Efficient Loads Surrogates for Waked Turbines in an Array

Kelsey Shaler, John Jasa, and Garrett E. Barter
National Renewable Energy Laboratory, Golden, CO 80401, USA
E-mail: kelsey.shaler@nrel.gov

Abstract. Accurately and efficiently predicting wind turbine structural loading is a crucial step in wind farm design. Without considering structural loading, wind farm optimization could negatively impact turbine fatigue and ultimate loads, especially for waked and partially waked turbines, which could result in higher maintenance costs and reduced turbine lifetime. However, predicting turbine loads throughout an array is a costly step, as these quantities require time-accurate results across long time histories, which is often intractable for large array optimization. Therefore, surrogate models that link array spacing to load outputs are often used, but the surrogates are then unique to the inflow conditions and array configurations in the training library. This work develops surrogate models for many wind turbine load outputs based solely on rotor plane velocity measurements, with no required input about array configuration or freestream inflow parameters. Surrogate models were constructed for many turbine quantities of interest (QoI), considering mean, standard deviation, ultimate, and fatigue loads. In general, most QoI statistics were accurately captured, as measured by predicted vs. actual correlation coefficient, confirming the suitability of the approach. Temporal mean values of the QoI required only temporal mean measurements of the rotor plane inflow velocity. However, accurate prediction of temporal standard deviation, ultimate, and fatigue values of QoI also required temporal standard deviations of the rotor plane velocity field. Poor surrogate performance was observed when too many correlated inputs were used, such as multiple velocity components. If the fewest inflow parameters are used to construct the surrogates, the average correlation coefficient value for all output QoI statistics is 0.89. Surrogates for standard deviations and damage equivalent loads (DELs) of turbine QoIs generally had lower accuracy and tower-base and shaft load channels posed the most difficult to capture accurately. The results suggest that these surrogates could be easily paired with analytic wake models, which are frequently used for pre-construction wind farm array optimization, to account for turbine loading in addition to power production. By including the optimal inflow conditions, the surrogate accuracy can improve to an average correlation coefficient value for all output QoI statistics of 0.92. This work has established the ability to build accurate surrogates for mean, standard deviation, ultimate load, and DEL turbine QoI values based on the rotor plane inflow velocity, and identified which inflow conditions lead to greater surrogate accuracy.

1 Introduction

Designing turbine array layouts is commonly done with lower-fidelity models that can rapidly estimate the flow field throughout the entire wind farm. While this approach has made great strides in maximizing power production potential of new farms, waking and partial waking of wind turbines can lead to increased turbine fatigue loads that are unevenly distributed through the array. Fully or partially waked inflow scenarios are also not covered in turbine design standards. Additionally,

accounting for wake-induced loading on the wind farm scale is computationally difficult with traditional aero-servo-hydro-elastic tools and prohibitive in a layout optimization setting. This is due to the large flow field, increased numbers of turbines, and longer time histories required to determine farm-level effects, especially for fatigue loading.

To better enable the inclusion of turbine loading into wind farm optimization processes, a surrogate model that links array spacing to load outputs is often used, with multiple examples found in the literature. Réthoré et al. used a lookup table of HAWC2 simulations within layout optimizations in TOPFARM. The lookup table assumes that a turbine's inflow profile is determined by its nearest upstream neighbor, so the lookup table varies downwind and crosswind distances of a two-turbine array [17]. This approach was later updated by Riva et al. [18] who combined a single turbine load surrogate model based on detailed inflow quantities [7] with analytical wake models to provide those values for all turbines in an array. Clark et al. [5] also relied on interpolations of a two-turbine simulation library and generalized to a larger array using FAST.Farm to provide the underlying aeroelastic load simulations. Dimitrov compared two surrogate approaches, polynomial chaos expansions and neural networks of aeroelastic simulations, for wake-induced loads of a parameterized, arbitrary layout and later demonstrated it against measured wind farm data [6, 9]. Stanley et al. [20] optimized array layouts that accommodated a maximum blade fatigue damage constraint that was calculated using corrections from a steady-state model to a time-domain model using the inflow wind speed, rotor rpm value, blade pitch angle, and wake turbulence intensity (TI).

These previous approaches are sufficient for their applications, but are bespoke in that they are tuned to the input conditions and/or turbine model of their training set. This work strives to generalize load surrogates by basing them on properties of the rotor plane velocity field. Specifically, this work uses statistical moments of vertical, horizontal, and $\pm 45^{\circ}$ profiles of the velocity components.The contributions in this work are the novel rotor plane velocity signal processing inputs to a loads surrogate model. The concept of tying turbine loads to rotor plane velocity profiles is not unique. Others have proposed using the rotor as a sensor of inferred inflow velocity states, averaged over rotor plane sectors, based on real-time load and power measurements. The goal was to inform operational control decisions to maximize power and redirect turbine wakes [1, 2, 15]. This paper essentially flips the inputs and outputs of the prior work, as the presented process starts with array flow models instead of real-time turbine load measurements. The objective is to provide a generic methodology to include wind turbine structural loads into wind farm optimization studies. This objective is accomplished by constructing surrogate models for several wind turbine load channel statistics based on inflow velocity field properties at the rotor plane, just upstream of a wind turbine. This will allow for the use of such surrogate loads in a wider variety of applications, including popular analytical wind farm wake models that model the mean flow, with less dependence on array geometry or inflow condition.

2 Approach

2.1 FAST.Farm Model

FAST.Farm is a multiphysics engineering tool that accounts for wake interaction effects on turbine performance and structural loading within wind farms [14]. FAST.Farm extends the National Renewable Energy Laboratory's OpenFAST tool, which solves the aero-hydro-servo-elasto dynamics of individual turbines, to include wake deficits, advection, deflection, meandering, and merging for wind farms. Ambient wind inflow is generated synthetically using TurbSim [13], which creates two-dimensional turbulent flow planes that are convected through the domain with the freestream velocity. Turbulence is simulated using the Kaimal spectrum and exponential coherence model. Turbulent inflows were generated for all of the parametric inflow conditions, including six turbulent seeds. To first focus on common turbine load channels and surrogate methodology development, monopile degrees of freedom and hydrodynamic loads were not simulated.

2.2 Surrogate Models

The most basic surrogate models are linear regression or piecewise linear interpolation, but these rarely capture nonlinear physical trends well. Instead, nonlinear surrogate models were constructed using different state-of-the-art methods implemented in the open-source Surrogate Modeling Toolbox [4]. Each of these methods supplies gradients of the outputs with respect to both the inputs and training data, which makes these models well-suited for later use in gradient-based optimization.

Multiple different surrogate models are studied in this paper, including inverse distance weighting (IDW) [19], radial basis functions (RBF) [16], Kriging with a partial least squares dimension reduction (KPLS) [3], simple neural network (NN) [10], and regularized minimal-energy tensor-product b-splines (RMTB) [12]. These methods were selected because they span the spectrum of training requirements and relative accuracy, with each having individual strengths and weaknesses as detailed in Table 1. The best surrogate method varies on a problem-by-problem basis, with the focus in this paper being to find the best methods for loads surrogate models.

Table 1. Considered surrogate models and associated strengths and weaknesses.

Surrogate model	Strengths	Weaknesses
IDW	Requires no training, has a low prediction cost, and generally has accurate surrogate prediction for loads	Derivatives at the training points are 0, which is problematic for later use in optimization
RBF	Simple surrogate that only has one tuning parameter	Prone to overfitting and producing undesired oscillations in the output model
KPLS	Well-suited for high-dimensional input spaces	Has accuracy issues when training points are too close to each other
NN	Good blend of accuracy and computational cost	Many tuning parameters
RMTB	Training cost scales well with the number of training points	Limited to approximately four input dimensions—could not be used for all input combinations considered in this work

All of these surrogate models have a different set of tuning parameters that affect how well the trained surrogate captures the truth data set. For example, the optimal distance order norm tuning parameter in IDW can be determined using a numerical optimizer that minimizes the percent error of a testing data set. These parameters were not exhaustively tuned for the comparison studies. Once the best-performing surrogate model was selected, its tuning parameters were optimized. The surrogate models studied here capture the state-of-the-art methods for nonlinear model prediction. The efficacy of these surrogate models to predict turbine loads are detailed in Section 3.

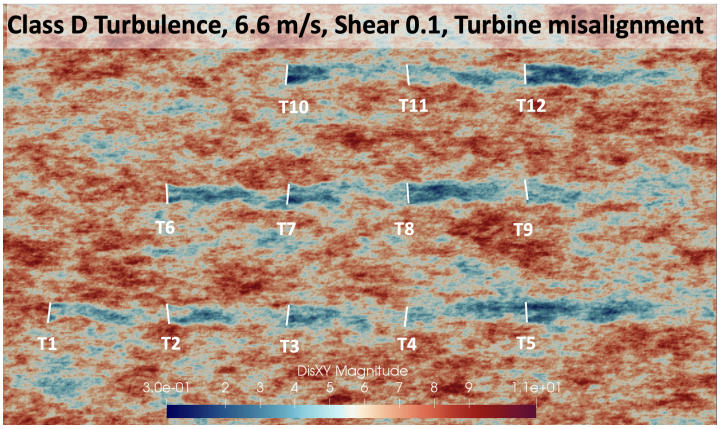


Figure 1. Flow visualization of the 12-turbine array in a turbulent FAST.Farm simulation (contours of velocity magnitude at hub height).

2.3 Surrogate Training Data Set, Inputs, and Outputs

Turbine load surrogate data were generated using a 12-turbine array with a different number of turbines per row to create varying wake overlap arrangements. The array, shown in Figure 1, featured

the International Energy Agency Wind Technology Collaboration Programme (IEA Wind) 15-MW turbine [8] with 7.72 D spacing in both the row and column directions. A parametric, design-of-experiments sweep was used across inflow hub-height velocity (V), inflow angle (Θ), shear exponent (α), turbulence intensity (TI), and turbulence seeds (S). The velocity values were selected to be centered around the rated conditions of 10.6 m/s. The inflow angle values include both positive and negative values to capture the non-symmetric effects on wake deflection, and thus the effect on downstream turbines. The total number of simulations performed in this work is $V \times \Theta \times \alpha \times TI \times S = 5 \times 21 \times 3 \times 3 \times 6 = 5,670$ different FAST.Farm simulations for the ranges shown in Table 2. The resulting temporal statistics were averaged across the seeds. Data were excluded if a FAST.Farm simulation failed to converge or became numerically unstable. With 12 turbines in each simulation, this yielded $5,670/6 \times 12 = 11,340$ unique rotor plane input fields and output time histories for all load channel inputs. From these simulations, load surrogate models were created for multiple load channels using time histories from all turbines and inflow cases. The load channels, or quantities of interest (QoI), used are listed in Table 3. The time histories were

Table 2. Surrogate training data parameter space.

Parameter	Values
Hub-Ht. Velocity [m/s]	6.6, 8.6, 10.6, 12.6, 16.6
Inflow angle [deg]	-20:2:20
Shear Exponent [-]	0.1, 0.2, 0.3
Turbulence Intensity [-]	Class B, C, D

Table 3. Surrogate output quantities of interest (QoI) and their vector components.

Quantity of Interest	Components	
Rotor power (*RotPwr*)		
Rotor speed (*RotSpeed*)		
Rotor torque (*RotTorq*)		
Rotor thrust (*RotThrust*)		
Blade-root moments	total bending (*RootMXY*)	pitching (*RootMzc1*)
Low-speed shaft moment	total bending (*LSSGagMXY*)	
Tower-top moments	total bending (*YawBrMXY*)	yaw (*YawBrMZp*)
Tower-base moment	total bending (*TwrBsMXY*)	yaw (*TwrBsMzt*)

summarized by their mean, standard deviation, ultimate load (maximum absolute value), and short-term fatigue DELs, though DELs and ultimate loads were not computed for the rotor quantities. These QoI statistics give 40 output dimensions for which surrogate models could be constructed.

Mean time-averaged and standard deviation statistics were calculated using standard methods for each time history and then averaged across all seeds. Ultimate loads were computed by averaging the sum of the maximum values from each seed. Short-term DELs of a given QoI were computed directly from the time series as an aggregate across all seeds using the NREL postprocessing tool, MLife. [11] Rainflow counting was used to bin a histogram of load cycle amplitudes over the time series. In this case, no Goodman or other mean-value correction was applied to the histogram amplitudes. The short-term DELs were computed from the histogram bins and the material-specific Wöhler exponent, m, from classical S-N fatigue theory. As is standard practice, $m = 4$ was used for loads on steel components (main shaft, tower, and monopile), and $m = 10$ was used for blade loads as a composite material. For fatigue of vector-component QoI, such as blade-root and tower bending moments, a load rose analysis was applied with a 10° increment, from which the maximum DEL around the rose was selected.

The thesis of this work is that rotor plane velocity profiles are suitable as generic inputs to a surrogate model for turbine loads. To that end, the potential surrogate input parameters are the velocity components at vertical, horizontal, and $\pm 45^\circ$ slices across the rotor plane. These slices are shown in Figure 2(c) on top of the time-averaged T1 rotor plane u-velocity. Standard deviations and absolute maximum values were also considered, resulting in three potential temporal statistics. Sampling the rotor plane velocity field in FAST.Farm is most easily done by exporting the

instantaneous inflow wind velocity of all 50 blade stations at each time step. These values neglect blade and induced velocity effects, and are purely the inflow values including any wake effects from upstream turbines. Then, for each blade station (radial position), the data is binned into 1° segments by the corresponding azimuthal position, as depicted in Figure 2(a). Here, each line shows the time-averaged value at each azimuthal angle for a given blade station. From here, the time-averaged planar velocity across varying rotor plane cross sections can be constructed by pulling out the values at specific azimuthal angles. For example, the horizontal rotor plane cross section is taken to be the time-averaged values from Figure 2(a) at 270° and 90° (0° is the blade pointing vertically up), as shown in Figure 2(b). This same process is repeated for the vertical cross section (0° and 180°); $+45^\circ$ cross section (225° and 45°); and -45° cross section (315° and 135°). On each slice, first, second, and third spatial moments of the velocity components are computed. The three velocity components, four slices, and three spatial moments give 36 possible turbine-inflow combinations to be used as input dimensions.

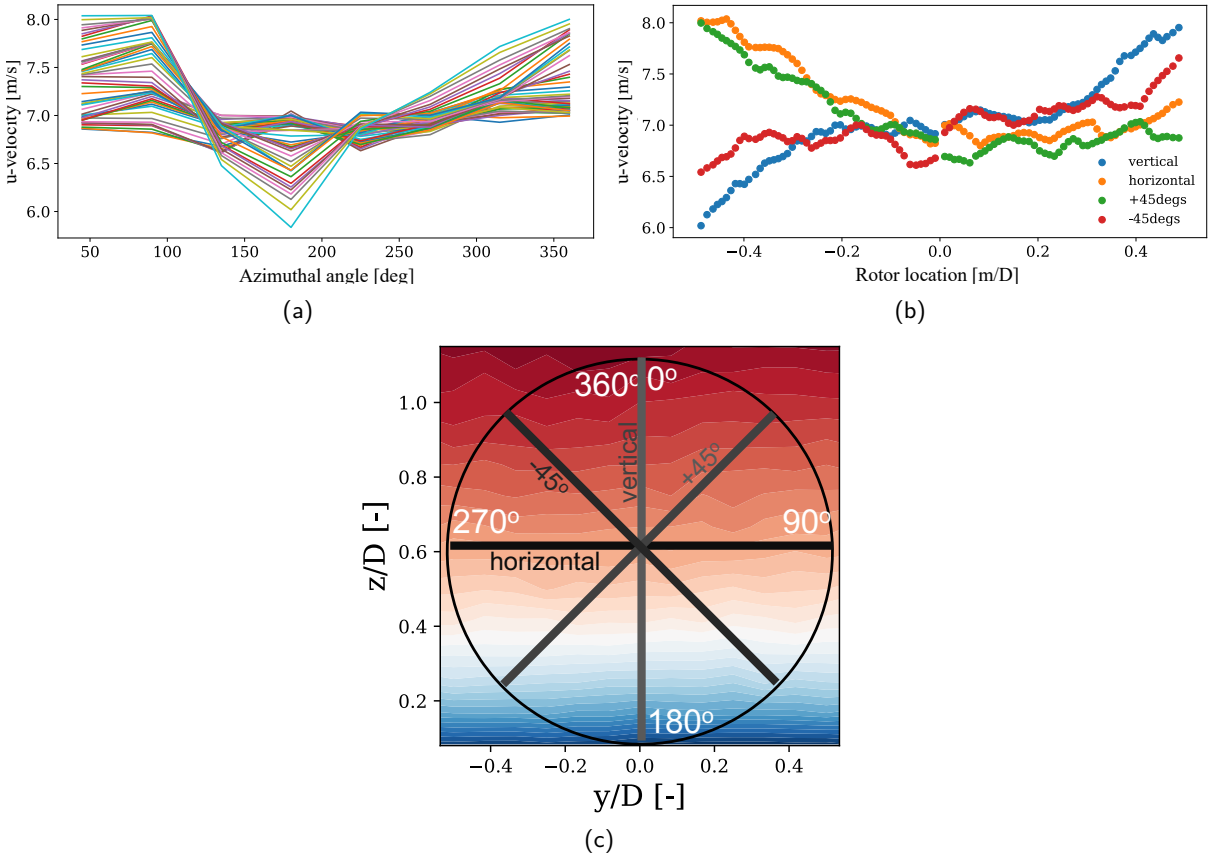


Figure 2. Inflow statistics generation process. (a) Inflow rotor plane and velocity profile "slice" visualization; (b) Time- and azimuthally-averaged results vs spatial location; and (c) Time- and azimuthally-averaged results vs spatial location.

To determine how much information about the inflow is required to construct accurate surrogates, many surrogates were created for different inflow parameter combinations. The tested input parameter combinations are summarized in Table 4. The rows in this table show all 24 available time-averaged surrogate input options, comprised of spatial moments for each velocity component and spatial slice. Each case number shows which of those 24 input options was used to construct a given surrogate, leading to 12 input parameter combinations using only temporal means. This

input parameter grouping was repeated for each temporal statistic, with the first set of 12 using only temporal means of the inflow quantities, the second set adding temporal standard deviations of the inflow quantities (in addition to the temporal means), and the final set adding temporal maxes of the inflow quantities (in addition to the temporal means and standard deviations). This test matrix resulted in a total of 36 input combinations or cases.

Table 4. Sample input parameter combinations to be tested for surrogate generation.

Component	Slice	Spatial Mom.	Time stat.	Case Number											
				1	2	3	4	5	6	7	8	9	10	11	12
u-velocity	horizontal	1st	mean	x	x	x	x	x	x	x	x	x	x	x	x
		2nd	mean			x	x	x	x			x	x	x	x
		3rd	mean					x	x					x	x
	vertical	1st	mean	x	x	x	x	x	x	x	x	x	x	x	x
		2nd	mean			x	x	x	x			x	x	x	x
		3rd	mean					x	x					x	x
	$+45°$	1st	mean		x		x		x		x		x		x
		2nd	mean				x		x				x		x
		3rd	mean						x						x
	$-45°$	1st	mean		x		x		x		x		x		x
		2nd	mean				x		x				x		x
		3rd	mean						x						x
v-velocity	horizontal	1st	mean							x	x	x	x	x	x
		2nd	mean									x	x	x	x
		3rd	mean											x	x
	vertical	1st	mean							x	x	x	x	x	x
		2nd	mean									x	x	x	x
		3rd	mean											x	x
	$+45°$	1st	mean								x		x		x
		2nd	mean										x		x
		3rd	mean												x
	$-45°$	1st	mean								x		x		x
		2nd	mean										x		x
		3rd	mean												x

2.4 Surrogate Generation and Testing

The parametric approach yielded $11,340$ rotor plane velocity fields, 36 input dimension combinations, and 40 output dimensions. The $11,340$ cases can be separated into different training and testing sets, in which the training set is used to generate the surrogate model, using one of the five surrogate approaches mentioned in Section 2.2, and then tested against the FAST.Farm results of the remaining cases. Through this process, mutually exclusive data sets were used to verify the accuracy of the trained surrogate models.

Surrogate model performance was measured by comparing the predicted load channel statistic against the true FAST.Farm value at every point in the testing data set. For this vector of comparison values, accuracy was quantified using the classical r^2 correlation coefficient and the average percentage error. The r^2 correlation coefficient measures the linear dependence between two variables and is therefore a useful measure of how well the surrogate tracks the trend of the truth values. The average percent error was used as the most intuitive measure of accuracy, though artificially large errors are predicted when the truth values are small.

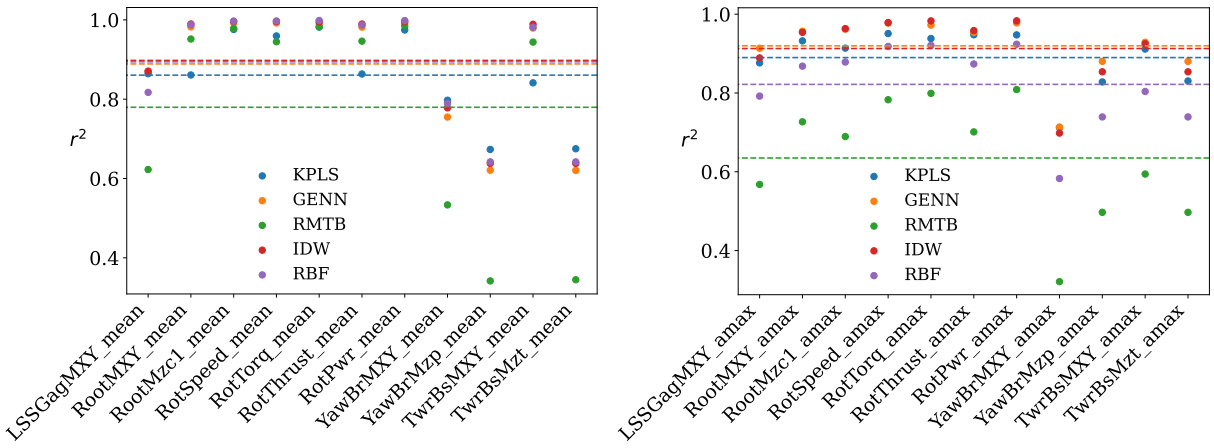
To help improve surrogate model accuracy, training points were first selected from the minimum and maximum inflow cases. Specifically, from the -20 and $+20$ degree inflow angle, 0.1 and 0.3 shear exponent, class B and D turbulence, and 6.6 and 16.6 m/s cases. These data points define

the bounds of the input space for the surrogate space, so always including them in the training data set eliminates extrapolation when querying the surrogate models.

3 Results

The accuracy of the surrogate model types listed in Section 2.2 is shown in Figure 3 and tabulated in Table 5. This comparison was performed using all 36 cases of rotor plane descriptor sets from Table 4. Each surrogate was trained with $1,000$ randomly selected points and tested using $1,000$ random other points for each surrogate model. Not all available training data were used to keep the computational time tractable. For each surrogate, the average r^2 values across the input cases are shown via scatter points for the temporal means and maxes of all considered QoI. The dashed lines indicated the average r^2 value across all QoI.

IDW and NN clearly separate themselves as better performers, with IDW having the highest average numerical r^2 performance across all output statistics (0.853) and the lowest computational cost (0.00112 seconds/surrogate) of all the surrogate models. The IDW method has better prediction qualities than other methods when the input data are irregularly spaced, which is the case for the velocity and moments presented here. The NN method had similarly high performance, but required $10,000\times$ longer to generate and test each surrogate. Although IDW predicts the actual QoI well, the gradients at each training point are zero by definition, which is physically unrealistic. Thus, for use in gradient-based optimization, special care must be used when adopting IDW to avoid numerical difficulties caused by this. Based on the results shown in Figure 3 and Table 5, the remainder of the paper focuses on results found using IDW.



(a) Means (b) Ultimate loads

Figure 3. Correlation coefficient of (a) temporal mean and (b) temporal max of QoI for several surrogate methods. Correlation coefficients for each surrogate method computed as the average of all 36 cases described in Table 4. Dashed lines represent average correlation coefficient for all QoI values. A higher r^2 value indicates a better capture of output trends.

To determine surrogate convergence, r^2 values are computed over a variable number of training points for all QoI statistics. Percent error results are not included in this paper because of space limitations and they did not show appreciably differrent results from those of r^2. These convergence plots are shown in Figure 4. For the sake of clarity, only Case 3 results are shown. For a given training point set, the r^2 values are computed against the same set of $1,000$ testing points. For all QoI statistics, increasing training points increases the accuracy of the surrogate, with rapidly decreasing benefit to including more points. For QoI temporal mean values, the r^2 values quickly converge to > 0.9 by 100 training points, with the exception of tower base yaw moments. For the

Table 5. Computational timings for surrogate generation and testing averaged per case and resulting average r^2 for each surrogate model and output subset. These values are for off-the-shelf surrogate models without extensive tuning.

Model	Timing [cpu-seconds]	r^2 means	r^2 std devs	r^2 maxes	r^2 DELs
IDW	0.001 12	0.896	0.841	0.879	0.795
RBF	0.0112	0.801	0.491	0.533	0.384
RMTB	1.51	0.813	0.708	0.781	0.613
KPLS	7.19	0.877	0.763	0.861	0.768
NN	18.2	0.878	0.832	0.875	0.792

tower-base yaw moment, r^2 continues to improve with additional training points and reaches up to 0.8. For QoI standard deviation, ultimate loads, and DEL results, r^2 convergence was reached by most QoI by 500 training points, at which point most QoI had an r^2 of at least 0.8. Exceptions to this were tower loads and rotor thrust, to different extents based on the statistical quantity. Surrogate convergence was overall reached by 500 training points, though some QoI, especially tower loads, benefited from additional samples. In addition to being slow to converge, the tower loads tended to be the least accurate for all QoI statistics, in particular the yaw moments. To capture the highest accuracy possible for this study, the remaining results are based on surrogates generated with 2000 training points.

Shown in Figure 5 are r^2 values for each considered input case (Cases 1–36) and each QoI. Cases are grouped such that the first 12 cases use only temporal means of the rotor plane velocity quantities, the next 12 cases add temporal standard deviations as potential input parameters, and the final 12 cases add temporal maximum values as potential input parameters. These groupings are designated by the red vertical dashed lines. For the mean QoI results, high r^2 values are observed for almost all input cases and all QoI. Exceptions to this are the same QoI that showed slow or poor convergence in Figure 4: tower-top total bending and yaw moments, tower yaw moments, and low-speed shaft bending. For these QoI, r^2 values are the highest overall when u-velocity temporal means and standard deviations of the horizontal and vertical planes are included (Cases 3, 15, and 27). Similar accuracy is reached when the corresponding v-velocity components are included, but the addition of this velocity component does not itself increase the accuracy. Thus, this peak in accuracy is likely due to the additional cross-plane slices, and not the additional velocity components or temporal statistics. The best and worst performing cases for each QoI-statistic combination, as measured by r^2 values, are identified in Table 6. When cross-referenced with Table 4, the rotor plane velocity features that are the key drivers of surrogate performance are evident. For the sake of brevity, not all QoI-statistic combinations will be explained in depth. Overall, it is apparent that the best r^2 values for all QoI statistics are quite close to 1.0, with some of the tower statistics topping out at 0.93. This underscores that the rotor plane velocity quantities can serve as accurate inputs for turbine load surrogates.

For the QoI standard deviation results, the best surrogate performance usually occurred with Case 13, which incorporated the first moment of the temporal standard deviation and mean along the horizontal and vertical cross planes. For the maximum values, the best performance does not occur when the temporal maximum rotor plane velocity quantities are included (Cases 25–36), but rather when just the temporal mean and standard deviation quantities are included (Cases 13 and 17). This supports the notion that the QoI maximum value can be thought of as a superposition of the mean value and some number of standard deviations. Therefore, if a set of rotor plane quantities can accurately predict both the mean and standard deviation of a QoI, then it should also be able to predict the maximum value. However, the temporal maximums are leveraged for the most accurate DEL surrogates of the tower load channels (Cases 25 and 26). DEL values also benefited from including the third moment (skew) of the rotor plane velocity slices. This is perhaps due to the more

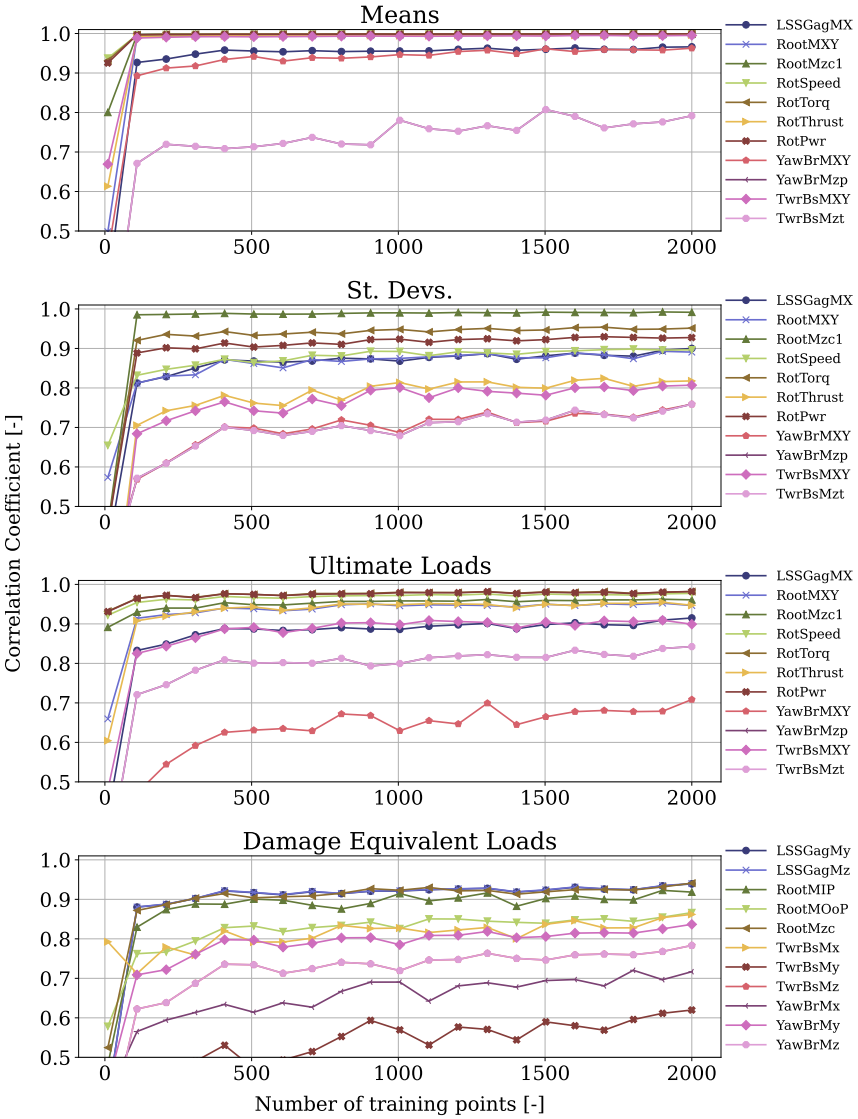


Figure 4. Prediction performance convergence as measured by r^2 values; only Case 3 results are shown.

complex calculation procedure for DELs where rainflow counting is used to bin load oscillations by amplitude.

Table 6 is also informative as it indicates when the surrogates do not perform well. Performance can drop significantly if either insufficient or excessive numbers of inputs are provided to the surrogates. Poor prediction accuracy occurs for the standard deviation, maximum, and DEL surrogates when they are limited to mean-value-only temporal statistics of the rotor plane velocity quantities. Cases 8 and 12 are commonly listed as the worst performing for these statistics and interestingly are the ones that also include the v-component of velocity parameters. For the QoI mean values, there was also a tendency to perform poorly when the surrogate was flooded with too many input parameters, either from statistics beyond the temporal mean or with v-component parameters. This was likely because the u- and v-velocity components are statistically correlated, so including them both in the surrogate model inputs detracted from the ability to fully capture the key, independent flow features. It is also possible that surrogate models other than IDW would be

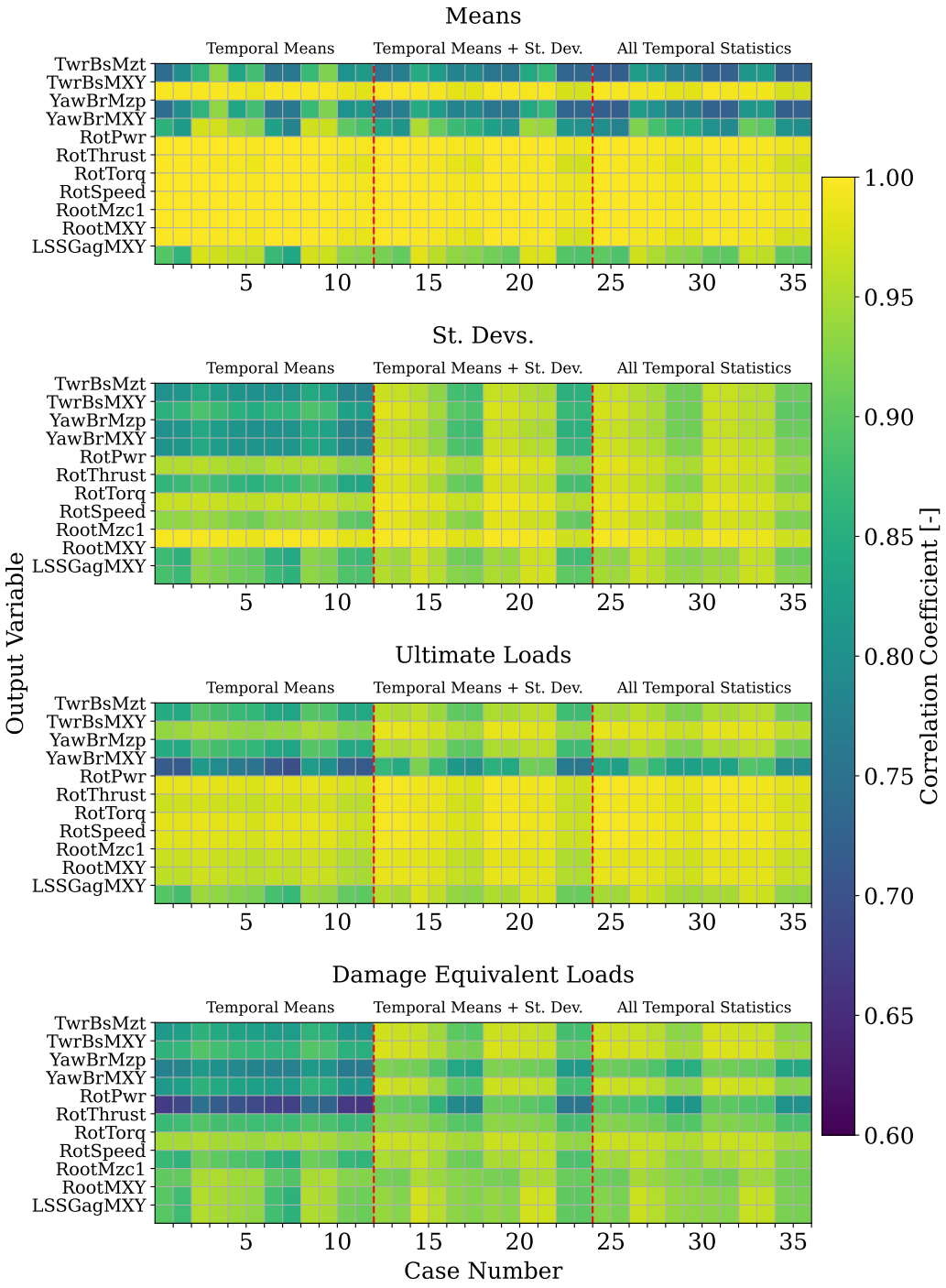


Figure 5. Prediction accuracy measured by r^2 for each QoI with varying inflow parameters. Each pixel corresponds to an inflow case and QoI combination, with each subplot showing a different QoI statistic. The vertical dashed lines indicate case groupings: the first 12 cases use only temporal means as potential input parameters; the next 12 cases add temporal standard deviations as potential input parameters; and the final 12 cases add temporal max values as potential input parameters.

Table 6. Cases that produce the best and worst surrogate models based on r^2 value.

QoI	Performance	Mean	St. Dev.	Max	DEL
Blade-root total bending moment	Best	1 (1.0)	17 (0.97)	13 (0.99)	4 (0.95)
	Worst	12 (0.98)	8 (0.84)	12 (0.95)	8 (0.89)
Blade-root pitching moment	Best	3 (1.0)	13 (1.0)	14 (0.99)	13 (0.97)
	Worst	12 (1.0)	36 (0.99)	35 (0.97)	12 (0.93)
Low-speed shaft total bending moment	Best	17 (0.98)	17 (0.97)	17 (0.98)	17 (0.97)
	Worst	8 (0.84)	8 (0.86)	8 (0.87)	8 (0.86)
Tower-top total bending moment	Best	3 (0.97)	13 (0.97)	17 (0.92)	26 (0.98)
	Worst	16 (0.78)	12 (0.78)	8 (0.7)	12 (0.85)
Tower-top yaw moment	Best	4 (0.93)	13 (0.97)	17 (0.96)	26 (0.97)
	Worst	14 (0.72)	12 (0.78)	8 (0.83)	12 (0.81)
Tower-base total bending moment	Best	3 (1.0)	13 (0.98)	13 (0.99)	25 (0.93)
	Worst	35 (0.97)	12 (0.82)	12 (0.92)	12 (0.87)
Tower-base yaw moment	Best	4 (0.93)	13 (0.97)	17 (0.96)	26 (0.97)
	Worst	14 (0.72)	12 (0.78)	8 (0.83)	12 (0.81)

less sensitive to the inclusion of the v-component parameters.

4 Conclusions

This work detailed an analysis of surrogate models built using rotor plane inflow statistics to predict mean, standard deviation, ultimate, and fatigue loads of wind turbines—both waked and non-waked. This paper focused on demonstrating the feasibility of this approach by identifying an appropriate surrogate model, understanding the requirements to achieve good predictive performance, and identifying which rotor plane velocity parameters were most appropriate for the prediction of various QoI statistics. The results suggest that these surrogates could be easily paired with analytic wake models, which are frequently used for pre-construction wind farm array optimization, to account for turbine loading in addition to power production.

Future work will focus on testing the generality of the surrogate models demonstrated in this paper, including performing this study with a randomly generated sample space of inflow conditions and turbine array layout instead of the parametric approach used here. This will likely allow for fewer training points to be needed while providing a more representative sample of potential inflow conditions that a wind farm can experience. The effects of yaw misalignment should also be considered in future work as well as nondimensionalizing the surrogates such that they are independent of the turbine to further increase the surrogate generalization. Finally, the aim is to incorporate these surrogates into lower-fidelity, steady-state wind farm simulation models, which are often used for array layout optimization, to allow for the inclusion of fatigue and ultimate loads in the wind farm design process. Examples of these types of models include open-source tools such as FLORIS and PyWake or commercial tools such as OpenWind, WindFarmer, WAsP, WindPRO, and others.

Acknowledgments

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Wind Energy Technologies Office. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes. The research was performed using computational resources sponsored by the Department of Energy's Office of Energy

Efficiency and Renewable Energy and located at the National Renewable Energy Laboratory.

References

[1] M. Bertelè, C. L. Bottasso, S. Cacciola, and et al. Wind inflow observation from load harmonics. *Wind Energy Science*, 2(2):615–640, 2017.

[2] C. L. Bottasso, S. Cacciola, and J. Schreiber. Local wind speed estimation, with application to wake impingement detection. *Renewable Energy*, 116:155–168, 2018.

[3] M. A. Bouhlel, N. Bartoli, A. Otsmane, and et al. Improving kriging surrogates of high-dimensional design models by partial least squares dimension reduction. *Structural and Multidisciplinary Optimization*, 53(5):935–952, 2016.

[4] M. A. Bouhlel, J. T. Hwang, N. Bartoli, and et al. A Python surrogate modeling framework with derivatives. *Advances in Engineering Software*, 135, September 2019.

[5] C. E. Clark, G. Barter, K. Shaler, and et al. Reliability-based layout optimization in offshore wind energy systems. *Wind Energy*, pages 1–24, 2021.

[6] N. Dimitrov. Surrogate models for parameterized representation of wake-induced loads in wind farms. *Wind Energy*, 22(10):1371–1389, 2019.

[7] N. Dimitrov, M. C. Kelly, A. Vignaroli, and et al. From wind to loads: wind turbine site-specific load estimation with surrogate models trained on high-fidelity load databases. *Wind Energy Science*, 3(2):767–790, 2018.

[8] E. Gaertner, J. Rinker, L. Sethuraman, and et al. Definition of the iea 15-megawatt offshore reference wind turbine. Technical Report NREL/TP-5000-75698, National Renewable Energy Laboratory, Golden, CO, March 2020.

[9] C. Galinos, N. Dimitrov, T. J. Larsen, and et al. Mapping wind farm loads and power production-a case study on horns rev 1. *Journal of Physics: Conference Series*, 753(3):032010, 2016.

[10] S. Haykin. Neural networks: a comprehensive foundation. 1999. *Mc Millan, New Jersey*, pages 1–24, 2010.

[11] G. J. Hayman. Mlife theory manual for version 1.00. Technical Report NREL/TP, National Renewable Energy Laboratory, Golden, CO, October 2012.

[12] J. T. Hwang and J. R.R.A. Martins. A fast-prediction surrogate model for large datasets. *Aerospace Science and Technology*, 75:74–87, 2018.

[13] B. J. Jonkman. Turbsim user's guide: Version 2.00.00. Technical report, NREL, Golden, CO, June 2016.

[14] J. Jonkman and K. Shaler. Fast.farm user's guide and theory manual. Technical Report NREL/TP-5000-78485, National Renewable Energy Laboratory, Golden, CO, April 2021.

[15] N. Namura. Wind shear estimation model using load measurement of wind turbine tower and surrogate model. *Wind Energy*, 23(2):327–339, 2020.

[16] M. Powell. The theory of radial basis function approximation in 1990. *Advances in numerical analysis*, pages 105–210, 1992.

[17] P.-E. Réthoré, P. Fuglsang, G. Larsen, and et al. Topfarm: Multi-fidelity optimization of wind farms. *Wind Energy*, 17(12):1797–1816, 2014.

[18] R. Riva, J. Liew, M. Friis-Møller, and et al. Wind farm layout optimization with load constraints using surrogate modelling. *Journal of Physics: Conference Series*, 1618(4):042035, sep 2020.

[19] D. Shepard. A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM National Conference*, ACM '68, page 517–524, New York, NY, USA, 1968. Association for Computing Machinery.

[20] A. P. J. Stanley, J. King, C. Bay, and et al. A model to calculate fatigue damage caused by partial waking during wind farm optimization. *Wind Energy Science*, 7:433–454, 2022.