Received: 17 September 2022 | Revised: 20 August 2023 | Accepted: 16 September 2023

DOI: 10.1002/we.2874

SPECIAL ISSUE PAPER

WILEY

Predicting wind farm operations with machine learning and the P2D-RANS model: A case study for an AWAKEN site

Coleman Moss[1] | Romit Maulik[2,3] | Patrick Moriarty[4] | Giacomo Valerio Iungo[1]

[1]Department of Mechanical Engineering, Wind Fluids and Experiments (WindFluX) Laboratory, The University of Texas at Dallas, Richardson, Texas, USA

[2]Pennsylvania State University, State College, Pennsylvania, USA

[3]Argonne National Laboratory, Lemont, Illinois, USA

[4]National Renewable Energy Laboratory, Golden, Colorado, USA

Correspondence
Giacomo Valerio Iungo, Department of Mechanical Engineering, Wind Fluids and Experiments (WindFluX) Laboratory, 800 West Campbell Rd, WT 10, Richardson, TX 75080-3021, USA.
Email: valerio.iungo@utdallas.edu

Funding information
National Renewable Energy Laboratory; Alliance for Sustainable Energy, LLC, Managing and Operating Contractor for the National Renewable Energy Laboratory for the U.S. Department of Energy; National Science Foundation (NSF) CBET, Fluid Dynamics CAREER program, Grant/Award Number: 2046160

Abstract

The power performance and the wind velocity field of an onshore wind farm are predicted with machine learning models and the pseudo-2D RANS model, then assessed against SCADA data. The wind farm under investigation is one of the sites involved with the American WAKE experimeNt (AWAKEN). The performed simulations enable predictions of the power capture at the farm and turbine levels while providing insights into the effects on power capture associated with wake interactions that operating upstream turbines induce, as well as the variability caused by atmospheric stability. The machine learning models show improved accuracy compared to the pseudo-2D RANS model in the predictions of turbine power capture and farm power capture with roughly half the normalized error. The machine learning models also entail lower computational costs upon training. Further, the machine learning models provide predictions of the wind turbulence intensity at the turbine level for different wind and atmospheric conditions with very good accuracy, which is difficult to achieve through RANS modeling. Additionally, farm-to-farm interactions are noted, with adverse impacts on power predictions from both models.

KEYWORDS
machine learning, RANS, SCADA data, wind farm, wind turbine

1 | INTRODUCTION

The efficient extraction of energy from wind is a challenging engineering problem that has been advanced to support more affordable wind energy production and larger global wind energy consumption.[1,2] While improvements in energy extraction are available in many facets of wind turbine design, such as blade material,[3] aerodynamic design,[4] turbine control,[5] and fault monitoring,[6] this work focuses on methods to quantify and, eventually, optimize effects due to the deployments of wind turbines in clusters, which are known as wind farms.[7,8]

Since wind turbines generate power by removing kinetic energy from the wind, the wind available for downstream turbines can have decreased speed and enhanced turbulence intensity, *TI*, compared to the upstream wind condition. This phenomenon is referred to as a wind turbine wake.[9,10] Turbines downstream of operating turbines experience this reduction of available kinetic energy and produce less power as a result.[11] Thus, the layout of wind farms must be designed in such a manner that turbine-to-turbine wake interactions are minimized. Wakes, however, are highly three-dimensional, vary with incoming wind conditions, specifically with atmospheric stability, and modify the downstream flow conditions, such as turbulence intensity.[12–14] Though slowing the wind speed impacts turbine power production, the increase in turbulence intensity due to wakes also impacts power production[15,16] as well as the fatigue loading on the turbine blades,[17] thus impacting turbine reliability.

Therefore, to optimize the layout of a wind farm, whether for power production, turbine longevity, or an optimal combination of the two, the propagation of wind turbine wakes must be predicted as well as their impact on power capture. These predictions must be repeated for varying environmental conditions given the variability of wakes to incoming flow conditions.[18,19] The large number of cases that must be predicted to design a wind farm demands accurate models with low computational costs.

Of the approaches to modeling wind farms, one successful approach is large eddy simulation (LES).[17,20,21] While LES can reproduce the complex, time-varying evolution of wind turbine wakes, the computational cost can be prohibitively high.[22] Using the Reynolds-averaged Navier–Stokes (RANS) modeling approach is another popular option to simulate wake propagation in wind farms,[23–25] with run times three or four orders of magnitude faster than LES.[22]

While LES and RANS models are physics-based, a promising alternative is machine learning (ML) modeling. ML methods are data-driven, though they can be coupled with physics-based models.[26] Fluid dynamics in general, and wind energy in particular, have many applications that are prime candidates for ML studies.[27] For instance, physics-informed neural networks can be used to quickly solve the governing equations for several fluid flows.[26] In atmospheric research, ML has been used for various applications such as turbulence predictions[28] and wind forecasting.[29] A suite of ML tools has been applied to many elements of the wind energy sector, such as modeling wind turbine wakes[30–33] and modeling wind turbine power capture.[34–37] To model the spatial variability of the wind field, most works rely on high-fidelity CFD simulations or light detection and ranging (LiDAR) measurements.[38,39] Prediction of turbine power is also performed in the context of wake steering[40] or in optimal input selection.[41] While information regarding the specific layout of a given wind farm and the position of turbines relative to each other can be captured by ML models, building ML-based modeling approaches that generalize to arbitrary turbine layouts is difficult without injecting physics-based models.[42] Other approaches to generalizing models involve using unique geometric parameters or advanced graph neural network models.[43,44] Generally lacking, however, are detailed predictions of individual turbine wind speed and *TI* using models trained on SCADA data, which are important to further understanding real farm performance.

With the above in mind, the purpose of this paper is to provide a case study using an experimental site of the American WAKE experimeNt (AWAKEN)[45] to compare the performance of RANS and ML models in predicting the wind field evolution through the wind farm, wake interactions, and power capture for different wind and atmospheric conditions. Specifically, a depth-averaged RANS model is used, called pseudo-2D RANS (P2D-RANS), which has been tuned and validated on a different wind farm.[46] The ML models used will be uniquely generated for the site under consideration, and the performance of the two approaches will be compared and discussed. Finally, potential power losses associated with neighboring wind farms will be discussed.

The remainder of the paper is organized as follows. Section 2 gives an overview of the experimental site and the data available for this analysis. Section 3 discusses the RANS solver used and the development of RANS surrogate models for faster solutions as well as the ML models used. The models are then assessed against Supervisory Control And Data Acquisition (SCADA) data of the wind farm in Section 4, focusing on the impact of wakes and farm-to-farm interactions on model accuracy. Section 5 offers concluding remarks on this study. For the interested reader, Appendices A and B contain details on the selection and tuning of ML models for SCADA-data filtering and modeling wind turbine performance, respectively.

2 | DATA SET OVERVIEW

2.1 | Wind farm under investigation

The wind farm under consideration is the King Plains wind farm in Oklahoma, USA. The wind farm consists of 88 turbines installed over relatively flat terrain with the layout shown in Figure 1A. Turbine specifications are withheld due to a standing nondisclosure agreement. SCADA data are available in averages and standard deviations calculated over 10-min periods. Available parameters of note to this project are wind speed

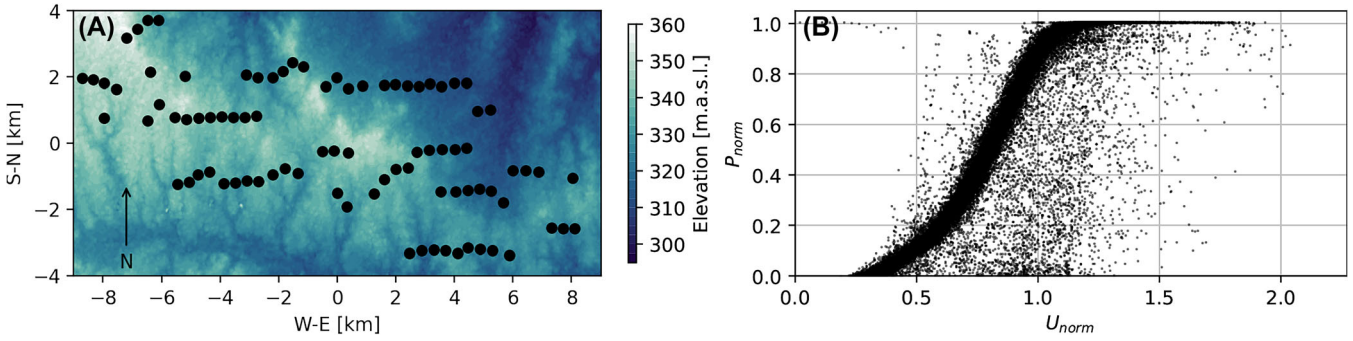

FIGURE 1 King Plains wind farm: (A) layout; (B) raw power curve for a sample wind turbine.

corrected by the manufacturer's nacelle transfer function, turbulence intensity (*TI*), turbine yaw angle, wind direction offset from yaw angle, and power. SCADA data are available for all turbines from 1 December 2020 until 31 December 2021. Figure 1B illustrates raw SCADA data in wind speed and power for a sample turbine from the farm. Wind speed and power are normalized by rated wind speed and power, which are also withheld. No meteorological tower is available for the site.

An important feature of this site is the presence of neighboring wind farms. Directly to the south lies a large wind farm at a distance of approximately 2.25 km that has the potential to induce slow-downs in the westerly turbines of the farm. Additionally, to the west, northwest, and north lie large farms at distances of 5.5, 16, and 15 km, respectively. Figure 2 illustrates the relative positions of these neighboring farms. The potential for farm-to-farm interactions will be investigated, as well as any impacts such interactions may have on turbine power production and model error.

To investigate the climatology of the site, as well as the modifications on the incoming flow field caused by the operation of the turbines, freestream measurements are necessary. Since there is no meteorological tower for this site, the freestream measurements are estimated by using reference conditions.[11] Using the International Electrotechnical Commission (IEC) standard for defining waked sectors of turbines,[47] the reference wind speed, *TI*, and wind direction are defined as the average of turbine wind speeds, *TI*s, and wind directions for all unwaked turbines, rejecting turbines with points farther than two standard deviations from the mean of the parameter under consideration. With the reference conditions defined, the climatology of the site can be investigated, as in Figure 3.

From Figure 3A,B, it is observed that reference wind speed and *TI* both follow fairly typical distributions. Figure 3C demonstrates a daily cycle in atmospheric stability while Figure 3D shows that south and south-south-east winds tend to dominate, with some winds coming from the north. Thus, for most of the farm's operating conditions, farm-to-farm interactions are minimized or restricted to a specific sector of the total farm. These interactions will be discussed later in Section 4.3. In general, the results of this climatology analysis closely follow those of a previous study at the Southern Great Plains ARM Observatory near the farm under consideration.[48]

2.2 | SCADA data filtering

In support of building accurate ML models, filtering SCADA data is considered. Since ML models are data-driven, outliers in the data have the chance to dilute the training data and negatively impact model performance. Additionally, for certain ML models, only specific regions of turbine operation may be of interest. Thus, the modeling task is simplified by rejecting unneeded regions of turbine operation. This region removal is also considered.

Starting with region removal, in the SCADA data there are numerous points with negative wind speed or power. Furthermore, the turbine yaw angle exceeds 360° or falls below 0° for some points. Therefore, the first step is to keep only points with positive wind speed and power and yaw angle between 0° and 360°. Several options for additional region removal are now available. If the interest is to model turbine behavior in an ideal state, waked sectors can be removed following the IEC standard.[47] If the impact and evolution of wakes are of interest, limiting the study to region two of the power curve would be effective, since wake impacts are minimal once the wind speed exceeds the rated wind speed and power production is fixed.[42,49] The inclusion of region one, or the region below the cut-in wind speed for the wind turbines, can be used to prevent

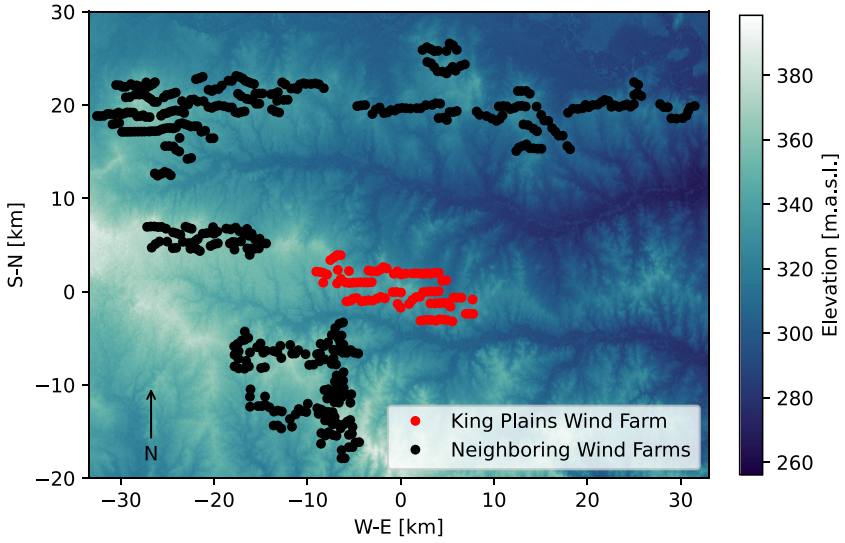

FIGURE 2 King Plains and neighboring wind farm layouts.

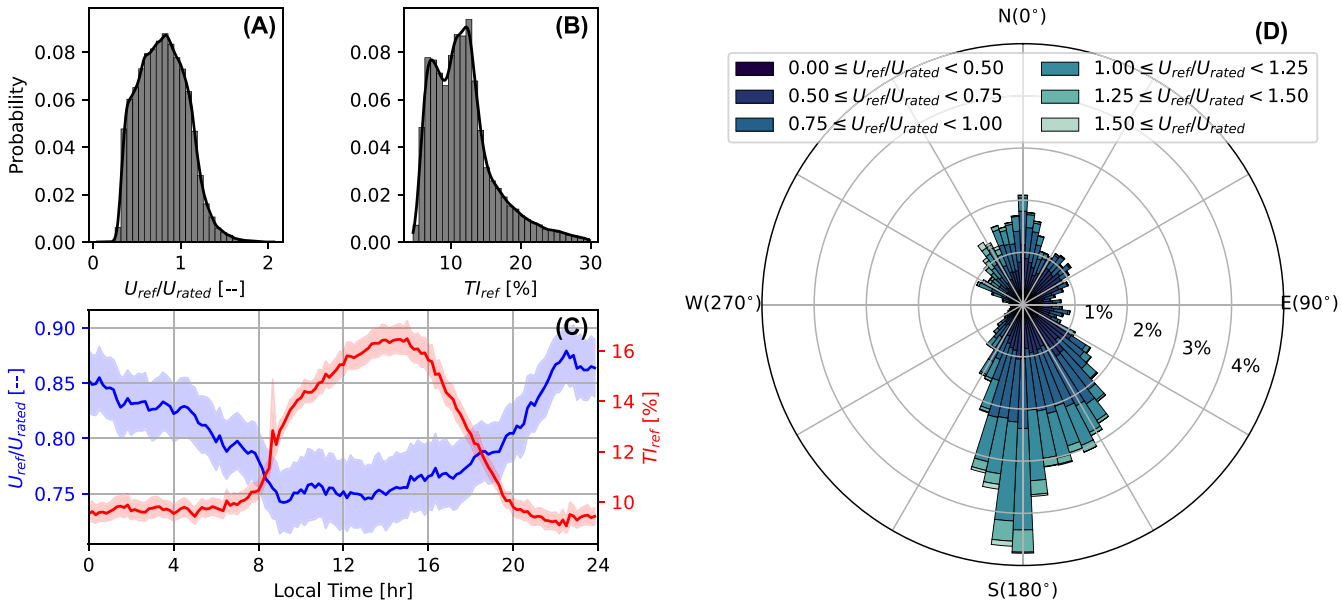

FIGURE 3 Climatology for the King Plains wind farm: histogram of reference wind speed normalized by turbine rated wind speed (A) and reference *TI* (B) with the kernel density of the estimated PDF superimposed; (C) daily cycle of reference wind speed normalized by turbine rated wind speed and reference *TI* with (95%) confidence interval calculated via bootstrapping; (D) windrose of reference wind direction and normalized reference wind speed.

boundary issues at low wind speeds, or it can be rejected to avoid including the more random power production in that region. As each region removal approach has a specific use, the approaches will be discussed in detail when needed.

Even after the rejection of non-physical samples, SCADA data can still have outliers in power that could be due to sensor fault, curtailment, maintenance, degradation of components, or other causes. For accurate assessments of power capture potential, as well as to avoid inducing possible model error, several filtering approaches are used to remove these points. Two ML filters previously proposed are implemented exactly as in the original works: an automatic Gaussian Process (GP) filter,[50] referred to as the GP filter, and a k-means clustering filter,[51] referred to as the k-means filter. The GP filter works by iteratively fitting a GP model to the power curve and rejecting points that fall too far from the fitted curve. On the other hand, the k-means filter works by clustering the data using the k-means clustering algorithm and rejecting points that fall too far from the cluster centroid by using the Mahalanobis distance to measure the distance between a given point and the cluster centroid. Furthermore, a simple statistical filter is applied. This filter bins the data into 0.5 m/s wide bins in wind speed. For each bin, the mean and standard deviation of power is calculated, and points further than two standard deviations from the mean power are rejected. This filter is referred to as the statistical filter.

Additionally, a novel filter is proposed also using ML following a similar approach to the GP filter. The filter trains an ML model over the data set to be filtered, using the entire data set. The inputs can be freely selected but the output is fixed as power. Then, the model is used to predict power for every point in the data set. The absolute difference between true and predicted power is ascribed to each point as an error. For the entire data set, the mean and standard deviation of error is calculated. Points with an error greater than a given multiple of standard deviations from the mean error are rejected. The model is then retrained over the cleaned data set and the process is repeated until a maximum number of iterations is reached or until the number of rejected points falls below a certain threshold. This filter assumes that the ML model is capable of capturing the governing physical phenomena in power production relating the variations of the inputs to the variation in the power output. As such, points that the model cannot predict with reasonable accuracy can be assumed to belong to different physical phenomena, such as the various causes of outliers already mentioned. The result of the iterative process is a data set containing points all belonging to the same physical phenomena, to some degree of certainty. This enables future models to learn more precisely the phenomena of interest.

To generate an accurate ML filter, the ML model selected for use in the filter should be capable of accurately predicting wind turbine power. Furthermore, ML filters need to have their hyperparameters tuned for optimal performance. For this application, a Random Forest (RF) model is found to be most suitable. For more details on the selection and tuning of this model, the reader is directed to Appendix A where Deep Neural Network (DNN) and RF models are compared for filtering, with the results being reported in Table A1.

After selecting the RF model and tuning its hyperparameters for optimal performance, it is used in the ML filtering approach, termed hereafter as the RF filter. All filters are applied to all turbines. Since the purpose is to prepare the turbine data sets for turbine models that will focus on performance during wake interactions, namely, during operations in region two, the turbine data sets are first cleaned using a region removal

rejecting wind speeds outside of region two. A portion of region three is included to give the model the appropriate information to capture the transition between regions two and three. Wakes are included since eventual turbine models will be trained and probed over regions including wakes. Statistics on the percentage of each turbine data set that is rejected for each filter are reported in Table 1 and the results of applying the filters to a sample turbine are shown in Figure 4.

From Figure 4A, it is clear that the statistical filter struggles to reject outliers given the large number of outliers, and thus the large standard deviations for each bin. Likewise, the k-means clustering filter in Figure 4B also struggles to reject all outliers. For both binning approaches, the standard deviations per bin are simply too high to allow for accurate filtering. The automatic GP filter in Figure 4C does a better job of removing outliers far from the power curve but still leaves some outliers near the curve. Finally, the RF filter in Figure 4D does the best job of rejecting the most outliers but could be too aggressive. Comparing these results with the results in Table 1, this supposition is borne out. The statistical and k-means filters have lower rejection rates, suggesting that they are not removing enough outliers when cross-referencing with the figure. Compared to the GP filter, the RF has significantly higher rejection rates.

While RF certainly is the most aggressive of the filters, it does the best job of removing outliers, even if it risks removing inliers as well. Since the data sets available are large, even the larger rejection rates of the RF filter likely do not threaten to reduce the data set size below a usable level, especially since the median rejection rate is only (15%). Furthermore, since the basic premise of the RF filter is to ensure that all points are predictable by ML methods to a certain level, it makes sense to use this filter to support ML models. Thus, the RF filter is selected for use.

3 | MODELING

In this section, the process of setting up the P2D-RANS model [46] for the wind farm under examination, as well as the suite of ML models necessary to reproduce farm operations, is considered. First, the RANS approach is discussed, including defining the initial set of environmental conditions to solve using the RANS solver, building surrogate models on top of the RANS results, and using intelligent resampling to improve the surrogate models. To build an ML modeling approach for the entire farm, however, it is necessary to predict the wind speed and *TI* at each turbine from the reference conditions, then predict turbine power from the predicted turbine wind speed and *TI*. The optimal ML model for each step is determined via cross-validation analysis across the farm, and the final pipeline is proposed.

3.1 | P2D-RANS simulations and surrogate modeling of the wind farm

To predict the performance of the entire wind farm, and especially to compare the model predictions against real SCADA data, a model must be able to simulate the farm performance on the input domain, in this case, represented by wind speed, *TI*, and wind direction, with a relatively high resolution. When using the P2D-RANS model, the user may set the freestream wind speed, direction, and *TI*, and then solve the wind field for the farm. Thus, if the P2D-RANS model is to be assessed against the SCADA data directly, it must solve every case in the SCADA data set. Even though the computational cost for the P2D-RANS model to solve a single case is relatively low, typically 5 min or less on consumer desktop hardware, these costs quickly compound when many thousands of cases must be solved. Thus, rather than using the P2D-RANS model directly, the P2D-RANS model will be used to solve a few selected input cases completely. Then, an RF surrogate model will be trained on top of the P2D-RANS solutions. Thus, farm performance for arbitrary inputs can be simulated using the RF surrogate model at a much lower computational cost than solving every case directly using the P2D-RANS model.

This approach creates an optimization problem: what is the smallest number of cases in wind speed, wind direction, and *TI* that should be sampled by executing P2D-RANS simulations to train an accurate surrogate model? Furthermore, some regions of the input domain experience greater variability in farm performance than others. For instance, when the wind direction is along the east-west line, the wake interactions are stronger than when the wind direction is along the north-south direction and more wake interactions occur at a given time, thus causing increased variability in farm power. These regions will likely require more sampled points to accurately recreate performance using the surrogate model. Therefore, the number of sampled points as well as their distribution in the input domain both need to be optimized.

TABLE 1 Statistics of rejection rates for all filters applied to all turbines.

Filter	Min	25th percentile	Median	75th percentile	Max
Statistical	4.25%	4.94%	5.10%	5.36%	6.65%
K-means	1.77%	3.32%	3.71%	3.98%	4.73%
GP	4.65%	6.07%	6.93%	7.60%	15.10%
RF	9.52%	13.17%	14.55%	15.63%	23.83%

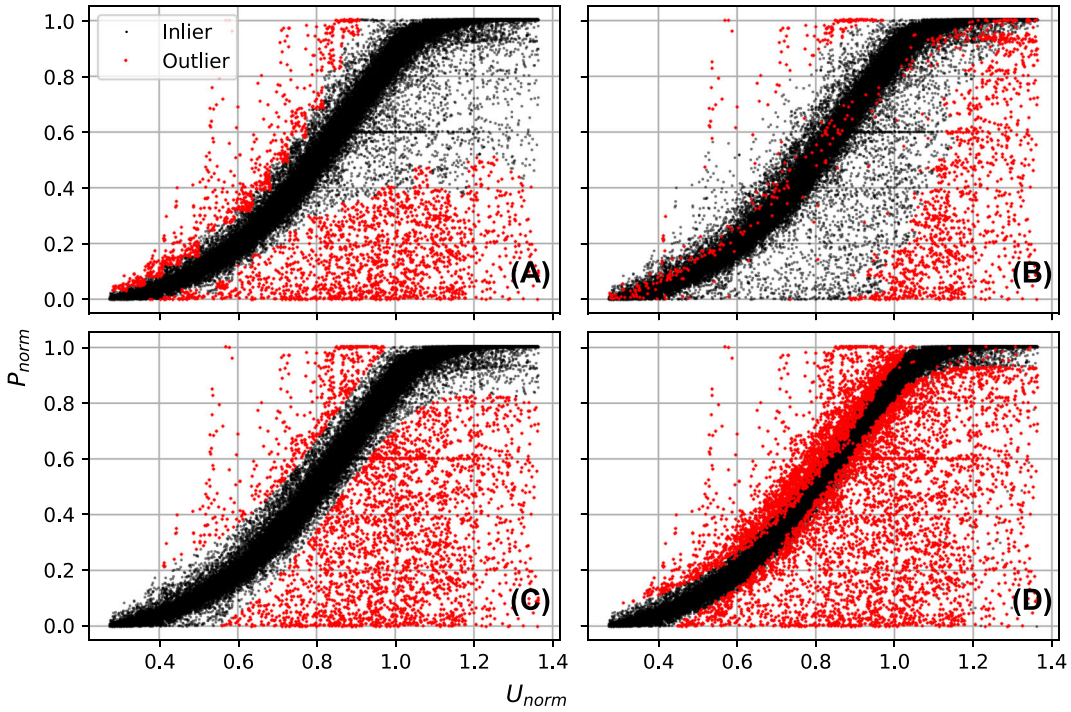

FIGURE 4 Application of (A) statistical, (B) k-means clustering, (C) automatic GP, and (D) automatic RF filters to a sample turbine.

As a first step, an initial test subset of the input parameters to be solved with the P2D-RANS model is defined. This subset can be defined manually or automatically using Latin hypercube sampling (LHS).[52,53] Since the interest of modeling is the wake interactions between turbines, the wind speed domain is restricted to region two only. Any transition between regions two and three is not needed as there is no threat of boundary issues in the RANS solution. A manual subset of the input parameters is defined with steps in wind speed of 1 m/s width, while *TI* inputs are (2.5%), (7.5%), (12.5%), (20%), and (30%), manually selected to give a higher resolution at lower *TI* values where wake interactions are more severe. Finally, wind direction is varied between 0° and 360° in steps of 15°. While it is known that wake effects take place within smaller sectors in wind direction,[47] this angular resolution was specifically chosen to give a coarse initial sampling to be improved upon with additional sampling. By making all possible combinations of the three input parameters, 1320 cases are defined. Furthermore, LHS is performed by setting the limits on wind speed, direction, and *TI* to be identical to the manual selection, then sampling 1320 cases randomly in the domain to keep the same number of points as the manual selection.

To run simulations of the down-selected cases, the P2D-RANS model is implemented exactly as in the original work.[46] The model was tuned using LiDAR measurements at a different wind farm. These tuned parameters, however, are not currently available at the King Plains site. Thus, a potential source of error is the lack of site-specific tuning for the P2D-RANS, which is applied as-is while adjusting for the specific turbine diameter and rated power. However, given that the terrain at both the training farm and the farm under consideration is relatively flat and that the turbine manufacturer is the same in both cases, any error due to a lack of tuning should be minimal. The lack of tuning is an appropriate allowance since generating a more accurate model would require more costly simulations, which makes the ML models more compelling from a computational standpoint. The un-tuned P2D-RANS is thus a good trade-off between computational cost and potential inaccuracy.

The settings of the P2D-RANS model were selected through a sensitivity analysis to provide good accuracy while maintaining reasonable computational cost. For any given case, however, convergence is not guaranteed. Several cases failed to converge when sampling the manual and LHS subsets. The P2D-RANS settings were therefore tuned iteratively to achieve convergence for as many points as possible. After this tuning, 1232 samples of the manual subset and 1225 points in the LHS subset converged during the P2D-RANS solution. While the P2D-RANS model solves the wind field over the entire farm, only the wind speed one diameter upstream of each turbine is sampled and recorded, as well as the turbine power, obtained by interpolating the sampled wind speed on the power curve. Since no manufacturer power curve is available, and to achieve better accuracy, the power curve used was the IEC binning curve[47] using RF-filtered data aggregated across all turbines. As a result, for each simulation, the farm power as well as individual turbine powers and wind speeds are available. The wake loss calculated using the RANS model over the wind farm as well as the individual turbine wind speeds, from which power can be calculated, for a sample case is illustrated in Figure 5.

First, the different sampling methods are compared to determine which method provides a better set of input points for the surrogate model. To investigate this, two RF models are used to predict total wind farm power for the inflow wind speed, direction, and *TI*. DeepHyper is used to

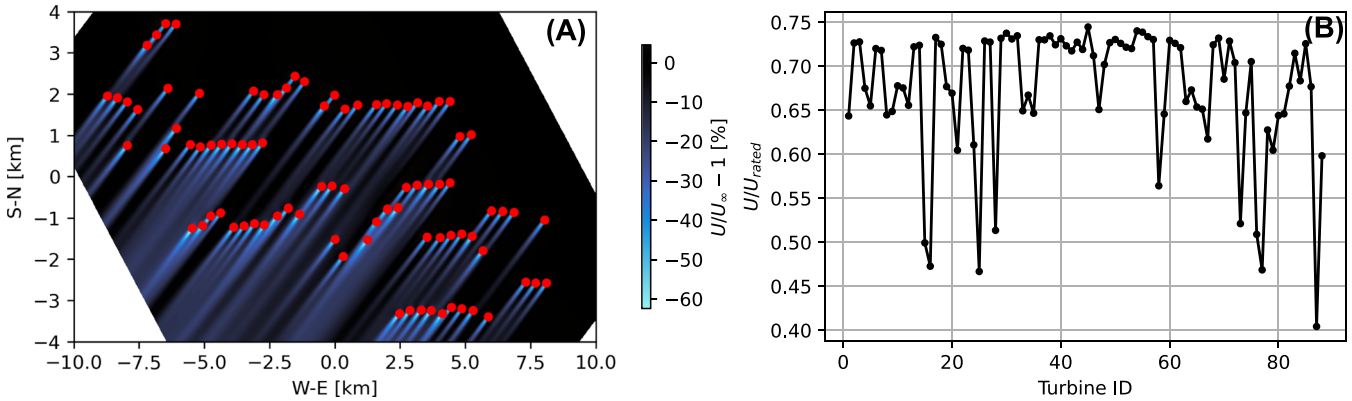

FIGURE 5 P2D-RANS simulation for a normalized wind speed of 0.73, wind direction of 50°, and *TI*= 10%. (A) Wind farm velocity field. (B) Rotor average wind speed for the various wind turbines.

optimize the hyperparameters of these models.[54] One model uses the manual data subset while the other uses the LHS one. Using five-fold cross-validation, the manual subset has a percent root-mean-squared error (PRMSE) of 5.319%, a percent mean absolute error (PMAE) of 3.909%, and an R^2 of 0.683. In this case, PRMSE and PMAE are simply the root-mean-squared error and mean absolute error normalized by the turbine-rated power. By contrast, the LHS subset has a PRMSE of 3.103%, a PMAE of 1.975%, and an R^2 of 0.981. This means that the LHS model generalized better when asked to predict new samples than the manual subset model did. For this reason, when additional sampling is needed hereafter, LHS is used. To take advantage of all sample points, however, the manual subset and LHS subset are combined, a new RF surrogate model is optimized, and the same five-fold cross-validation is performed. In this case, the model had a PRMSE of 2.256%, a PMAE of 1.289%, and an R^2 of 0.986. Thus, though the manual subset does not generalize well on its own, the surrogate model benefits more from the addition of the manual samples to the LHS points than it is negatively affected by the addition of those samples.

Now that the surrogate model is defined, the error of the model is investigated to enable intelligent resampling of the input domain and improvement of the model accuracy. To do this, the surrogate model trained on the manual and LHS subsets is used to predict the farm power for all the training samples. This farm power is compared against the RANS farm power, and the absolute percent difference between the two is ascribed to each sample as an error. Then, an additional RF surrogate model is optimized to predict the original model's error from the input points. This error surrogate model is then probed to generate Figure 6A–D. The error in the power surrogate model is larger when the wind direction is perpendicular to the prevailing wind directions for the farm, that is when the wind direction is along the east-west line.

In an attempt to reduce the power surrogate model's error, the input domain is resampled by testing additional 1000 cases with the P2D-RANS model. LHS is used and is allowed to sample over the entire wind speed and *TI* range. 500 points are sampled between wind directions of 70° and 115° and another 500 points are sampled between wind directions of 250° and 300°. These additional points are combined with the initial manual subset and initial LHS subset and a new power surrogate model is optimized and trained. This new model has a PRMSE of 1.495%, PMAE of 0.959%, and R^2 of 0.9947 after cross-validation. Training another error surrogate model on the new power surrogate model and probing generates the plots in Figure 6E–H. Compared to Figure 6A–D, the peak error is reduced. Additionally, as *TI* increases, the error bands associated with wind directions along the east-west line decrease for higher wind speeds. Finally, an error band emerges for all wind directions at low wind speeds.

The second and final resampling is done with an additional 1000 LHS points. These points are split between two regions. The first region receives 300 points and allows all wind speeds, *TI* below 7.5%, and the same wind direction bands as the first resample. The second region receives 700 points for wind speeds under 5 m/s, all wind directions, and *TI* above 7.5%. These resampled points are added to all preceding points and a final power surrogate model is trained, along with a corresponding error model. This final error model results in the plots in Figure 6I–L. Once again, the error is reduced compared to Figure 6E–H, with the low wind speed band especially showing improvement. The final power surrogate model has a PRMSE of 1.499%, PMAE of 0.961%, and R^2 of 0.9946 from cross-validation, showing that the improvements of the additional sampling were greatly reduced. Thus, no additional points are sampled, and the current set of points is used.

All samples are used in training despite the third resampling appearing to reduce the accuracy of the surrogate model. This is because, in general, ML models improve when trained on larger data sets. While the full set of all sampled points has a lower score, it also includes a higher proportion of points specifically chosen because they represent challenging cases to predict. Thus, it should be no surprise that adding additional challenging cases could eventually decrease the model's performance. However, since the decrease in the score is very marginal, it seems better to include additional points rather than stopping after the first resample.

Finally, it is investigated whether modeling the wind farm power directly with a single surrogate model is more or less accurate than modeling each turbine and then summing across the models to get the total farm power. In this case, since individual turbine power is a result of the P2D-

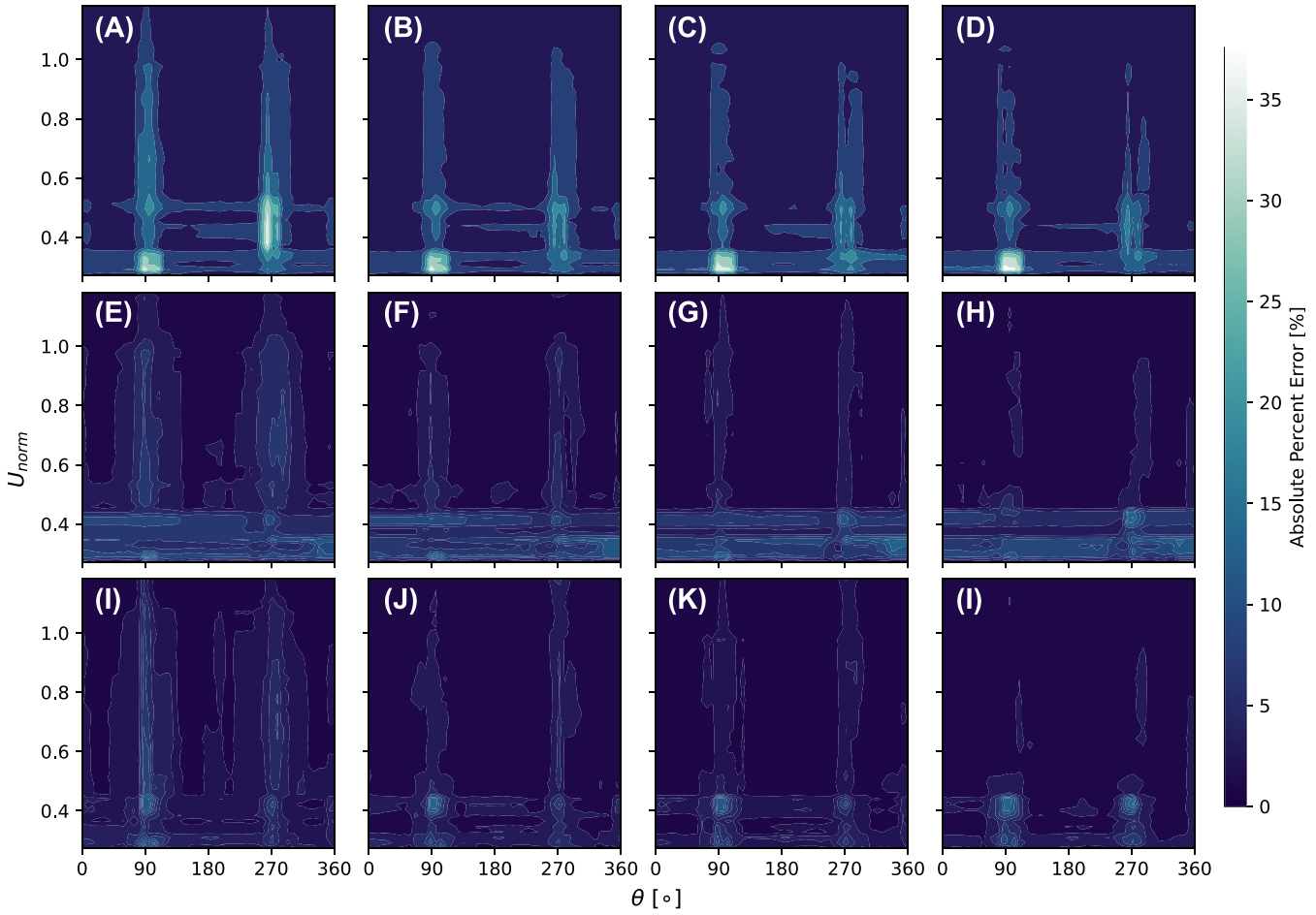

FIGURE 6 Predicted surrogate model error for models trained on the manual subset and first LHS subset (A–D), trained on initial data with first resample (E–H), and trained on initial data with first and second resample (I–L), varying *TI* at 2.5% (A, E, I), 7.5% (B, F, J), 12.5% (C, G, K), and 20% (D, H, L).

RANS simulation, each turbine can be modeled directly with power as a function of the inflow conditions. RF models are used for surrogate modeling but they are not optimized on a turbine basis, as the improvement from the individual optimization is negligible. Performing five-fold cross-validation as prior results in a PRMSE of 0.750%, PMAE of 0.392%, and an R^2 of 0.9989. The results of all the initial manual and LHS subsets individually, combined, and adding in all the additional samples are summarized alongside the individual turbine results in Table 2. To model farm power, modeling each turbine individually is the best approach, and will be adopted hereafter.

3.2 | Machine learning modeling of the wind farm

ML modeling of wind farm performance from reference conditions encompasses three different models for each turbine. First, a unique ML model is used to predict each turbine's wind speed from the reference wind speed, direction, and *TI*. Then, another ML model is used to predict turbine *TI* using the reference wind conditions along with the predicted turbine wind speed. Finally, the last ML model predicts turbine power from predicted turbine wind speed and *TI*. By chaining these models together, many levels of turbine interactions can be detected.

First, the model for turbine wind speed must be selected. Common machine learning models, namely the Gaussian process (GP), support vector machine (SVM), deep neural network (DNN), and RF models, are considered again here. The DNN models are built using Keras with the TensorFlow backend[55] while the GP, SVM, and RF models all use Scikit-Learn,[56] all in Python. The wind speed models are trained on raw SCADA data. The data is not filtered since the filtering methods require either statistical similarity of the data set or a definition of nominal behavior, such as a power curve, which can be used to detect outliers. Yet in the case of the wind speed measurement, wake interactions increase wind speed variability and make statistical filtering unfeasible. On the other hand, nominal behaviors are not easy to define for wind speed at the individual turbine level. Thus, the data is left unfiltered.

TABLE 2 Five-fold cross-validation results for P2D-RANS surrogate models.

Method	PRMSE [%]	PMAE [%]	R^2 [-]
Manual subset	5.319	3.909	0.693
LHS subset	3.103	1.975	0.981
Manual + LHS	2.256	1.289	0.986
First resample	1.495	0.959	0.9947
Second resample	1.499	0.961	0.9946
Individual turbines	0.750	0.392	0.9989

Five-fold cross-validation is applied to all 88 turbines and the average PRMSE, PMAE, normalized mean absolute error (NMAE), and R^2 for each turbine is calculated. PRMSE and PMAE are identical to the quantities defined for the surrogate model training in Section 3.1, while the definition of the NMAE is detailed in Appendix A. From this analysis, the RF model is found to perform best. For the exact scores of the models, the reader is referred to Table B1 in Appendix B.

Subsequently, turbine *TI* is predicted from reference wind speed, direction, and *TI*, as well as the local predicted wind speed. For training, the real local turbine wind speed is used to avoid mixing error sources. When applied to farm simulations, however, this local wind speed will be predicted by the wind speed model. Five-fold cross-validation is applied to all 88 turbines and the average PRMSE, PMAE, NMAE, and R^2 for each turbine is calculated. Across the farm, the minimum, 25th percentile, median, 75th percentile, and maximum of each of these metrics are reported in Table B2 in Appendix B for the GP, SVM, DNN, and RF models, as for the wind speed models previously. Once again, the RF model is selected as the best-performing model.

To develop models to predict turbine power from local turbine wind speed and *TI*, which will be provided by the wind speed and *TI* models discussed above, the data being used are filtered as discussed in Section 4.1. Before selecting an ML model for power by cross-validation across the farm, the RF and DNN models are optimized using DeepHyper, taking advantage of the fact that power data can be aggregated across all turbines and a single, optimal model be defined. After optimizing both the RF and DNN models, five-fold cross-validation is applied across the farm using the GP, SVM, RF, and DNN models. The results of the cross-validation are reported in Table B3 in Appendix B.

From this analysis, the effects of filtering are clearly seen, as the scores in Table B3 are much lower than in Table A1. Secondly, the RF is no longer clearly the best model, as in some cases, the SVM outperforms the RF. For instance, the minimum MAE and NMAE of the SVM are smaller than those of the RF. For some turbines, however, the SVM struggles to predict, as shown by its higher maximum scores compared to the GP and RF models. Additionally, the RF median performance is always better than the SVM. Thus, the RF model is selected to ensure better overall performance and to avoid individual turbines having a much higher error compared to the farm overall. In summary, to predict farm performance from reference wind conditions, RF models will be used first to predict turbine wind speed, then turbine *TI*, and finally turbine power. For reference, Table 3 lists the hyperparameters of the machine learning models used. The settings optimized by DeepHyper are listed but do not necessarily represent the absolute optimal hyperparameters for each instance, given that DeepHyper does not use a deterministic solver to optimize the settings. Thus, while the given hyperparameters are accurate, there is no guarantee that running DeepHyper again would produce identical results.

4 | MODEL APPLICATION

With the P2D-RANS-based surrogate model and the RF models defined above, the developed modeling approach is now applied to the King Plains wind farm and the accuracy of the models is investigated. Of particular interest is whether the accuracy of the models is sufficient to capture the effects of farm-to-farm interactions.

4.1 | Farm power predictions

To start, the RANS surrogate model and RF models are used to predict total farm power for all timestamps in the reference-condition time series, and the predicted power is compared to the real power. Statistical parameters, such as RMSE, MAE, NMAE, and R^2, are computed and reported in Table 4, with RMSE and MAE being normalized by the rated power of the farm, that is, the rated power of the turbines times the number of turbines, and reported as PRMSE and PMAE. Additionally, the linear regression between true and predicted values for each case is illustrated in Figure 7. It is noticed that there is a large number of points for both the RANS and ML approaches where the model overestimates the farm power, especially when the farm power is zero. Since the reference conditions time series used to produce this analysis were not filtered, these points likely correspond to curtailment conditions that were removed during filtering. Thus, given the environmental conditions, the models

TABLE 3 Hyperparameters of machine learning models used. Note that entries marked by an asterisk were not optimized by DeepHyper and used either setting found to produce good results by a manual sensitivity analysis in the case of the DNN models or the default settings in the case of the RF models.

Model use/type	Layers	Neurons	Depth	Estimators
DNN / Filter	6	280		
RF / Filter			7	275
RF / RANS Surrogate, Manual Grid			92	54
RF / RANS Surrogate, LHS Grid			81	26
RF / RANS Surrogate, Manual + LHS Grid			12	27
RF / RANS Surrogate, First Resample			16	40
RF / RANS Surrogate, Second Resample			17	188
DNN / WS Model*	5	150		
DNN / TI Model*	5	150		
DNN / Power Model	8	175		
RF / WS Model*			100	N/A
RF / TI Model*			100	N/A
RF / Power Model			7	1900

TABLE 4 Error analysis comparing total farm power prediction accuracy using the RANS surrogate model and RF models for unfiltered and filtered data.

Method	PRMSE [%]	PMAE [%]	NMAE [%]	R^2 [–]
		Unfiltered		
P2D-RANS	16.00	7.737	20.97	0.719
ML	16.48	6.730	18.21	0.702
		Filtered		
P2D-RANS	5.319	4.191	11.60	0.955
ML	2.619	1.894	5.27	0.989

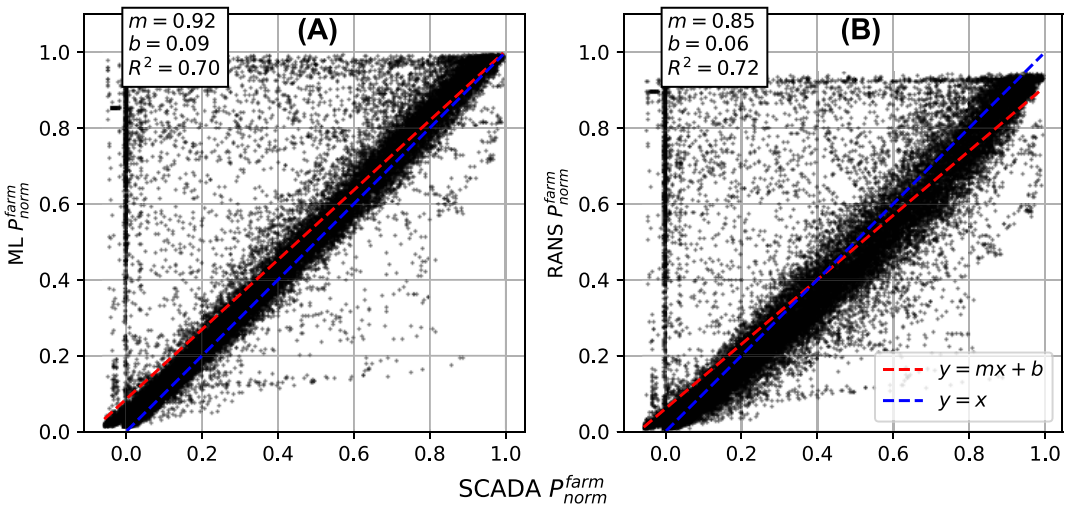

FIGURE 7 Predicting total farm power from unfiltered time series of reference conditions using (A) RF models and (B) RANS surrogate model and comparing with SCADA data.

reproduced ideal turbine behavior rather than derated behavior. To better assess the model accuracy, the time series of reference conditions are restricted to only those points where every turbine is considered to be an inlier according to the GP filter. The GP filter is chosen instead of the RF filter because the RF filter is too strict, reducing the time series to fewer than 2000 points. The GP is a good balance of including points and

removing probable derated points. The RF models are trained on data sets excluding the points to be tested, ensuring that the result is an accurate assessment of the model over new data. Table 4 reports the error metrics after outliers are removed and the regression results are reported in Figure 8.

Predicting using filtered points, that is, predicting points where all the turbines are performing nominally, greatly improves the error metrics of the models. While the RANS surrogate model performs better for the unfiltered set by a relatively small margin, the ML approach performs significantly better over the filtered set. This is indicative of an important feature of ML modeling: the model will generally reproduce the behavior of the training set. Since the RF models were trained on filtered data, they reproduce nominal turbine behavior. The RANS surrogate model, however, could be considered a more general or average performance model, since it lacks tuning of the wake profiles to this specific farm. Thus, considering unfiltered data, the RANS surrogate model performs better reproducing average behavior than the RF models reproducing nominal performance even for underperforming points. This highlights the importance of filtering so that the training set has only points characteristic of the desired behavior to be reproduced by the models.

4.2 | Turbine power predictions

To further investigate the model performance, the models are used to predict power for individual turbines. For the ML approach, each turbine data set filtered using the RF filter is used. Each set is split into a training and testing set such that the training set has 80% of the total points and the testing set has an equivalent distribution of waked and unwaked points to the total set as well as an equivalent distribution of power. The RF models are trained on the training set and then used to predict power for the testing set to ensure the model is being assessed over new data. For the RANS surrogate model, the same splitting approach is used, and the models predict power for only 20% of the available filtered turbine data. While this step is unnecessary for the RANS surrogate model as the test is already a blind test, this ensures that the same number of points are used in the ML and RANS analysis, thus keeping the two comparisons consistent. RMSE, MAE, NMAE, and R^2 scores are calculated for each modeling technique for all the considered data as well as the waked and unwaked sets separately. Again, RMSE and MAE scores are normalized by turbine-rated power and reported as PRMSE and PMAE. These scores are reported in Table 5 while Figure 9 illustrates the results of the regression analysis.

The task of predicting the performance of an individual turbine from reference conditions alone is a far more difficult task than that of predicting the performance of the entire farm. Once again, the RF models outperform the RANS surrogate model by a significant margin. Additionally, considering the RANS surrogate model, it seems that the overall performance is somewhat an average between waked and unwaked performances, with waked conditions being substantially more difficult to predict. For the RF models, however, the change in metrics from all, to waked, to unwaked is much smaller. The PRMSE score changes almost negligibly, less than (1%) from case to case. This suggests that the RF models can better capture the features describing waked behavior from the data than the RANS surrogate model can reproduce these features. In other words, the variability caused by wakes can be included into the RF models, whereas the RANS surrogate model struggles to reproduce this behavior accurately. This also shows that the inputs selected for the RF models contain the necessary features for both waked and unwaked performance.

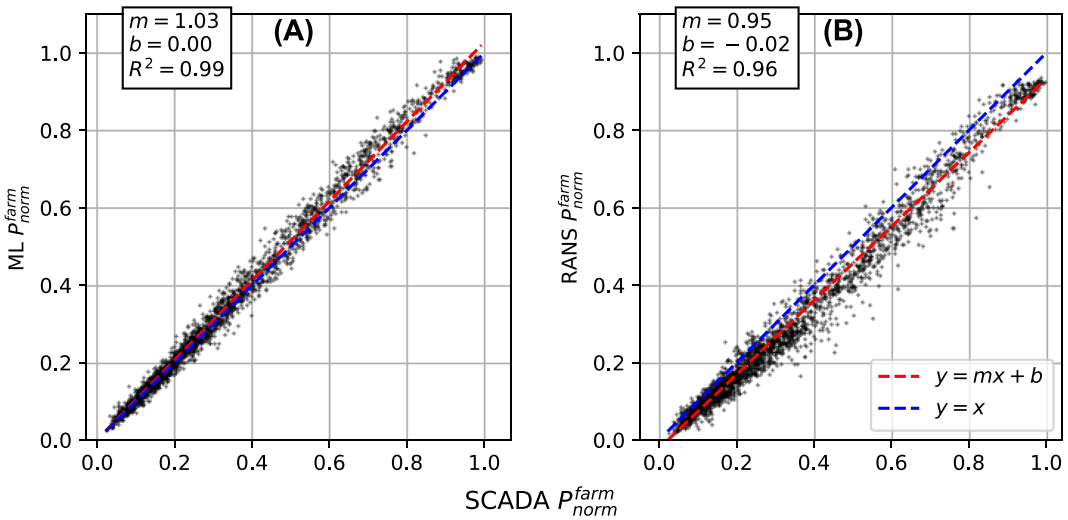

FIGURE 8 Predicting total farm power from filtered time series of reference conditions using (A) RF models and (B) RANS surrogate model and comparing with SCADA data.

TABLE 5 Accuracy scores for RANS and RF models predicting individual turbine power for all conditions, waked conditions, and unwaked conditions.

		PRMSE [%]	PMAE [%]	NMAE [%]	R^2 [–]
RANS	All	13.11	8.855	22.097	0.83228
	Waked	14.04	9.681	25.793	0.80209
	Unwaked	11.52	7.585	17.250	0.87260
RF	All	9.696	6.374	15.907	0.90822
	Waked	9.728	6.509	17.343	0.90501
	Unwaked	9.647	6.166	14.024	0.91072

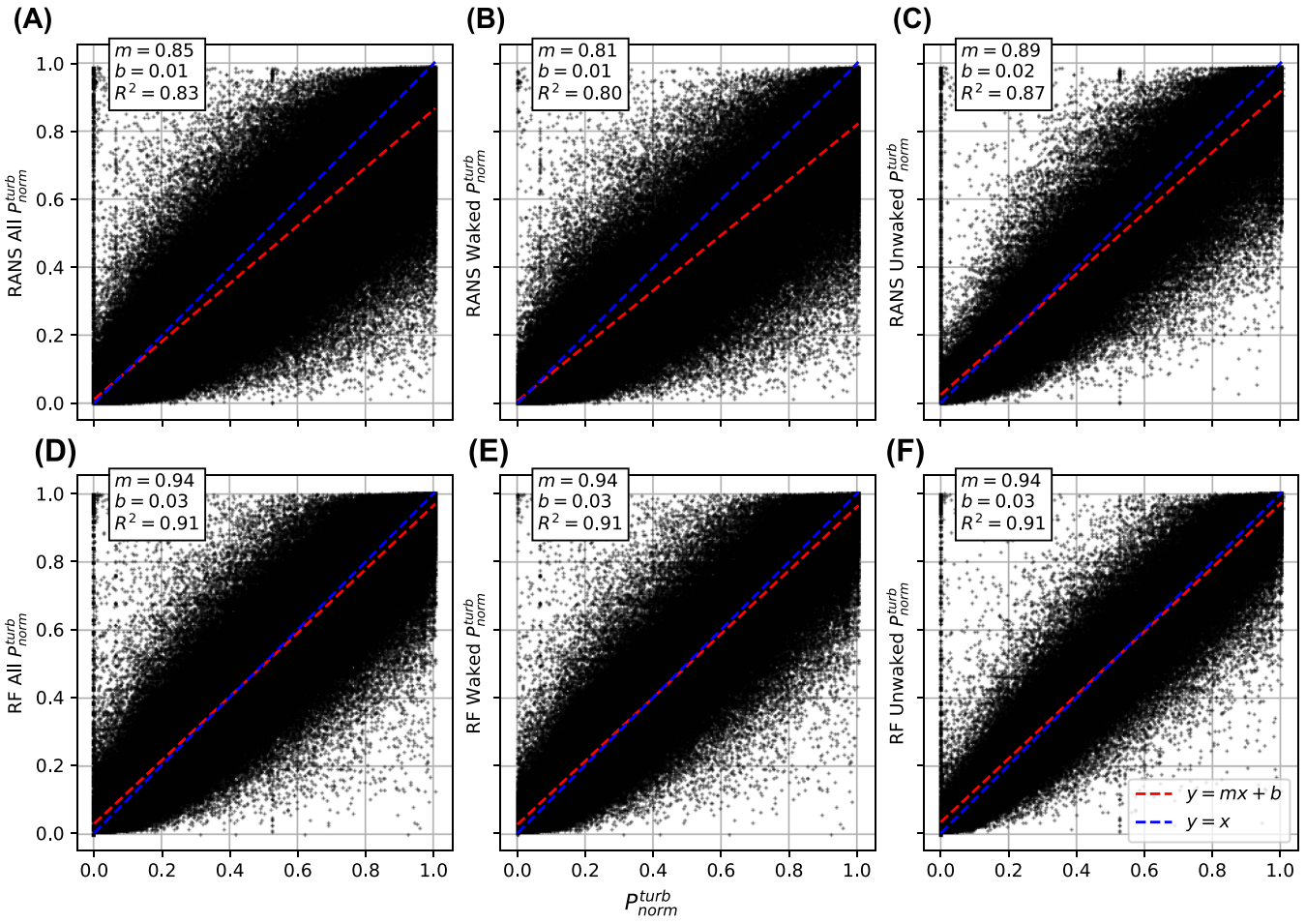

FIGURE 9 Predicting individual turbine power from reference conditions using the RANS surrogate model (A–C) and the RF models (D–F) for all conditions (A, D), waked conditions (B, E), and unwaked conditions (C, F) and comparing with SCADA data.

4.3 | Farm-to-farm interactions

As noted before, this specific site has a high likelihood of farm-to-farm interactions. These interactions could impact the reference conditions. Since the reference conditions are used as inputs to the RANS surrogate models, the assumption is that the RANS freestream inputs map to the reference wind conditions of the farm. Under farm-to-farm interactions, however, this assumption may not be valid due to increased variability in inflow conditions. While the RF models may be able to extract this increased variability as a feature in the data and accurately reproduce it, the RANS models do not have this capability. To investigate this further, the farm-to-farm phenomenon is investigated first in the context of the SCADA data and then in the context of induced error in predictions of the RANS surrogate and RF models.

4.3.1 | Impact of farm-to-farm interactions from SCADA data

First, to highlight intra-farm variability and possible farm-to-farm variability, the following analysis is applied. For each turbine, SCADA data is selected where the reference wind speed is between 7.5 and 8.5 m/s, and the reference *TI* is between 7.5% and 12.5%. Bins are defined in the reference wind direction running from 0° to 360°, each 45° wide. This is to ensure that the turbines are operating in region two and have the right wind speeds and atmospheric stability to experience strong wake interactions. For each turbine, the average wind speed is calculated. Then, for each wind direction bin, the percent difference between the turbine's average wind speed and the median of the averaged wind speeds is calculated for all the turbines. The results are illustrated in Figure 10.

Supposing that the site is free of farm-to-farm or topographic effects on the freestream flow, the first turbines to interact with the freestream flow should have higher than median wind speed, while the turbines downstream should see a decrease in wind speed, eventually experiencing wind speed lower than the median wind speed. Thus, if upstream turbines experience wind speed that is lower than the median wind speed, then the freestream flow has likely experienced induced variability. The subplots that highlight behavior expected of undisturbed freestream conditions are Figure 10B,C,F. In all these cases, the turbines with higher wind speed tend to fall further upstream while turbines experiencing reduced wind speed fall further downstream. The remaining subplots all show some evidence of farm-to-farm interactions.

In Figure 10A,E, turbines on the north row between 0 and 5 km east are all registering median or lower than median wind speed despite being the first row exposed to the freestream. In Figures 10D,H, the front row—now on the south—turbines behave similarly. In both cases, turbines east of the 0 km line tend to have a positive wind speed differential while those west of the line tend to have a negative wind speed differential. All the turbines should ideally be experiencing unperturbed freestream, thus the variability in wind speed can be attributed to freestream variability. Finally, Figure 10G continues the trend of upstream turbine variability, with the turbines north of the 0 km line and west of the −5 km line tending to have positive wind speed differentials while turbines south of the 0 km line tend to have negative differentials. In all cases of likely freestream modification, it is not clear what exactly the source of the modifications is, but given the magnitude of the wind speed differentials, farm-to-farm interactions in the directions noted are certainly a possible explanation.

An objection may be raised to the above analysis that some panels such as Figure 10A,E could be reporting flow speedups on the sides of the wind farm due to flow redirection around the farm rather than farm-to-farm interactions. This would explain why turbines or even the middle of the first row. If this was the case, then the behavior in question would be due purely to effects within the considered farm. To check if this is the case, the P2D-RANS model is used to simulate the farm for wind from the north with a wind speed of 8 m/s and *TI* of 10% to match the bins examined as closely as possible. The variation of the wind field with respect to freestream is simulated. Additionally, for each turbine, the percent difference of the turbine wind speed from the median turbine wind speed across the farm is calculated. The results of this analysis are reported in

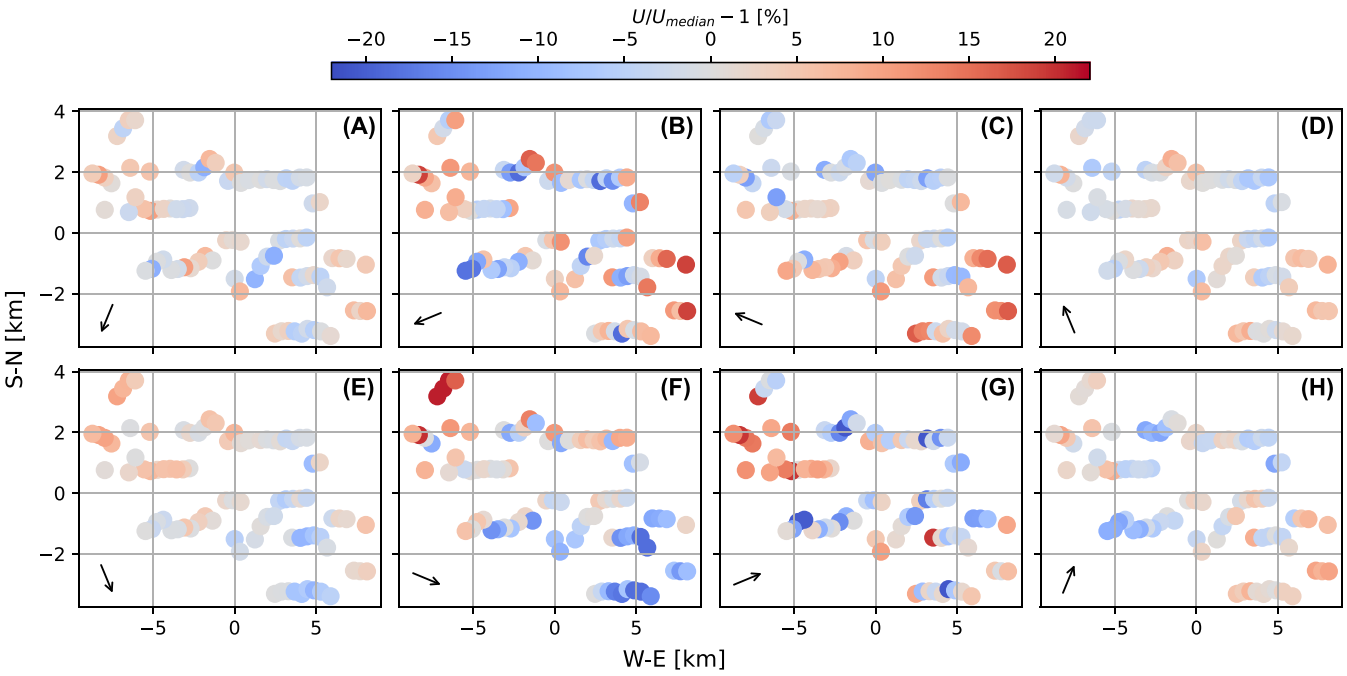

FIGURE 10 Average power for each turbine binning with reference wind speed between 7.5 and 8.5 m/s, reference *TI* between 7.5% and 12.5%, and reference wind direction between the following: (A) 0° and 45°, (B) 45° and 90°, (C) 90° and 135°, (D) 135° and 180°, (H) 180° and 225°, (G) 225° and 270°, (F) 270° and 315°, (E) 315° and 360° (arrow pointing along bin center, wind moving from tail to tip).

Figure 11. Note that, to focus on variability in turbines impacted directly by the freestream flow, percent variations in the turbines smaller than −8% are saturated.

Comparing Figure 10A,E with Figure 11, it is clear that the RANS simulation does not reproduce the variability across the row of freestream turbines that exists in the real data. The freestream turbines have an almost constant wind speed. Additionally, any small variability in the freestream turbines can be easily attributed to speedups due to the wakes of neighboring turbines, which are also seen to cause larger speedups in some select middle-row turbines. Thus, it is concluded that the variations in wind speed in the SCADA data are most likely due to topographic or farm-to-farm effects rather than intra-farm effects.

Considering the variability effects and farm-to-farm effects, the impact of atmospheric conditions is investigated. Since Figure 10D,H seem likely candidates for farm-to-farm interactions, the data is restricted to reference wind directions between 157.5° and 202.5°, thus making a bin 45° wide and centered on 180°. Three *TI* bins are used on reference *TI*, restricting it to the following regions: (2.5%–7.5%), (7.5%–12.5%), and (12.5%–17.5%). Four bins are used in reference wind speed, restricting it to the following regions: 3–5 m/s, 5–7 m/s, 7–9 m/s, and 9–11 m/s. For each bin, the average wind speed is calculated at each turbine, as well as the percent difference from the median average wind speed for all turbines. The result is reported in Figure 12.

From Figure 12, the variability in freestream turbines in the southerly rows is highest in panel (B). The magnitude of the variability appears to decrease with increasing reference wind speed as well as with increasing reference *TI*. For the highest *TI* bin, the variability due to changes in reference wind speed is nearly negligible. For the highest wind speed bin, however, the variability due to changes in the reference *TI* still has a noticeable impact on the freestream turbines' variability. This behavior is strongly typical of wind turbine wakes.[11,12,14]

To summarize, the wind farm under consideration is bordered by neighboring wind farms that pose a strong possibility of farm-to-farm interactions. Moreover, when the reference wind direction aligns the neighboring farms and the King Plains farm, the upstream turbines exhibit strong variability in incoming wind speed. This variability decreases in magnitude with increasing reference wind speed and increasing reference *TI*. Performing RANS simulations under identical inflow conditions, however, does not reproduce the variability. For this reason, the variability is likely due to effects external to the wind farm rather than internal effects creating speedups along the edges of the farm. Though the freestream could be modified by topographic effects, given the relatively flat terrain and the magnitude of the variability, the variability is likely due to freestream modifications caused by farm-to-farm interaction.

4.3.2 | Impact of farm-to-farm interactions on model accuracy

Now that the occurrence of farm-to-farm interactions has been established and investigated, the impact on model predictions is considered. First, the farm power prediction results from Figure 8 are considered. The reference conditions are binned in wind speed bins 1 m/s wide and wind direction bins 10° wide. For each bin, the RANS and RF NMAEs are calculated. Figure 13 plots the results, as well as the RANS NMAE minus the RF NMAE, given as ΔNMAE.

Considering Figure 13A, clear spikes are difficult to discern for the RF model. On the other hand, the RANS surrogate model has comparatively large spikes in error around reference wind directions of 90° and 270°, corresponding to the wind directions with the highest intra-farm wake interactions. Taking the difference in the NMAEs highlights the fact that these spikes are primarily present in the RANS result. This suggests that the RF models are better at capturing the higher variability due to increased wake interactions. Additionally, in the region between reference wind directions of 90° and 270°, the difference is almost always positive and tends to lie between 0% and 10%. This corresponds to the region with the expected strongest farm-to-farm interactions. This could indicate that the RF models do a better job of capturing or correcting both the impact of farm-to-farm wake interactions on the reference conditions as well as on the variability within the farm.

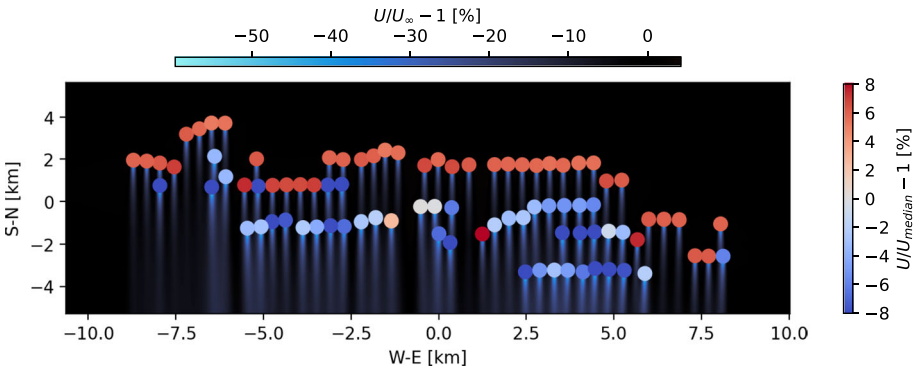

FIGURE 11 RANS simulation of the wind farm for wind from the north at 8 m/s with a *TI* of 10%.

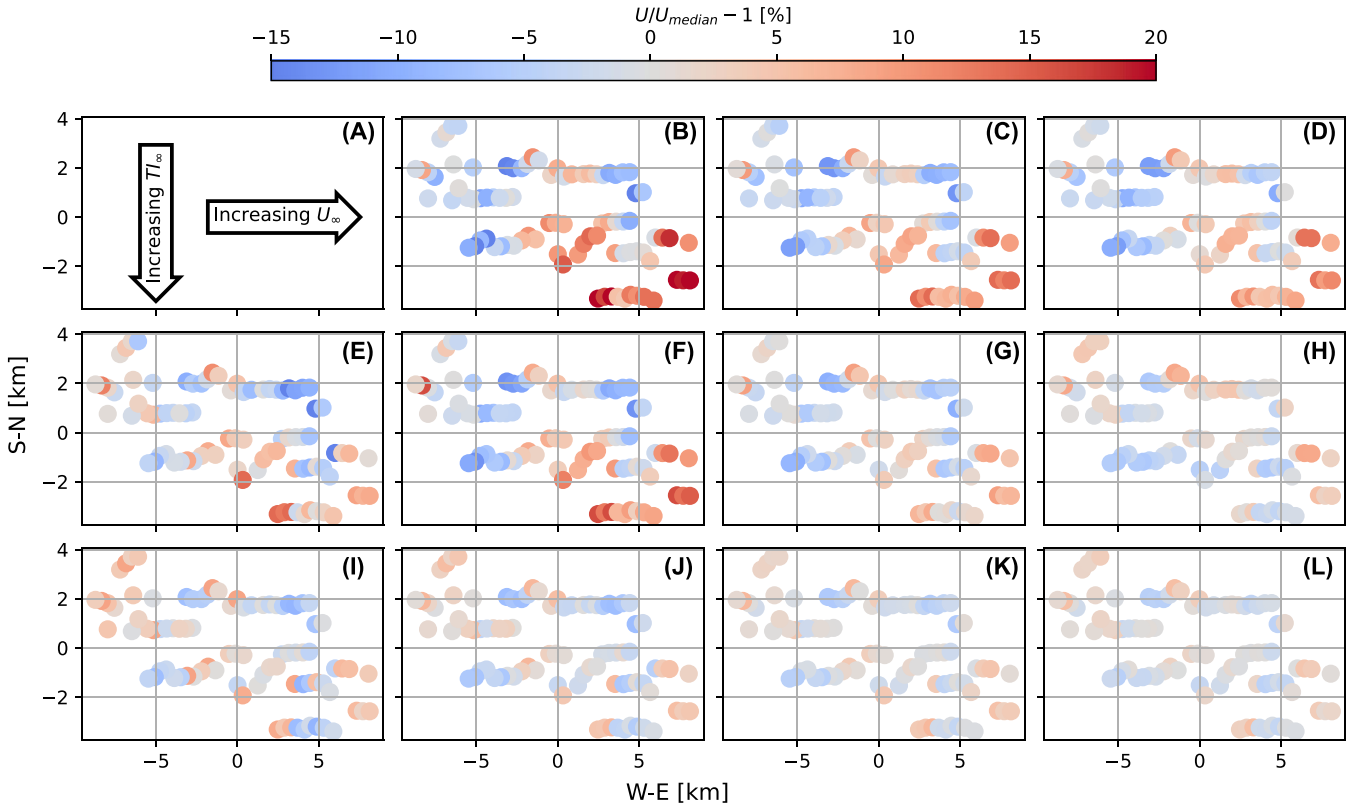

FIGURE 12 Percent deviation from median wind speed for SCADA data between reference wind directions of 157.5°−202.5°. Reference wind speed is binned using the following intervals: [3 m/s, 5 m/s) (A, E, I), [5 m/s, 7 m/s) (B, F, J), [7 m/s, 9 m/s) (C, G, K), and [9 m/s, 11 m/s) (D, H, J). Reference *TI* is binned using the following intervals: [2.5%, 7.5%) (A–D), [7.5%, 12.5%) (E–H), and [12.5%, 17.5%) (I, J). All bins have SCADA data available except panel (A).

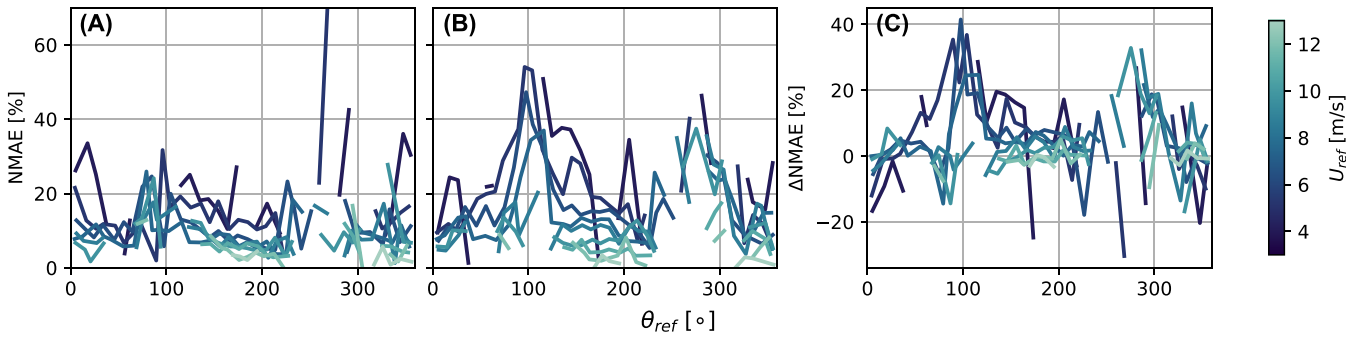

FIGURE 13 Bin NMAE of farm power predictions for RF (A) and RANS (B) models as well as the RANS NMAE minus the RF NMAE for each bin (C).

To investigate how the RANS and RF model errors vary with wind direction at the turbine level, each model predicts local turbine power for the reference conditions time series, as in Figures 9A,D. The data are binned into wind direction bins 45° wide from 0° to 360°. For each bin, for each turbine, the NMAE between true and predicted power is calculated. Across the farm, the percent difference between a given turbine's NMAE and the minimum NMAE of the bin for all turbines is calculated to isolate the variability of error within the farm. The minimum NMAE of each bin is also indicated as the background color of each subplot. Finally, for each bin, the percent error in total power prediction is reported as ΔP, given in Equation (1). These results are all reported in Figure 14.

$$\Delta P = \frac{\Sigma_i^N P_{i,SCADA} - \Sigma_i^N P_{i,pred}}{\Sigma_i^N P_{i,SCADA}}, \quad (1)$$

where N is the number of turbines and P_i is the sum of the power for the ith turbine for the respective bin.

The results in Figure 14 show that the minimum NMAE between the models tends to be either similar or the RF tends to have lower minimum NMAEs. Thus, differences in the model accuracy are due to variations across the farm and not in the minimum NMAE of the models. In general, turbines with higher NMAE compared to the farm minimum are more waked. The notable exception occurs in Figure 14H, where turbines on the west edge of the farm on the south row have a much higher NMAE than those along the east side.

The RF model, however, slightly reproduces this trend but is much more consistent in Figure 14P across the southern row. The RF generally has much more consistent NMAEs across the farm, with wake-associated peaks typically having a much smaller magnitude than their RANS counterparts. Once again, this suggests that the RF model does a better job of reproducing wake-induced variability. This is confirmed with the ΔP metric, which indicates the RANS surrogate model tends to underestimate power compared to the SCADA for each bin between 8% and 27%, suggesting the RANS wake model is predicting slowdowns that are too large. On the other hand, the RF models fluctuate between positive and negative ΔP but with a magnitude always equal to or lower than 2%, demonstrating a much more accurate capture of wake-induced power variations.

It appears that possible farm-to-farm interaction is causing the NMAE variability in the southern row for the RANS surrogate model in Figure 14D without causing the same in the RF case in Figure 14P. This is further investigated by subtracting the RF NMAE for each turbine for each bin from the corresponding RANS NMAE. This analysis highlights the fact that the RANS surrogate model does not recreate waked behavior as accurately as the RF models (see Figure 15). Furthermore, between Figure 15D and 15H, the southern row of the farm switches from nearly constant across the row to a larger difference in NMAEs toward the west edge of the farm. This indicates that the RF is predicting those turbines' power more accurately. Supposing farm-to-farm interactions, this result could indicate that the RF models can still capture farm behavior despite inflow variability not captured by the reference conditions, while the RANS surrogate model cannot.

Finally, to investigate the impact of farm-to-farm interactions on power predictions for a specific case, following an approach previously used to test farm-to-farm model performance,[57] the RANS and RF models are used to predict power for all turbines for a constant wind speed of 8 m/s, *TI* of 10%, and varying wind direction between 140° and 220° in 1° steps. For each wind direction, the RANS and RF models predict

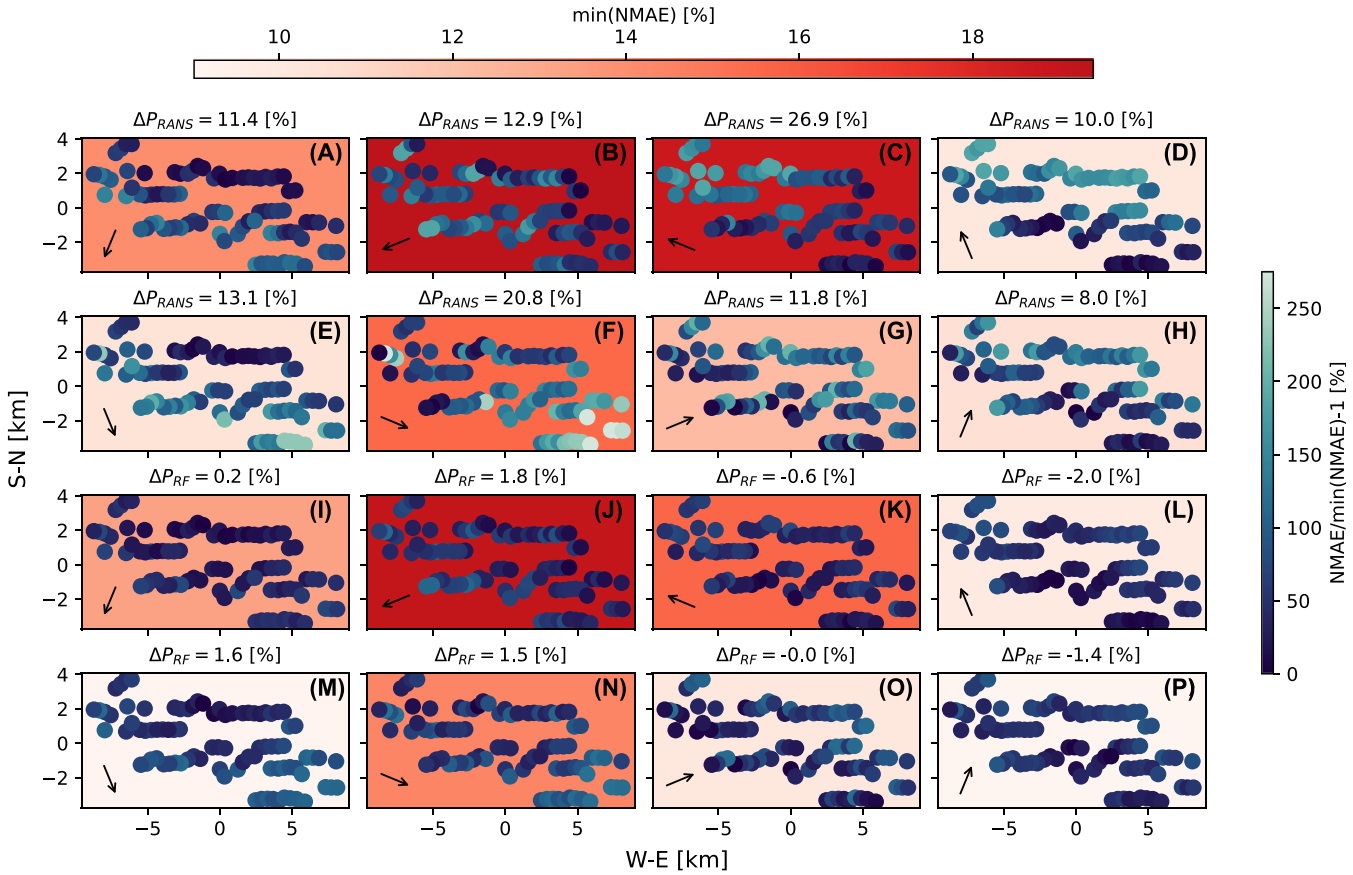

FIGURE 14 Bin NMAE of turbine power predictions for RANS (A–H) and RF (I–P) models, reference wind direction between the following: (A, I) 0° and 45°, (B, J) 45° and 90°, (C, K) 90° and 135°, (D, L) 135° and 180°, (H, M) 180° and 225°, (G, N) 225° and 270°, (F, O) 270° and 315°, (E, P) 315° and 360° (arrow pointing along bin center, wind moving from tail to tip). The subplot color indicates the minimum NMAE for the subplot.

turbine powers. Additionally, the SCADA data is binned in non-overlapping bins in reference wind direction 5° wide, with reference wind speed restricted between 7.5 and 8.5 m/s and reference *TI* between 7.5% and 12.5%. For each turbine, the average power is calculated from the SCADA data for each bin. Conversely, for the two models, average turbine power is calculated for each turbine using the predicted powers for each wind direction but weighting the powers by the frequency of occurrence of the wind directions in the SCADA data for the bin under consideration. Thus, for each bin, average turbine powers are computed from SCADA data, RANS predictions, and RF predictions. Furthermore, two clusters of turbines in the south row are defined, illustrated in Figure 16A.

Given the location of the wind farm to the south, the westerly cluster is expected to experience farm-to-farm interactions for certain wind directions. To verify this, the median power of each cluster is calculated for each wind direction bin, and the ratio of the east and west cluster medians is plotted in Figure 16B. Additionally, the median ratios for the clusters are calculated and plotted for the two models. Finally, the power loss due to farm-to-farm interactions is investigated. Ideal power is defined using the power curve and assuming a freestream wind speed of 8 m/s. The total SCADA power, as well as the total predicted RANS surrogate and RF powers, are compared to the ideal power, summing across all the turbines in both clusters, and the percent difference is plotted in Figure 16C, with positive values indicating a lower power than ideal.

Interpreting Figure 16B, it is clear from the SCADA data that freestream modification is occurring due to farm-to-farm interactions. The impacted turbines are producing as much as 50% less power than the unimpaired turbines. As expected, the RANS surrogate model does not capture any of this variability. Of course, the RANS simulates the farm performance by assuming a clean inflow, so it should not be a surprise that it does not recreate variability. Interestingly, the RF models do recreate a similar trend in variability to the SCADA data but have a much larger magnitude of variability, indicating that in some cases the impacted turbines produce as much as 100% less power. When variability between the clusters is low, such as for wind directions before 160°, the models are both similar in performance to the SCADA data. When the freestream is being modified, the models perform almost equally poorly, though in opposite ways. Care needs to be taken in defining inflow conditions in cases such as this one where strong inflow modifications are occurring.

Considering Figure 16C, power losses of up to (15%) occur in the SCADA data due to farm-to-farm interactions. The RANS, of course, predicts powers much closer to the ideal powers. The RF models, on the other hand, actually predict power increases over the ideal case. Thus, even though the RF models can recreate some variability trends that are more typical of reality than the RANS predictions, there can still be errors in the magnitude of the predicted powers. Since the previous analysis has shown that the RF models outperform the RANS surrogate model generally in turbine and farm power, this analysis highlights how model performance can vary when small regions are considered.

From this analysis, it is clear that both models have room for improvement. Neither model accurately matches the SCADA variability, with one under-predicting and one over-predicting the magnitude. Furthermore, neither model accurately predicts the actual power loss due to farm-to-farm interactions. Thus, though the models are suited to predicting farm performance for the entire farm, as previously shown, even under farm-to-farm interaction cases, they are not currently capable of investigating farm-to-farm interaction effects in a localized manner. Further improvement of the freestream condition definitions may help with this, with future work perhaps even coupling ML and RANS models to more accurately characterize freestream modifications and the resulting power variability or losses.

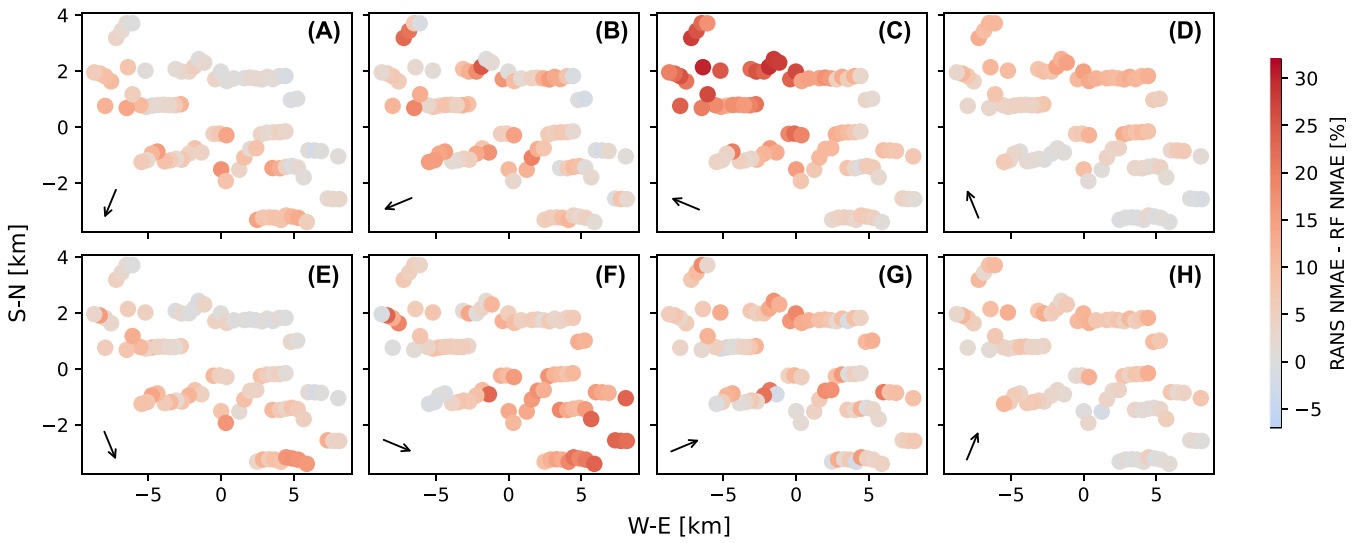

FIGURE 15 RANS bin NMAE for predicted turbine power minus RF bin NMAE, reference wind direction between the following: (A) 0° and 45°, (B) 45° and 90°, (C) 90° and 135°, (D) 135° and 180°, (H) 180° and 225°, (G) 225° and 270°, (F) 270° and 315°, (E) 315° and 360° (arrow pointing along bin center, wind moving from tail to tip).

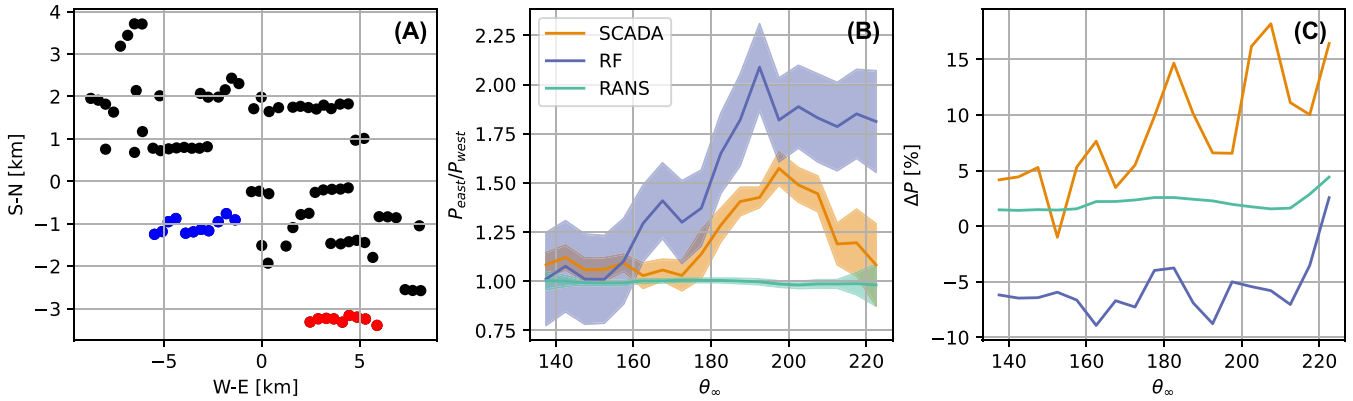

FIGURE 16 (A) Selection of east (red) and west (blue) turbine clusters, (B) the ratio of median cluster power between east and west for different models varying wind direction, and (C) ΔP for the sum of averaged turbine powers for each bin compared with the ideal power. The shaded region is one standard deviation interval.

5 | CONCLUSIONS

In this paper, a comparison of machine learning (ML) models and the pseudo-2D RANS model[46] to simulate operations of the King Plains wind farm, which is a site of the American WAKE experimeNt (AWAKEN), has been presented. This wind farm is composed of (88) turbines installed on relatively flat terrain in Oklahoma. Interactions with neighboring wind farms are also very likely, given large installments of turbines at a short distance from the farm under consideration. SCADA data is available for just over one year from December 2020 to December 2021. Several methods of filtering the SCADA data have been considered, including automatic ML filters, and a filtering method using a Random Forest (RF) model for turbine power. After filtering the data, ML models have been used to predict local turbine wind speed and turbulence intensity (*TI*) for specific inflow conditions, turbine power is then predicted from the local wind speed and *TI*. Performing a comparative analysis across the farm, the RF algorithm has been selected for each model used. When possible, models are optimized using the DeepHyper package. Chaining the RF models allows individual turbine power to be predicted from approximated inflow conditions.

To generate predictions with the RANS model, an initial subset of inflow conditions in wind speed, wind direction, and *TI* has been sampled by solving each combination of inflow conditions with the RANS model. An RF model is then trained on top of the RANS results to create a surrogate model for farm power capture. Another RF model has been trained to predict the error of the power surrogate model given specific inflow conditions. With this model, the input domain has been intelligently resampled to reduce error in the surrogate model. This resampling procedure has then been applied until the power surrogate model no longer benefits from additional points simulated with the P2D-RANS model. Individual surrogate models have been trained for each turbine. Thus, from user-supplied inflow conditions, the surrogate models can predict turbine power performance without requiring additional RANS simulations.

Once the ability to simulate farm performance for generic inflow conditions has been established for both RANS and ML approaches, a series of analyses have been performed to evaluate model performance. First, both models have been used to predict the total farm power, i.e. the sum of power across all turbines, for a time series of inflow conditions. Removing points with derated turbines and comparing the predictions against SCADA data, the RANS surrogate model has entailed a normalized mean absolute error (NMAE) of 11.6% while the RF models have an NMAE of 5.3%. Subsequently, the models have been used to predict individual turbine power for all inflow conditions. Keeping only SCADA points with nominal performance and calculating scores across all turbines, the RANS surrogate model and RF models have produced scores of 22.1% and 15.9% NMAE, respectively.

Investigating the model error further, the NMAE for the RANS surrogate model and RF models have been compared for the filtered time series on total farm power, considering model error a function of inflow wind speed and wind direction. RF error is found to be relatively constant across wind directions while the RANS error increases when the wind direction is perpendicular to the prevailing wind direction for the farm and intra-farm wake interactions are maximized. Furthermore, considering the NMAE errors for both models at the turbine level as a function of wind direction, it is seen that turbines under waked conditions typically have higher errors than turbines that are unwaked. While this trend is repeated in both models, the difference in errors is much greater for the RANS surrogate model than the RF models.

Finally, to consider the potential effects of farm-to-farm interactions, the difference in NMAEs for each model at each turbine is considered as a function of wind direction. The RANS surrogate model has had a worse performance for all wake conditions but can perform worse at specific turbines that may be experiencing farm-to-farm interactions. Finally, for a specific inflow wind speed and *TI*, the ratio of median power between clusters of turbines for SCADA, RANS, and RF powers has been considered to highlight localized farm-to-farm interactions. The RANS surrogate model has no capability to reproduce the flow variability while the RF model reproduces twice as much variability as the SCADA data.

Additionally, neither RANS nor RF accurately predicts the power loss due to farm-to-farm interactions. This suggests that though the RF model is adept at predicting performance across the farm, both the RF models and the RANS surrogate model break down when trying to predict farm-to-farm interactions with accurate estimates.

From this analysis, it has been concluded that the RF models are better to recreate turbine performance under waked conditions, especially wakes within the farm. Farm-to-farm wakes, on the other hand, are difficult to predict and quantify related effects on power capture. Thus, not only does the RANS wake model have room for improvement, but both models can improve performance under variable inflow conditions. Future work could consider modifications to the freestream inputs or even coupling of RANS and ML to advance the model capabilities for complex and variable inflow conditions.

ACKNOWLEDGEMENTS

This article was developed based upon funding from the Alliance for Sustainable Energy, LLC, Managing and Operating Contractor for the National Renewable Energy Laboratory for the U.S. Department of Energy. CFM and GVI are also supported by the National Science Foundation (NSF) CBET, Fluid Dynamics CAREER program, Award No. 2046160, program manager Ron Joslin.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/we.2874.

DATA AVAILABILITY STATEMENT

The data used in this manuscript are covered by a non-disclosure agreement.

ORCID

Coleman Moss https://orcid.org/0000-0002-4188-6915
Giacomo Valerio Iungo https://orcid.org/0000-0002-0990-8133

REFERENCES

1. Kaldellis JK, Zafirakis D. The wind energy (r)evolution: a short review of a long history. *Renew Energy*. 2011;36(7):1887-1901.
2. Gielen D, Boshell F, Saygin D, Bazilian MD, Wagner N, Gorini R. The role of renewable energy in the global energy transformation. *Energy Strategy Rev*. 2019;24:38-50.
3. Mishnaevsky L, Branner K, Petersen HN, Beauson J, McGugan M, Sørensen BF. Materials for wind turbine blades: an overview. *Materials*. 2017;10(11):1285.
4. Schubel PJ, Crossley RJ. Wind turbine blade design. *Energies*. 2012;5(9):3425-3449.
5. Novaes Menezes EJ, Araújo AM, Bouchonneau da Silva NS. A review on wind turbine control and its associated methods. *J Cleaner Product*. 2018;174:945-953.
6. Liu WY, Tang BP, Han JG, Lu XN, Hu NN, He ZZ. The structure healthy condition monitoring and fault diagnosis methods in wind turbines: a review. *Renew Sustain Energy Rev*. 2015;44:466-472.
7. Santhanagopalan V, Letizia S, Zhan L, Al-Hamidi LY, Iungo GV. Profitability optimization of a wind power plant performed through different optimization algorithms and a data-driven RANS solver. In: American Institute of Aeronautics and Astronautics 2018 Wind Energy Symposium; 2018.
8. Shakoor R, Hassan MY, Raheem A, Wu Y-K. Wake effect modeling: a review of wind farm layout optimization using Jensen's model. *Renew Sustain Energy Rev*. 2016;58:1048-1059.
9. Stevens RJAM, Meneveau C. Flow structure and turbulence in wind farms. *Ann Rev Fluid Mech*. 2017;49(1):311-339.
10. Porté-Agel F, Bastankhah M, Shamsoddin S. Wind-turbine and wind-farm flows: a review. *Bound-Layer Meteorol*. 2020;174(1):1-59.
11. El-Asha S, Zhan L, Iungo GV. Quantification of power losses due to wind turbine wake interactions through SCADA, meteorological and wind LiDAR data. *Wind Energy*. 2017;20(11):1823-1839.
12. Iungo GV, Porté-Agel F. Volumetric LiDAR scanning of wind turbine wakes under convective and neutral atmospheric stability regimes. *J Atmosph Ocean Technol*. 2014;31(10):2035-2048.
13. Iungo GV. Experimental characterization of wind turbine wakes: wind tunnel tests and wind LiDAR measurements. *J Wind Eng Ind Aerodyn*. 2016;149:35-39.
14. Zhan L, Letizia S, Valerio Iungo G. LiDAR measurements for an onshore wind farm: wake variability for different incoming wind speeds and atmospheric stability regimes. *Wind Energy*. 2020;23(3):501-527.
15. Clifton A, Wagner R. Accounting for the effect of turbulence on wind turbine power curves. *J Phys: Conf Ser*. 2014;524:012109.
16. Bardal LM, Sætran LR. Influence of turbulence intensity on wind turbine power curves. *Energy Procedia*. 2017;137:553-558.
17. Churchfield M, Lee S, Moriarty P, Martinez L, Leonardi S, Vijayakumar G, Brasseur J. A large-Eddy simulation of wind-plant aerodynamics. 50th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition; 2012.
18. González JS, Gonzalez Rodriguez AG, Mora JC, Santos JR, Payan MB. Optimization of wind farm turbines layout using an evolutive algorithm. *Renew Energy*. 2010;35(8):1671-1681.
19. DuPont B, Cagan J, Moriarty P. An advanced modeling system for optimization of wind farm layout and wind turbine sizing using a multi-level extended pattern search algorithm. *Energy*. 2016;106:802-814.
20. Nilsson K, Ivanell S, Hansen KS, Mikkelsen R, Sørensen JN, Breton S-P, Henningson D. Large-eddy simulations of the Lillgrund wind farm. *Wind Energy*. 2015;18(3):449-467.

21. Creech A, Früh W-G, Maguire AE. Simulations of an offshore wind farm using large-Eddy simulation and a torque-controlled actuator disc model. *Surv Geophys*. 2015;36(3):427-481.
22. van der Laan MP, Baungaard M, Kelly M. Inflow modeling for wind farm flows in RANS. *J Phys: Conf Ser*. 2021;1934(1):012012.
23. Barthelmie RJ, Hansen K, Frandsen ST, et al. Modelling and measuring flow and wind turbine wakes in large wind farms offshore. *Wind Energy*. 2009;12(5):431-444.
24. Tabib M, Rasheed A, Kvamsdal T. LES and RANS simulation of onshore Bessaker wind farm: Analysing terrain and wake effects on wind farm performance. *J Phys: Conf Ser*. 2015;625:012032.
25. Iungo GV, Viola F, Ciri U, Rotea MA, Leonardi S. Data-driven RANS for simulations of large wind farms. *J Phys: Conf Ser*. 2015;625:012025.
26. Eivazi H, Tahani M, Schlatter P, Vinuesa R. Physics-informed neural networks for solving Reynolds-averaged Navier–Stokes equations. *Phys. Fluids*. 2022;34(7):075117.
27. Vinuesa R, Brunton SL. Enhancing computational fluid dynamics with machine learning. *Nat Comput Sci*. 2022;2(6):358-366.
28. Bodini N, Lundquist JK, Livingston H, Moriarty P. How generalizable is a machine-learning approach for modeling hub-height turbulence intensity? *J Phys: Conf Ser*. 2022;2265(2):022028.
29. Huang Y, Li J, Hou W, Zhang B, Zhang Y, Li Y, Sun L. Improved clustering and deep learning based short-term wind energy forecasting in large-scale wind farms. *J Renew Sustain Energy*. 2020;12(6):066101.
30. Ti Z, Deng XW, Zhang M. Artificial neural networks based wake model for power prediction of wind farm. *Renew Energy*. 2021;172:618-631.
31. Ashwin Renganathan S, Maulik R, Letizia S, Iungo GV. Data-driven wind turbine wake modeling via probabilistic machine learning. *Neural Comput Applic*. 2022;34(8):6171-6186.
32. Baungaard M, Abkar M, van der Laan MP, Kelly M. A numerical investigation of a wind turbine wake in non-neutral atmospheric conditions. *J Phys: Conf Ser*. 2022;2265(2):022015.
33. Zehtabiyan-Rezaie N, Iosifidis A, Abkar M. Data-driven fluid mechanics of wind farms: a review. *J Renew Sustain Energy*. 2022;14(3):032703.
34. Marvuglia A, Messineo A. Monitoring of wind farms' power curves using machine learning techniques. *Appl Energy*. 2012;98:574-583.
35. Pei S, Li Y. Wind turbine power curve modeling with a hybrid machine learning technique. *Appl Sci*. 2019;9(22):4930.
36. Pandit RK, Infield D, Carroll J. Incorporating air density into a Gaussian process wind turbine power curve model for improving fitting accuracy. *Wind Energy*. 2019;22(2):302-315.
37. Zhou J, Guo P, Wang X-R. Modeling of wind turbine power curve based on Gaussian process. In: 2014 International Conference on Machine Learning and Cybernetics, Vol. 1; 2014:71-76.
38. Purohit S, Ng EYK, Syed Ahmed Kabir IF. Evaluation of three potential machine learning algorithms for predicting the velocity and turbulence intensity of a wind turbine wake. *Renew Energy*. 2022;184:405-420.
39. Iungo GV, Maulik R, Renganathan SA, Letizia S. Machine-learning identification of the variability of mean velocity and turbulence intensity for wakes generated by onshore wind turbines: Cluster analysis of wind LiDAR measurements. *J Renew Sustain Energy*. 2022;14(2):023307.
40. Sun H, Qiu C, Lu L, Gao X, Chen J, Yang H. Wind turbine power modelling and optimization using artificial neural network with wind field experimental data. *Appl Energy*. 2020;280:115880.
41. Optis M, Perr-Sauer J. The importance of atmospheric turbulence and stability in machine-learning models of wind farm power production. *Renew Sustain Energy Rev*. 2019;112:27-41.
42. Howland MF, Dabiri JO. Wind farm modeling with interpretable physics-informed machine learning. *Energies*. 2019;12(14):2716.
43. Yan C, Pan Y, Archer CL. A general method to estimate wind farm power using artificial neural networks. *Wind Energy*. 2019;22(11):1421-1432.
44. Park J, Park J. Physics-induced graph neural network: an application to wind-farm power estimation. *Energy*. 2019;187:115883.
45. Moriarty P, Hamilton N, Debnath M, et al. American WAKE experimeNt (AWAKEN). *Tech. Rep*.; 2020. NREL/TP-5000-75789, 1659798, MainId: 5894.
46. Letizia S, Iungo GV. Pseudo-2D RANS: a LiDAR-driven mid-fidelity model for simulations of wind farm flows. *J Renew Sustain Energy*. 2022;14(2):023301.
47. Wind energy generation systems - part 12-1: power performance measurements of electricity producing wind turbines, international standard 61400-12-2. *Int Electrotechn Commiss*. 2013;3:2017-3.
48. Krishnamurthy R, Newsom RK, Chand D, Shaw WJ. Boundary layer climatology at ARM Southern Great Plains. *Tech. Rep*. PNNL-30832, Pacific Northwest National Lab. (PNNL), Richland, WA (United States); 2021.
49. Burton T, Jenkins N, Sharpe D, Bossanyi E. *Wind energy handbook*. First: Wiley; 2011.
50. Manobel B, Sehnke F, Lazzús JA, Salfate I, Felder M, Montecinos S. Wind turbine power curve modeling based on Gaussian processes and artificial neural networks. *Renew Energy*. 2018;125:1015-1020.
51. Yesilbudak M. Implementation of novel hybrid approaches for power curve modeling of wind turbines. *Energy Convers Manag*. 2018;171:156-169.
52. Virtanen P, Gommers R, Oliphant T, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*. 2020;17:261-272.
53. Garland NA, Maulik R, Tang Q, Tang X-Z, Balaprakash P. Efficient training of artificial neural network surrogates for a collisional-radiative model through adaptive parameter space sampling. arXiv arXiv:2112.05325; 2021.
54. Maulik R, Egele R, Lusch B, Balaprakash P. Recurrent neural network architecture search for geophysical emulation. In: SC20: International Conference for High Performance Computing, Networking, Storage and Analysis; 2020:1-14.
55. Abadi M, Agarwal A, Barham P, et al. TensorFlow: large-scale machine learning on heterogeneous systems. https://www.tensorflow.org/, Software available from tensorflow.org; 2015.
56. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830.
57. Nygaard NG, Steen ST, Poulsen L, Pedersen JG. Modelling cluster wakes and wind farm blockage. *J Phys: Conf Ser*. 2020;1618(6):062072.
58. Yang Y. Consistency of cross validation for comparing regression procedures. *Ann Stat*. 2007;35(6):2450-2473.

How to cite this article: Moss C, Maulik R, Moriarty P, Iungo GV. Predicting wind farm operations with machine learning and the P2D-RANS model: A case study for an AWAKEN site. *Wind Energy*. 2024;27(11):1245-1267. doi:10.1002/we.2874

APPENDIX A: MODEL SELECTION AND TUNING FOR ML FILTERING

This section considers the selection and tuning of the ideal ML model for use in the ML filter. Since the application of the models is power prediction, the goal of filtering is to isolate turbine performance from off-design performance. Thus, the inputs to the models used in filtering should be selected to focus on turbine performance. From the available SCADA data, wind speed and *TI* are selected. Yaw angle or wind direction should not be included because wake effects are captured by decreases in wind speed and increases in *TI*. Adding yaw angle or wind direction begins to add farm layout information to the model by correlating power changes to wind direction and implying the existence of upstream turbines. The model would thus be capturing both turbine performance and layout-induced variations. This is not desired for filtering, thus yaw angle and wind direction should not be used as inputs. Additionally, the misalignment between the turbine rotor and the incoming wind, termed yaw misalignment, is not considered an input. Though this parameter has a marked impact on power production, not all SCADA data available contain this parameter. Thus, the method is made more general by not considering it as input. In this case, the effect of excluding yaw misalignment as a filtering input is small since the turbines are, on average, very aligned, with 75% having a yaw misalignment of 2° or less. Finally, the filter should reject points with large power variability caused by high yaw misalignment.

With the inputs selected, the next step in the ML filter is to select the ML model used. A Random Forest (RF) regression model is used along with a Deep Neural Network (DNN). Other models considered included a Support Vector Machine (SVM) and Gaussian Process (GP), but these models were not comparable to the RF and DNN in terms of accuracy and are not reported here. The DNN models are built using Keras with the TensorFlow backend[55] while the GP, SVM, and RF models all use Scikit-Learn,[56] all in Python. These models are sensitive to hyperparameter tuning and need to have their individual parameters optimized. To avoid optimizing the models on each turbine, which would be computationally expensive, the data from all the turbines are aggregated to form a single data set. Combining all the turbines in this manner creates a data set with over 4,000,000 points, which is too large to efficiently train models for optimization. The data set is therefore downsampled to 20,000 for optimization, keeping an identical power distribution to the initial data set. The RF and DNN models are then optimized to predict power from wind speed and *TI* on this data set using the DeepHyper Python package.[54] When optimizing the RF model, the maximum depth of the RF model and the number of estimators used, the two most important hyperparameters according to the Scikit-Learn documentation, are optimized.[56] On the other hand, the DNN model is to be a fully-connected sequential dense network. The selected parameters to optimize are the number of neurons, the number of hidden layers, and the activation function used. In each case, the R^2 of each model was maximized by the DeepHyper optimizer, which was allowed to run for 200 iterations. The top three resultant RF models all had a maximum depth of 7. The models had a different number of estimators, however, at 191, 565, and 73, with only marginal differences in the R^2. Taking the average of the number of estimators, the final RF model has 7 for the maximum depth and 275 estimators. Averaging across the top five results for the DNN search, the optimized DNN model had 280 neurons, 6 layers, and the exponential linear unit (ELU) activation function.

Now that the RF and DNN models have been optimized to predict turbine power from raw SCADA wind speed and *TI*, they are to be compared across all the turbines to select the most accurate model for use in the ML filter. Five-fold cross-validation is used to test each model on each turbine.[58] The root mean squared error (RMSE), mean absolute error (MAE), normalized mean absolute error (NMAE),[46] and R^2 scores are averaged across each of the five folds and reported for each turbine. The minimum, 25th percentile, median, 75th percentile, and maximum of each of these metrics across the 88 turbines in the farm are reported in Table A1. NMAE is calculated according to Equation (A1). When reporting the errors, the RMSE and MAE are normalized by the turbine-rated power and are reported as percent RMSE and MAE, or PRMSE and PMAE.

$$NMAE = \frac{\Sigma_i^N |True_i - Pred_i|}{\Sigma_i^N |True_i|} \tag{A1}$$

From Table A1, it is clear that the RF is the better model, as it outperforms the DNN in every metric. Thus, the ML filter will use the optimized RF. The filter is set to require points to have an error smaller than three standard deviations from the mean error. The filter retrains on and re-filters the data set until either 50 iterations are reached or the number of rejected points is smaller than 25 (values chosen via sensitivity analysis).

TABLE A1 Statistics of cross-fold validation applied to all turbines for filtering models.

Metric	Model	Min	25th percentile	Median	75th percentile	Max
PRMSE [%]	RF	10.37	12.41	12.76	13.15	15.19
	DNN	10.98	13.17	13.67	15.05	59.31
PMAE [%]	RF	4.66	5.99	6.28	6.73	9.08
	DNN	5.00	6.86	7.74	9.34	48.4
NMAE [%]	RF	9.846	13.42	14.12	15.37	19.50
	DNN	10.34	15.61	17.27	20.65	100.0
R^2 [-]	RF	0.7956	0.8437	0.8535	0.8611	0.9087
	DNN	-1.991	0.7982	0.8321	0.8437	0.8974

APPENDIX B: MODEL SELECTION FOR ML MODELING

Optimal ML models are to be selected for predicting turbine wind speed, *TI*, and power considering the GP, SVM, DNN, and RF models as candidates. For each output, PRMSE, PMAE, NMAE, and R^2 are calculated for each turbine. Across the farm, the minimum, 25th percentile, median, 75th percentile, and maximum of each of these metrics is reported in Tables B1,B2,B3 for wind speed, *TI*, and power predictions respectively. RMSE and MAE are calculated as typically and then normalized by the rated wind speed and thus are reported as PRMSE and PMAE.

TABLE B1 Statistics of cross-fold validation applied to all turbines predicting turbine wind speed from reference wind speed, direction, and *TI*.

Metric	Model	Min	25th percentile	Median	75th percentile	Max
PRMSE [%]	GP	6.455	8.182	8.636	9.000	13.55
	SVM	6.273	7.909	8.182	8.545	13.64
	DNN	7.091	8.545	8.909	9.455	34.82
	RF	5.636	6.818	7.182	7.545	11.91
PMAE [%]	GP	4.82	6.00	6.36	6.55	9.27
	SVM	4.55	5.64	5.91	6.27	8.73
	DNN	5.36	6.27	6.64	7.09	31.7
	RF	4.18	5.00	5.27	5.55	7.91
NMAE [%]	GP	6.22	7.80	8.18	8.61	10.96
	SVM	5.85	7.24	7.72	8.06	10.28
	DNN	6.96	8.11	8.63	9.11	41.92
	RF	5.29	6.47	6.93	7.15	9.28
R^2 [-]	GP	0.74	0.88	0.89	0.90	0.93
	SVM	0.73	0.89	0.90	0.91	0.94
	DNN	−2.34	0.87	0.88	0.89	0.92
	RF	0.80	0.91	0.92	0.93	0.95

TABLE B2 Statistics of cross-validation applied to all turbines predicting turbine *TI* from reference wind speed, direction, and *TI*, as well as turbine wind speed.

Metric	Model	Min	25th percentile	Median	75th percentile	Max
RMSE [%]	GP	2.29	2.78	3.19	3.91	89.06
	SVM	2.22	2.79	3.31	3.90	106.09
	DNN	2.60	5.50	7.36	10.14	93.48
	RF	2.10	2.52	2.89	3.36	86.65
MAE [%]	GP	1.52	1.86	1.98	2.17	6.37
	SVM	1.44	1.76	1.86	2.04	24.38
	DNN	1.80	4.31	5.93	8.00	10.47
	RF	1.37	1.64	1.73	1.89	3.92
NMAE [%]	GP	11.63	14.89	15.55	16.89	40.14
	SVM	10.96	14.01	14.73	15.72	160.96
	DNN	14.35	32.95	47.87	62.35	74.63
	RF	10.46	13.01	13.58	14.55	24.02
R^2 [-]	GP	−0.15	0.59	0.69	0.74	0.81
	SVM	−10.38	0.59	0.67	0.73	0.83
	DNN	−3.76	−1.99	−0.95	−0.16	0.76
	RF	0.14	0.68	0.74	0.78	0.85

TABLE B3 Statistics of cross-validation applied to all turbines predicting turbine power from reference wind speed, direction, and *TI*.

Metric	Model	Min	25th percentile	Median	75th percentile	Max
PRMSE [%]	GP	2.026	2.483	3.075	3.848	11.25
	SVM	1.9806	2.463	2.843	3.723	26.76
	DNN	2.723	3.317	9.635	14.00	33.93
	RF	1.893	2.202	2.391	2.778	5.500
PMAE [%]	GP	1.471	1.711	1.931	2.450	5.411
	SVM	1.420	1.644	1.834	2.279	10.54
	DNN	1.862	2.410	8.070	10.84	26.35
	RF	1.429	1.656	1.798	2.097	4.220
NMAE [%]	GP	3.433	3.923	4.419	5.584	13.28
	SVM	3.177	3.764	4.168	5.182	25.88
	DNN	4.392	5.610	17.789	24.628	62.81
	RF	3.192	3.776	4.151	4.803	9.632
R^2 [-]	GP	0.8994	0.9868	0.9918	0.9947	0.9965
	SVM	0.4349	0.9881	0.9930	0.9947	0.9966
	DNN	−0.5396	0.4555	0.7911	0.9901	0.9933
	RF	0.9735	0.9933	0.9951	0.9959	0.9970