241	42	84
Longitudinal	83	2	4
Lateral	27	21	21
Vertical	23	1	1
Usable Samples	909	909	909
Time Duration	1818 s	1818 s	1818 s

The wind field utilizes 51,543 grid points. For the wind at turbines set, we extract 42 points from the hub-height horizontal plane near the rotor disks. The sparse wind dataset expands this to 84 points by adding two intermediate cross-sections downstream of the first turbine, also at hub height. In this work, we investigate complexity reduction within a horizontal grid specifically for the wind field from [10] which does not have a vertical shear. However, our findings suggest that a similar reduction in complexity might be achievable for three-dimensional wind fields with a vertical shear component, representing a promising direction for future research.

As shown in Figure 1, the selected virtual sensors capture both the wake development and the adjacent free-stream flow.



(a) Plan view (X-Y plane) of the wind field at hub-height ($z = 119$ m).



(b) Cross-sectional view (Y-Z plane) at the rotors of WT1 and WT2 and two intermediate locations.

Figure 1. Simulated wind field at the 500th verification sample. Black arrows represent the full wind field; red arrows indicate the 84 positions used for the sparse wind data.

In the figure, the dense wind field is represented by black arrows, while the sparse points are highlighted in red. The locations for these sparse sensors were determined through a sensitivity analysis. For each cross-section, we selected 21 lateral points covering a 250 m span at hub height. Although the second turbine is located at $5D$, we found that measurements slightly upstream at 862.5 m yielded the highest VAF value. Similarly, the intermediate locations at 112.5 m and 625 m downstream of the first turbine were identified as the most informative regions for power prediction. While we present the best-performing configuration here, several other sparse combinations also outperformed the dense-measurement model, see Section 4.2. This suggests that sparse sampling improves prediction by excluding localized turbulent noise inherent in the full grid. The turbine signals are resampled and synchronized with the 909 usable wind field snapshots. The resampled yaw, power, and rotor speed signals as well as the wind speed measurements are subsequently used for identification and validation of the DMD models.

3. Koopman-based Identification

The Koopman framework represents finite-dimensional nonlinear systems as infinite-dimensional linear systems by mapping state data into a higher-dimensional space using lifting functions [12]. These functions are nonlinear transformations that map the original wind speed data into a higher-dimensional state space, allowing the framework to capture nonlinear flow physics within a linear dynamical structure. In this work, DMD is used to find a finite matrix approximation of this infinite-dimensional Koopman operator. Further methodological details are provided in [13, 14].

The Koopman-based DMD identification process treats the wind farm as a discrete-time nonlinear dynamical system which is approximated by a discrete-time linear model. The state vector for this system is formed by flattening the longitudinal wind measurements into a single array. The dimension of this state changes depending on the measurement configuration: the full wind field vector contains 51,543 elements, while the sparse and turbine-only sets contain 84 and 42 elements, respectively. The size of this vector is reduced via dimensionality reduction to ensure that all important information is retained, and the rotor speeds are then attached to it. In all cases, the control input is the yaw angle of the upstream turbine, and the outputs are the power signals from both turbines. To prepare the data for identification, we apply the following scaling methods to maintain physical consistency: turbine rotor speeds and power outputs are standardized as z-scores by subtracting the mean and dividing by the standard deviation. For the wind measurements, however, we scale the values by the standard deviation of the total velocity magnitude without subtracting the mean. This approach is chosen as removing the mean would distort the physical relationship between the dominant longitudinal flow and the much smaller lateral and vertical wind components. Figure 2 summarizes the identification pipeline.

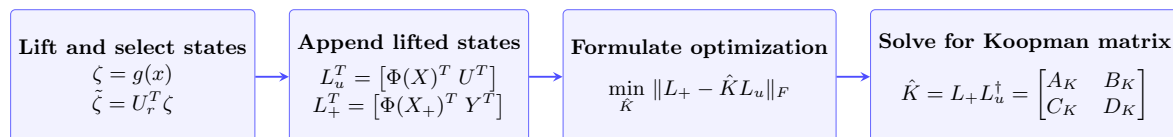


Figure 2. Pipeline of the DMD identification process from lifting and state selection to the least-squares solution of the Koopman matrix.

This workflow moves from state lifting and dimensionality reduction to the final matrix solution, with the specific computational steps detailed in the following subsections.

3.1. Lifting Function and State Selection

We apply lifting functions to the original wind states to capture the nonlinear physics of the wind farm. We consider three variants of the lifted state ζ : (i) unmodified wind speed; (ii) quadratic terms, which relate to turbine thrust; (iii) cubic terms, which are proportional to turbine power; and (iv) data augmented with both quadratic and cubic terms. However, due to the high dimensionality of these lifted states, especially for the wind field case, we use Singular Value Decomposition to project the data into a reduced-order space of dimension n_r with a state $\tilde{\zeta}$. Following the methodology in [3], we test all truncation ranks up to a $n_r=100$ and select the one with the the maximum VAF value of the predicted power for the downstream turbine. The VAF value between a measured signal vector $\mathbf{y}$ and its prediction $\hat{\mathbf{y}}$ is defined as:

$$\text{VAF}(\mathbf{y}, \hat{\mathbf{y}}) = \max \left(0, \left(1 - \frac{\text{var}(\mathbf{y} - \hat{\mathbf{y}})}{\text{var}(\mathbf{y})} \right) \right) \cdot 100\%$$

where $\mathbf{y} = [y_1, y_2, \dots, y_{n_O}]^T$ and $\hat{\mathbf{y}} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_{n_O}]^T$ represent the sequences of measured and predicted values at the n_O discrete sample times, respectively. A VAF value of 100% indicates a perfect model fit, while lower values represent the portion of the signal's variation left unexplained by the model.

We incorporate the rotor speeds of the turbines, ω_1 and ω_2 , to further enhance the model. These signals undergo their own lifting process, including quadratic and cubic terms, moving averages, and time derivatives; see [7] for details. Depending on how many lifted states are included, this results in a number n_{g_ω} of additional states ranging from two, if only the rotation speeds are included, to 24, if all lifting functions are used. These additional turbine-specific states are appended to the reduced-order wind state vector $\tilde{\zeta}$ forming the vector $\tilde{\zeta}_\omega$ which is used for system identification.

In addition to VAF, we use the Normalized Root Mean Square Error (NRMSE) to quantify the absolute difference between the predicted and true wind farm power:

$$\text{NRMSE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{\sqrt{\frac{1}{n_O} \sum_{i=1}^{n_O} (y_i - \hat{y}_i)^2}}{\text{mean}(\mathbf{y})} \cdot 100\%.$$

While VAF measures how well the model captures dynamic fluctuations and phase, NRMSE assesses point-by-point deviations and accounts for any mean offset or bias. Utilizing both metrics ensures that the identified models are not only dynamic but also accurate in their absolute power predictions, thus providing a suitable foundation for Model Predictive Control.

3.2. Koopman Matrix Calculation

We use $n_O = 909$ time samples to construct two data matrices, L_u and L_+ , each containing $n_O - 1$ data samples to identify a linear system for power prediction. These matrices organize the lifted states ζ , inputs u , and outputs y into time-shifted pairs:

$$L_u = \begin{bmatrix} \tilde{\zeta}_{\omega,1} & \cdots & \tilde{\zeta}_{\omega,n_O-1} \\ u_1 & \cdots & u_{n_O-1} \end{bmatrix}, \quad L_+ = \begin{bmatrix} \tilde{\zeta}_{\omega,2} & \cdots & \tilde{\zeta}_{\omega,n_O} \\ y_1 & \cdots & y_{n_O-1} \end{bmatrix}.$$

The Koopman matrix $\hat{K}$ is then identified by solving the minimization problem:

$$\min_K \|L_+ - K L_u\|_F^2.$$

The analytical solution is found via the Moore-Penrose pseudoinverse (†):

$$\hat{K} = L_+ L_u^\dagger = \begin{bmatrix} A_K & B_K \\ C_K & D_K \end{bmatrix}$$

where the Koopman matrix can be split into the state-space matrices A_K, B_K, C_K , and D_K . These matrices are used to predict the turbine power $\hat{y} = [P_1, P_2]^T$ based on the yaw control input $u = \gamma$ and the lifted states $\tilde{\zeta}_\omega$ through the state-space representation:

$$\tilde{\zeta}_{\omega,k+1} = A_K \tilde{\zeta}_{\omega,k} + B_K u_k, \quad \hat{y}_k = C_K \tilde{\zeta}_{\omega,k} + D_K u_k$$

This linear model is trained on the SOWFA identification data, and its accuracy is evaluated by predicting the power output for the unseen verification dataset.

4. Results and Discussion

This section evaluates the power prediction accuracy of the different identified models on the verification data of the three different measurement sets, provides a sensitivity analysis and discusses future research directions.

4.1. Model Performance: VAF and NRMSE Values and Time-Series Comparison

Table 2 summarizes the VAF values and NRMSE values for power predictions across all three measurement configurations. The upstream turbine P_1 is predicted with near-perfect accuracy (VAF > 99.8%) in every case, as its power is driven by the steady inflow. The primary challenge is predicting the downstream turbine P_2 , where the model must capture the complex, time-varying wake dynamics generated by the upstream yaw changes.

Table 2. VAF values for P_1 and P_2 and NRMSE for P_{WF} , both in percent, using different measurement sets and lifting functions. Bold values indicate the best-performing models for the downstream turbine for four different data categories.

Datasets	Lifting Func.	$n_g(\zeta)$	$n_r(\tilde{\zeta}) + n_{g_\omega}(\omega)$	VAF(P_1)	VAF(P_2)	NRMSE
Wind field	Lin.	51,543	51 + 2	99.87	89.39	4.89
	Lin. + quad.	103,086	25 + 2	99.87	90.22	4.94
	Lin. + cubic	103,086	26 + 2	99.87	90.87	4.85
	Lin, quad. + cub.	154,629	26 + 2	99.87	90.86	4.85
Sparse wind data	Lin.	84	17 + 2	99.87	92.28	4.06
	Lin. + quad.	168	19 + 2	99.86	91.85	4.18
	Lin. + cubic	168	13 + 2	99.86	92.88	3.81
	Lin, quad. + cub.	252	13 + 2	99.86	92.92	3.81
Wind at turbines	Lin.	42	41 + 2	99.89	81.89	6.93
	Lin. + quad.	84	82 + 2	99.86	86.01	5.18
	Lin. + cubic	84	82 + 2	99.89	86.19	5.14
	Lin, quad. + cub.	126	100 + 2	99.86	86.79	5.05
Lin., quad. + cubic wind at turbines + Koopman	$n_{g_\omega} = 4$	126	90 + 4	99.89	87.02	4.97
	$n_{g_\omega} = 6$	126	90 + 6	99.89	86.56	5.15
	$n_{g_\omega} = 12$	126	53 + 12	99.88	87.34	5.04
	$n_{g_\omega} = 24$	126	68 + 24	99.89	88.14	4.92

Table 2 also illustrates the dimensionality reduction achieved through Singular Value Decomposition across the different measurement configurations. For the full wind field, the state space is compressed from over 150,000 lifted features down to 26 states to which two rotor speeds are appended. This massive reduction indicates a high degree of spatial redundancy in the wind field, where neighboring flow points are strongly correlated. In contrast, the reduction ratio is

significantly lower for the sparse and turbine-only datasets and shows that they contain much less redundant information. These measurements are already spatially filtered and contain less redundant information. This also explains why the optimal truncation rank n_r remains relatively high for the turbine-only case to capture the necessary dynamical fluctuations. The wind field can be represented by a much smaller rank relative to its original size, but the turbine-only set cannot be reduced further without losing critical information required to characterize the system dynamics.

The sparse wind data achieved a remarkably high accuracy of 92.92%, surpassing the full wind field model by approximately 2%. This counter-intuitive result suggests that the dense three-dimensional grid may introduce localized turbulent noise that hinders the identification process. By restricting measurements to specific near-wake and far-wake cross-sections, the DMD process focuses on the most relevant flow gradients. This is supported by the fact that the sparse models consistently require fewer selected states ($n_r + n_{g_\omega} = 15$) than those relying on the full wind field ($n_r + n_{g_\omega} = 28$). Furthermore, for these two datasets, the impact of adding complex Koopman lifting functions to the rotor speed was negligible, suggesting that a well-placed spatial grid provides sufficient information to linearize the system dynamics implicitly.

In contrast, for the most spatially restricted dataset (wind at turbines), lifting functions are essential. Augmenting these sparse measurements with physically motivated states, such as quadratic and cubic rotor speeds and their time averages, boosted P_2 accuracy from 81.89% for the linear case to 88.14% for the fully augmented model. This indicates that when spatial information is scarce, nonlinear temporal states can compensate for the missing flow data.

The NRMSE values generally correlate with the VAF results, with the sparse wind data achieving the lowest overall error at 3.86%. In comparison, the wind field dataset resulted in a higher NRMSE of 4.85%. While the wind at turbines set initially exhibited the highest error (6.93%), the application of 24 Koopman lifting functions successfully reduced the NRMSE to 4.92%. This performance closely approaches that of the full wind field, confirming that nonlinear temporal lifting is a valuable tool for minimizing prediction deviations when spatial resolution is limited. Figure 3 visualizes the performance of the identified models. While the VAF values for the wind field and sparse wind data models are close, the latter produces a time series that is visually more consistent with the reference data. The first three subplots of Figure 3 show the yaw control input and individual turbine power P_1 and P_2 , while the bottom subplot depicts the total farm power P_{WF} . The wind field models track rapid transients effectively but exhibit unrealistic fluctuations during time periods where the power output P_2 is higher due to the yaw offset of the first turbine, specifically between 100 s and 500 s. These oscillations likely stem from localized turbulent noise in the dense grid. The sparse models provide smoother estimates while maintaining relatively good tracking accuracy of the power transients occurring as a result of the yaw step changes. However, it should be noted that these models exhibit slightly more damped dynamics during sudden input changes (e.g., at 650 s and 950 s). Nevertheless, they achieve a lower NRMSE by avoiding the noisy response during the power plateaus observed in the full wind field model response.

The results in Figure 3(b) illustrate the critical role of temporal lifting when data is restricted to turbine locations. Models relying solely on rotor-disk measurements without additional Koopman features deviate significantly during transients. However, by incorporating 24 lifting functions on the rotor speeds, the model accuracy is improved and the two quantification criteria are nearly as good as for the models built on the whole wind field. This confirms that when spatial resolution is limited, nonlinear temporal lifting is a valuable tool for minimizing prediction deviations and capturing the characteristic time delays of the wake.

While both the SOWFA training and verification datasets utilize 0° to 20° yaw step changes

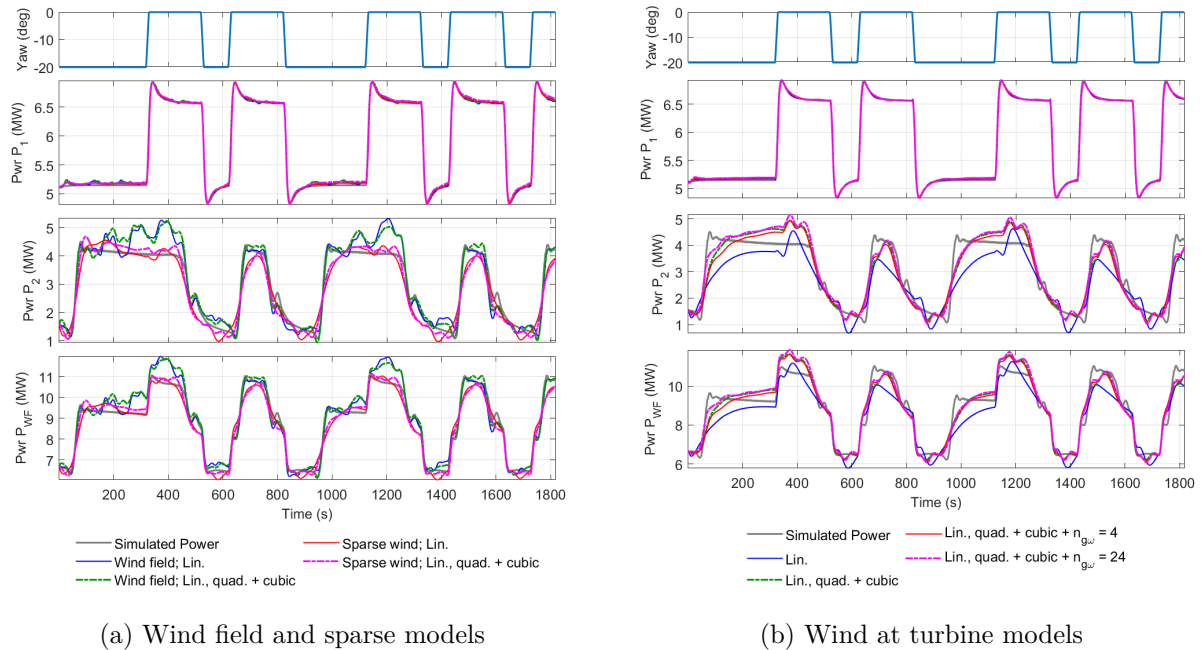


Figure 3. Power prediction verification over 1800s. Grey solid lines denote SOWFA ground truth data. (a) Comparison between the 51,543-point grid (blue and green, existing) and the proposed 84-point sparse measurements (red and pink, proposed). (b) Performance of measurements at the rotor disks, demonstrating accuracy gains from applying lifting functions (green, red and pink) compared to the baseline using raw wind data and rotor speed only (blue).

rather than broad-spectrum excitation, this approach reflects traditional wind farm control where turbines realign based on time-averaged wind directions. The 20° amplitude was selected to induce significant, observable wind field and power transients. Future work should evaluate the performance of these Koopman models under more varied excitations to further explore their frequency-dependent characteristics.

4.2. Analysis of Measurement Selection and Lifting

A sweep of 1,155 longitudinal sensor combinations was performed to identify the optimal sparse measurement configuration, see the code available in [11]. The selection was based on the highest VAF value achieved using a combination of linear, quadratic, and cubic lifting functions. The search varied sensor locations across three zones: the upstream inflow (fixed at index 1), the near-wake (indices 5–15), the mid-wake (indices 40–60), and the region near the downstream turbine (indices 68–72). Of these combinations, 386 outperformed the full wind field dataset in terms of VAF. The optimal configuration was found at indices 1, 10, 51, and 70. Further implementation details and the complete dataset are available in the repository [11].

Once the optimal longitudinal positions were fixed, the impact of the lateral measurement width was investigated. The discussed results were obtained with the lateral range set to 4–24. It was originally investigated if the lateral range of the turbine rotor would be sufficient. However, the lateral range of 7–21, which covers the rotor, only achieved a value of 89.53% compared to 92.66%. The 4–24 configuration was ultimately selected for the sparse wind dataset, as a further expansion to the 1–28 range yielded an only negligibly higher VAF of 92.85% for the cubic case, but would have increased the total number of measurement points. Consequently, the 4–24 range was chosen as a balance between sensing sparsity and prediction fidelity.

Applying Koopman temporal lifting to the sparse wind data (linear, quadratic, and cubic wind) marginally increased the VAF from 92.66% to 92.74%, while the wind field set remained unchanged at 90.86%. The most significant gain occurred in the quadratic-linear sparse configuration, rising from 92.19% to 92.55%, a negligible improvement of 0.39%.

It should also be noted that the presented sensitivity study does not aim to replicate specific lidar scan patterns. Instead, the investigation focused on determining the minimum number of spatial measurements required to maintain high prediction accuracy and identifying the sensitivity of the model to sensor placement. By isolating these requirements, we demonstrate the potential for significant data reduction in the identification process. This fundamental analysis provides the groundwork for future work, where these optimal measurement zones can be mapped onto realistic sensing hardware, such as scanning lidars or fixed sensor mast arrays.

Future research will focus on three primary directions: (i) integrating past power and rotor speed values into the state vector to better capture transient dynamics; (ii) generating DMD models directly from experimental measurement data; and (iii) evaluating Koopman-based MPC in closed-loop environments such as FAST.Farm. This final direction is particularly important, as previous studies indicate that open-loop metrics like VAF and NRMSE are not always reliable indicators of closed-loop tracking performance. Furthermore, while the current study utilizes high-fidelity SOWFA data, the wind field was modeled without vertical shear to isolate the impact of sensor sparsity. Future research should investigate the validity of the proposed DMD modeling approach under more complex inflow conditions. Utilizing experimental field measurements will be essential to evaluate prediction quality under varying vertical velocity gradients and turbulence intensities, ensuring the results are reproducible in real-world environments.

5. Summary

This paper evaluates Koopman-based power prediction for wind farms under realistic sensing constraints using SOWFA simulation data. By comparing models built on the full wind field and on strategically placed sparse wind data, we demonstrate that a sparse configuration can outperform a full-field model. We conjecture that localized sampling effectively filters turbulent noise present in high-resolution datasets, leading to more physically consistent dynamics. Furthermore, while the wind at turbines dataset initially shows lower fidelity, quadratic and cubic lifting functions applied to the wind data and nonlinear temporal lifting on rotor speeds significantly recovers accuracy, reaching VAF values nearly equivalent to the full wind field model. These results confirm that physically motivated state lifting enables high-fidelity Koopman modeling even with sparse measurements, supporting its viability for real-time wind farm control.

References

- [1] Annoni J, Gebraad P and Seiler P 2016 Wind farm flow modeling using an input-output reduced-order model *2016 american control conference (acc)* (IEEE) pp 506–512
- [2] Chen Z, Doekemeijer B M, Lin Z, Xie Z, Si Z, Liu J and van Wingerden J W 2020 Data-driven modeling & analysis of dynamic wake for wind farm control: A comparison study *2020 Chinese automation congress (CAC)* (IEEE) pp 5326–5331
- [3] Cassamo N and van Wingerden J W 2020 On the potential of reduced order models for wind farm control: A koopman dynamic mode decomposition approach *Energies* **13** 6513
- [4] Cassamo N and van Wingerden J W 2021 Model predictive control for wake redirection in wind farms: a koopman dynamic mode decomposition approach *2021 American Control Conference (ACC)* (IEEE) pp 1776–1782

- [5] Churchfield M, Lee S, Moriarty P, Martinez L, Leonardi S, Vijayakumar G and Brasseur J 2012 A large-eddy simulation of wind-plant aerodynamics *50th AIAA aerospace sciences meeting including the new horizons forum and aerospace exposition* p 537
- [6] Boersma S, Doekemeijer B, Vali M, Meyers J and van Wingerden J W 2018 A control-oriented dynamic wind farm model: Wfsim *Wind Energy Science* **3** 75–95
- [7] Sharan B, Dittmer A and Werner H 2022 Real-time model predictive control for wind farms: a koopman dynamic mode decomposition approach *2022 European Control Conference (ECC)* pp 1006–1011
- [8] Dittmer A, Sharan B and Werner H 2022 Data-driven adaptive model predictive control for wind farms: A koopman-based online learning approach *2022 IEEE 61st Conference on Decision and Control (CDC)* (IEEE) pp 1999–2004
- [9] Dittmer A, Sharan B and Werner H 2023 Koopman model predictive control for wind farm yield optimization with combined thrust and yaw control *IFAC-PapersOnLine* **56** 8420–8425
- [10] Cassamo N 2020 Code for thesis Koopman Dynamic Mode Decomposition for Wind Farm Control https://github.com/nassircassamo/IODMD_SOWFA
- [11] Dittmer A 2026 Code for: A Simulative Analysis of Physically-Motivated Koopman Lifted States https://github.com/antjedittmer/2025_SOWFA_Koopman
- [12] Kaiser E, Kutz J N and Brunton S L 2020 Data-driven approximations of dynamical systems operators for control *The Koopman Operator in Systems and Control* (Springer) pp 197–234
- [13] Arbabi H, Korda M and Mezić I 2018 A data-driven koopman model predictive control framework for nonlinear partial differential equations *2018 IEEE Conference on Decision and Control (CDC)* (IEEE) pp 6409–6414
- [14] Proctor J L, Brunton S L and Kutz J N 2018 Generalizing koopman theory to allow for inputs and control *SIAM Journal on Applied Dynamical Systems* **17** 909–930