Uncertainty Quantification in Wind Plant Energy Estimation

Preprint

Anna E. Craig

National Renewable Energy Laboratory

*Presented at American Institute of Aeronautics and Astronautics
SciTech Forum
San Diego, California
7-11 January 2019*

**NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC**

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-5000-72741
January 2019

Uncertainty Quantification in Wind Plant Energy Estimation

Preprint

Anna E. Craig

National Renewable Energy Laboratory

Suggested Citation

Craig, Anna E., 2019. *Uncertainty Quantification in Wind Plant Energy Estimation: Preprint*. Golden, CO: National Renewable Energy Laboratory. NREL/CP-5000-72741. <https://www.nrel.gov/docs/fy19osti/72741.pdf>.

**NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC**

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-5000-72741
January 2019

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

NOTICE

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. This work was supported by the Laboratory Directed Research and Development (LDRD) Program at NREL. The views expressed herein 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.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

U.S. Department of Energy (DOE) reports produced after 1991 and a growing number of pre-1991 documents are available free via www.OSTI.gov.

Cover Photos by Dennis Schroeder: (clockwise, left to right) NREL 51934, NREL 45897, NREL 42160, NREL 45891, NREL 48097, NREL 46526.

NREL prints on paper that contains recycled content.

Uncertainty Quantification in Wind Plant Energy Estimation

Anna E. Craig*

National Renewable Energy Laboratory, Golden, CO, 80401

Data characterizing a wind power plant's production are important for the validation, calibration, and refinement of the wake loss models used to simulate performance for preconstruction energy estimates. In this process, there is the potential for significant variation in the methods used by an analyst for, first, the preparation of the operational production data for comparison to model data and, second, the generation of analogous production data from simulations. In this work, the uncertainty introduced by these method variations is quantified by considering ensembles of analysis methods. Specifically, 1920 method variations were developed for preparing the operational data as a predictable production benchmark and simulations taking as input sample sets of the observed wind resource ranging in size from 4% to 125% of a full year were run using 3 engineering wake loss models in simulations run in the FLOW Redirection and Induction in Steady State software package.

The benchmark and simulation input uncertainties place lower bounds on the precision of the calibration between the simulation model estimates and operational benchmark data. Even within subsets of method variations (representative of what might be seen within a single company), the uncertainty in the predictable expected production is estimated to be between 1% and 3% of the median predictable expected production, while the uncertainty in the simulated expected production remains at best on the order of 2%. To put this into context, the simulated expected production is very closely related to the annual energy production metric commonly used to quantify a plant's predicted performance and a 3% error in the preconstruction annual energy production estimate can result in a \$17 million loss due to additional financing costs for a typical 100-MW Texas plant.

I. Introduction

DATA collected during the operation of wind plants are used in many diverse contexts, such as identifying malfunctioning or underperforming turbines [1], evaluating the effectiveness of plant repowering options [2], providing operational reforecasts for increased accuracy and decreased uncertainty over the preconstruction production estimates [3], and providing training data sets or benchmarks for calibration and refinement of power production models (e.g., time-series energetic availability models, turbine wind flow/wake loss models, net power production models) [4, 5]. The broad range of possibilities introduces the question of consistency and reproducibility of the operational data analysis used by different analysts or within different companies. Examples of such choices for operational data analysis include flagging of "noise" in the data (e.g., sensor malfunction flagging, abnormal operation flagging, multiple versus single flagging algorithms applied to the data, and the order of flagging algorithms applied), deciding what data sources are considered (e.g., wind speed, wind direction, turbulence intensity, air density, wind shear, wind veer, blade pitch, neighboring turbine operational status, curtailment logs, status codes for the turbines, and power), selecting what models to fit to the data (e.g., parametric versus nonparametric versus machine-learned, number of data channels to take as input to the model, and whether or not to generate multiple models for different operational conditions), characterizing the concurrently observed wind resource (e.g., combinations of nacelle measurements, meteorological tower measurements, and spatial/temporal resolution of the wind resource as observed), and interpolating missing or flagged data (e.g., time-series interpolation, correlation with associated assets, or dropping missing data).

The present work quantifies the variation introduced into an analysis by the choices made in developing the analysis methodology. To accomplish this, ensembles of methodologies were developed, all designed to achieve the same analysis result, but varying in the choices made to achieve that result. Each method in the ensemble may be considered representative of what could be developed by a "reasonable" analyst.

Again, there are many possible goals that an analyst may have for processing the operational data and each analysis goal will be best achieved by different choices made within the analysis method. For the purposes of this study,

*Director's Fellow

methodologies designed to use the operational data to validate, calibrate, and refine low-fidelity wake loss models were developed. To use the data in this context, the full analyst method was divided into two submethods; the first concentrates on the preparation of turbine-level power production data and the second concentrates on the generation of analogous simulation turbine-level production data.

The goal of the first submethod is taken to be the adjustment of the raw, supervisory control and data acquisition (SCADA) reported power production of a turbine to match what could have been predicted preconstruction via a simulation model, that is, adjusting for “unpredictable” curtailment and availability losses but retaining the “predictable” production under different wind conditions (including wake losses). As illustrated in Fig. 1, this adjustment is accomplished by preprocessing the data to flag the “unpredictable” data, fitting the “predictable” data to a power curve model, and then evaluating the model, taking as input the “observed” wind resource over the operational period. The details associated with the diverse range of choices made to develop methods for accomplishing the goal are presented in section II.B.

The goal of the second submethod is taken to be the characterization of the “observed” wind resource to be used as input to the simulation model and then the use of simulation models to generate a prediction of the power production by the turbines over the operational period. By considering the observed wind resource as input to the simulation model, the potential for discrepancy between the simulated production and operational-data-derived predictable production due to wind resource forecasting uncertainty is effectively removed, thus allowing the accuracy of the wake loss simulation model to be evaluated independently from the accuracy of the preconstruction wind resource forecasting models. Again as illustrated in Fig. 1, this adjustment is accomplished by preprocessing the data to flag any “noise,” selecting data channels to provide a spatial/temporal characterization of the “observed” wind resource, sampling the wind resource characterization for input to the simulation model, and then evaluating the model. The details associated with the diverse range of choices made to develop methods accomplishing goal are presented in section II.C.

Fig. 1 Submethods required for comparison of operational and simulation data sets.

It should be emphasized that the goal of this work was not to develop new methods nor to directly evaluate the accuracy of existing methods within each of the submethodologies. Rather, the goal of the present work may be considered more of a “meta-analysis,” creating an ensemble of analysis methods representing the possible variations introduced by the analyst in the processing of the operational data and generation of the simulation data. By comparing the final analysis results across the ensembles of methods, the uncertainty in the results due to analyst choice may be quantified. The results of this uncertainty quantification analysis are presented in section III.

II. Methods

This section details the development of ensembles of methodologies to use operational data to validate, calibrate, and refine low-fidelity wake loss models. This analysis goal is broken down into two subanalyses—preparation of turbine-level power production data and generation of analogous simulation turbine-level production data (see Fig. 1). Ensembles of possible variations were developed for each submethod and each individual method variation may be considered representative of what could be developed by a “reasonable” analyst.

A. Data set descriptions

The plant, denoted z01, for which the first ensemble of methods was developed consisted of 100 turbines and four permanent meteorological towers. Data were available for the years 2014–2016 and included SCADA measurements

from each of the turbines and environmental measurements from the permanent meteorological towers, both reported as averages over 10-minute intervals. The second plant, denoted o46, consisted of 101 turbines and one permanent meteorological tower. Data were available for the years 2015–2018 and included SCADA measurements and a general status code for each of the turbines and environmental measurements from the permanent meteorological tower, both reported as averages over 10-minute intervals. The third plant, denoted h21, consisted of 137 turbines and four permanent meteorological towers. Data were available for the years 2014–2018 and included SCADA measurements and a general status code for each of the turbines and environmental measurements from the permanent meteorological tower, both reported as averages over 10-minute intervals. These data were supplemented by time-stamped status code changes for each turbine and hourly external historical meteorological data; in this case, the National Aeronautics and Space Administration’s Modern Era-Retrospective Analysis for Research and Applications-2 (NASA MERRA-2) data set was selected.

B. Predictable production evaluation

The goal of the first submethod is taken to be the adjustment of the raw, SCADA-reported power production of a turbine to match what could have been predicted preconstruction via a simulation model. In other words, data associated with losses not dependent on the wind resource (such as curtailment and availability losses) would be considered “unpredictable,” while all other data, especially including waked conditions, should be “predictable.” As illustrated in Fig. 2, following the collection of data, the goal of the data preprocessing stage was to remove erroneous measurements or measurements recorded during times of unpredictable operational conditions. In the processing stage, a power-production model was then fit to the predictable power production data for each turbine. The power-production model was then evaluated using a time series of wind resource inputs to yield the predictable power production over the full operational time period. Some of the method characteristics that are left to analyst choice are indicated in red in Fig. 2.

Fig. 2 Method flow for estimating a turbine’s predictable production from raw production data. Characteristics of the method that are subject to analyst decision are identified in red.

1. Data preprocessing

For this ensemble of methods, the SCADA measurement channels considered are the output power and nacelle wind speed. Data are treated on an individual turbine basis and are assumed to be time-history independent. It was assumed that all reasonable analysts would perform initial quality control of the data—for example, flagging measurements that did not change sufficiently over an extended period of time and thus could indicate a sensor failure, yielding at the end a similar set of flagged data. This source of potential variation between analysts was not examined.

Inspired by the International Electrotechnical Commission (IEC) standards and with input from industry partners [6–8], two core statistics-based algorithms were developed for flagging wind speed and power data. Note that data previously flagged by the quality control flags were removed from consideration so that the statistics were not skewed by the inclusion of outliers. The first flagging algorithm was flagged power measurements that, within wind speed bins (over some range of wind speeds), fall outside a threshold from the mean power within the bin. The second flagging algorithm was flagged wind speed measurements that, within power bins (over some range of powers), fall

outside a threshold from the mean wind speed within the bin. The thresholds may either be standard deviation of the measurements within the bin or a set scalar. These two threshold methods are treated as distinct algorithms, resulting in four possible algorithms in total, as indicated in Fig. 3. The parameters for each of the four algorithms include (1) minimum bin value, (2) maximum bin value, (3) bin width, and (4) threshold value. For each of these parameters, it is possible to define a range of possible reasonable values, for which the resulting algorithm will either remove more data (“tightest” parameter set, maximum/minimum of range for each parameter), less data (“loosest” parameter set, minimum/maximum of range for each parameter), or a medium amount of data (“base” parameter set, typically taken as the mean value over the possible range for each parameter). This resulted in three possible parameter sets for each of the four possible algorithms, yielding 12 possible variations. Only the data sets filtered with the “base” variations from the first pass of flags were then further processed, again applying one of the four flagging algorithms with one of the three parameter sets. Thus, overall there were 12 sets of single-pass flagged data and 48 sets of two-pass flagged data for a total of 60 data sets, as indicated in Fig. 3. The flagging algorithm parameters used and the resulting data flagged by each algorithm are provided in the appendix, section IV.A, as are examples of the data flagged and the remaining cleaned data as determined by random selections of the preprocessing methods.

Fig. 3 Predictable production evaluation—development of methodologies.

2. Data modeling

Applying all the flags computed in one variation of the data preprocessing stage yields a remaining data set that may be considered representative of predictable production and fit with a power curve model, $P_{turbine}(x_t)$. Four power curves were considered here—an IEC-inspired wind speed-binned step function [8], a logistics five parameter model [9, 10], a generalized additive model (which, when taking only one input, reduces to an n-spline model) [6, 7, 11], and a machine-learned ensemble extra trees regression model [6, 7, 11]. Details for these models may be found in the appendix, section IV.B. Additionally, noting the sensitivity of the model to the relative data sparsity at high wind speeds, an “asymptoted” modification of each model was also developed. For the asymptoted models, the base model was used for inputs between a lower bound wind speed and an upper bound wind speed. For all input wind speeds below the lower bound wind speed, the power production was set to zero. For all wind speeds above the upper wind speed, the power was set to a constant value, which was taken to be the median value predicted by the base model over some range of input wind speeds surrounding the upper wind speed bound. Examples of the predictable data sets to which the

curves were fit and the resulting power curve fits (as evaluated over a simple range of wind speeds) are given in Fig. 4 for a random turbine from each plant.

Fig. 4 Power curve models as fit to cleaned data sets. The blue hexbin plots indicate the cleaned data (with darker shades indicating more data) for a randomly selected turbine from each plant and with a random preprocessing method applied. The percentages in the upper left indicate the data remaining as a percentage of the total available data for the turbine. Each curve color indicates a different power curve model type, with the solid dashed line being the base model and the dashed line being the asymptote-enforced modification of the base model.

To evaluate the expected performance of a turbine over the operational time period, it is necessary to have a complete and reliable time history of the input measurements required by the power curve model. Unfortunately, it is often the case that the required inputs are incomplete over the operational time, due, for example, to sensor malfunction, such as may be determined in the quality control flagging of the data. Additionally, when nacelle measurements are used as the input (as was the case in this study) the measurements of the sensors may be highly inaccurate during times of abnormal turbine operation, even if the sensor itself is operating normally (e.g., if a turbine is nonrotating due to curtailment rather than low incoming wind speed, the anemometer measurement could be significantly different depending on whether it is in or out of a blade’s wake.) For this reason, it is generally agreed that some subset of the data flagged as abnormal operation should be imputed from other measurement channels. In this ensemble of methods, four possible time series of the wind resource were developed and used as input to the models.

The first time series of inputs was developed by taking a baseline estimate for the plant-wide wind resource as derived from the on-site meteorological towers* and applying a linear nacelle-transfer function (NTF) used to convert the “incoming” wind speed measurements to the equivalent nacelle-based wind speed measurements, which are used as input to the fit power curve models. This time series of inputs is referred to as the “full_NTF” imputation method. (Note that none of the nacelle-based wind speed measurements are directly used.) The second time series of inputs was composed primarily of the nacelle-based wind speed measurements, but the time stamps flagged by any of the quality control algorithms were replaced by the associated “full_NTF” input value. This method is referred to as the “QCgap_NTFfilled” imputation method. Similarly, the third time series of input was also composed primarily of the nacelle-based wind speed measurements, but the time stamps flagged by any of the quality control algorithms were replaced in this case by using a cascading hierarchy of linear transfer functions with other wind speed measurements in the array. The hierarchy of proxy measurements was determined by ordering the cross-correlation coefficients as computed between the (quality controlled) wind speed measurement of interest and all other (quality controlled) wind speed measurements in the array. This method is referred to as the “QCgap_crossCorrFilled” imputation method. Note that these three input time series are not dependent on the preprocessing of the data, but only on the quality control algorithms applied to the data. The final time series of inputs was specialized for each preprocessing of the data. In this case, the inputs were again composed primarily of the nacelle-based wind speed measurements, but the time stamps flagged by any of the quality control algorithms or specific preprocessing variation were replaced by the associated

*The method used to develop this baseline estimate will be discussed in section II.C

“full_NTF” input value. This method is referred to as the “PPgap_NTFfilled” imputation method. The distributions of the full time-series inputs as computed by the four methods are compared in the appendix, section IV.C.

3. Model evaluation

Fitting each of the 60 possible preprocessing data sets with each of the 8 possible power curve models and evaluating with each of the possible 4 time series of inputs yields 1920 estimates of the predictable operational production for each turbine in the array, as indicated in Fig. 3.

C. Simulated production generation

The goal of the second submethod is the characterization of the “observed” wind resource to be used as input to the simulation model and then the use of simulation models to generate a prediction of the power production by the turbines over the operational period. By considering the observed wind resource as input to the simulation model, the potential for simulation-predicted power production discrepancy (compared to the operational data) due to wind resource forecasting uncertainty is effectively removed, thus allowing the accuracy of the wake loss simulation model to be evaluated independently from the accuracy of the preconstruction wind resource forecasting models. Again, as illustrated in Fig. 5, this adjustment is accomplished by preprocessing the data to flag any “noise,” selecting and combining data channels to provide a spatial/temporal characterization of the “observed” wind resource, sampling the wind resource characterization for input to the simulation model, and then evaluating the model. Some of the method characteristics that are left to analyst choice are indicated in red in Fig. 5.

Fig. 5 Method flow for generating analogous simulated production for each turbine over the operational period. Characteristics of the method that are subject to analyst decision are identified in red.

1. Data preprocessing

For this ensemble of methods, the meteorological tower measurements considered are the 10-minute mean and standard deviations of wind speed, wind direction, temperature, and pressure. (If multiple sensors for a given measurement type are available, all are considered.) Data are again treated on an individual sensor basis and are assumed to be time-history independent. Again, it was assumed that all reasonable analysts would perform initial quality control of the data and this source of variation was not considered in the present work.

2. Sensor measurement combination

In the present work, measurements are combined across sensors and across towers to determine a single “observed” free-stream wind resource impinging on the plant as a whole, assuming spatial and temporal homogeneity across the plant. (That is, neglecting that for very large arrays the incoming wind resource may be distinct from one side of the array to the other and that there is temporal lag between the resource arriving at the leading edge of the plant and the trailing edge.) Furthermore, only hub-height measurements are considered in this variation of the analysis.

The first stage in characterizing this wind resource is therefore to determine the incoming direction, so that the upstream (unwaked) sensors may be identified and prioritized as the more accurate measurements for that time step. Historical reanalysis data is used to provide an initial first estimate for the wind direction at each time step and a simple geometric wake expansion model is used to determine which (if any) towers are anticipated to be upstream (unwaked) given that wind direction. (Note that, for computational efficiency, the waked/upstream condition for each tower was computed for each possible wind direction, in 1-degree increments. The wind direction at each time step was then rounded to the nearest 1-degree and the associated waked/upstream condition for that wind direction assigned to each tower for that time step.)

A “priority combination” method was then used to combine wind direction sensor measurements. In this algorithm,

at each time step, if there are no “priority” measurements available (i.e., no measurements from upstream sensors), then all available measurements for that time step are considered (i.e., upstream and waked conditions). The median and interquartile range of the measurements are then computed. If there is only one measurement or the interquartile range across the measurements is below a set threshold, the median value is accepted. Otherwise, the measurements outside some threshold of the interquartile range limits (by default taken to be in inner fence, as defined by the National Institute of Standards and Technology (NIST)/SEMATECH e-Handbook of Statistical Methods [12]) are removed from consideration and the median and interquartile range of the remaining measurements recomputed. If the new interquartile range across the measurements is below a set threshold, the new median value is accepted. Otherwise, the disagreement between the sensors is taken to be irreconcilable and, without additional information to indicate which sensor should be trusted, the value of the measurement for that time step is taken to be unknown (i.e., set to NaN).

Having applied this algorithm to the wind direction sensors, if the resulting estimate of wind direction is relatively complete, the classification of the towers as upstream/waked at each time step is recomputed and the wind direction estimate by combining across sensors is repeated. If the combined wind direction estimate is poor, the original estimate from the reanalysis data is retained. With either the reanalysis or refined estimate of upstream/waked conditions for each tower, the same priority combination method for combining the sensors is applied, still prioritizing upstream sensors, to determine a single estimate for the wind speed, turbulence intensity, air density, wind shear, and wind veer measurements.

A refinement of this baseline estimate has been developed using the upstream (unwaked) turbine nacelle-based measurements of wind speed and direction. However, the algorithms were developed using the z01 plant data and have not yet been generalized enough to allow robust application to the data from the other plants and represents an area of future work.

3. Resource sampling

Having obtained a single characterization for the wind resource experienced by the plant during the operational time frame, it then remains to provide this resource as input to the plant simulation. In the case of computationally cheap simulations, it may be possible to run a simulation of the plant production for each time step in the operational period and, indeed, this was done in the present case where simple engineering models were used for the simulations. However, for more computationally expensive simulations, it may be necessary to sample from the observed wind resource to attempt to capture the array performance over the operational period with fewer simulations.

In the present work, samples of the wind resource were developed by randomly selecting (with equal probability and without replacement) directly from the time series of the observed wind resource (thus ensuring that the conditional dependencies between different measurements of the wind resource were maintained.) Input sets of 2048, 4096, 8192, 16,384, 32,768, and 65,536 samples per set were constructed. In terms of the temporal annual coverage of the wind resource, these input set sizes represent (for 10-minute averaged data) – 3.9%, 7.8%, 15.6%, 31.2%, 62.3%, and 124.7%, respectively. Because each of the data sets examined contained approximately 3 years of data, multiple mutually exclusive (nonoverlapping) simulation input sets of each size may be constructed—at least 77, 38, 19, 9, 4, and 2 nonoverlapping sample sets for each of the input set sizes, respectively. The wind speed distributions as computed by different sample sets is given in Fig. 6. By comparing production predictions computed via different input sets of the same size (annual coverage), the variability due to the input sample may be evaluated for each level of annual coverage.

Fig. 6 Wind speed kernel density estimate (KDE) distributions as computed over different samples of the input resource (blue lines) and over the full operational period (red line) for increasing numbers of samples (by row) and for each project (by column).

4. Simulate plant performance

In this work, three engineering wake loss models were considered, as implemented within the FLOW Redirection and Induction in Steady State (FLORIS) software package [13]—the Jensen model, the Multizone model, and the Gauss model. The full time series of the observed wind resource over the operational period was used as inputs to each of the three models and the results partitioned into sets of different sizes *a posteriori*.

III. Results

The cumulative distribution function (*CDF*) of a turbine's production (*P*) taken over the operational period quantifies the probability that the turbine is producing less than or equal to a certain power level; therefore $CDF(P) = 0.5$ is the turbine's "expected" power production ($E(P)$) at a given time. In considering an ensemble of *CDFs* (e.g., as computed via different preprocessing/curve fitting/imputing methods for the predictable production estimate or via different simulation input sample sets/wake loss models for the simulated production estimate), the **input uncertainty** is defined as the difference between the worst and best *CDFs*. In comparing two *CDFs* of different derivations (e.g., the simulated production estimate against the predictable production estimate), the area validation metric quantifies the uncertainty between the two production estimates by comparing the median rank-based *CDFs* and not only their mean values. The **model uncertainty** is therefore defined as the absolute area between the best/worst *CDFs* for two production estimates and accounts for the possibility of getting the "right answer" for the "wrong reasons" (e.g., overestimating the probability of lower productions "compensating" for underestimating the probability of higher productions when considering the total production over some period.) For the purposes of either calibrating the parameters of or refining the physics behind a simulation model, it is critical to be obtaining the "right answer" for the "right reasons" to ensure that the model will be generalizable to other plants, especially those for which predictable production estimates are unknown (as is the case in generating preconstruction energy estimates). These classifications of uncertainty are illustrated in Fig. 7(a). In using the simulation model to predict the production of an arbitrary plant (e.g., in generating a preconstruction energy estimate), the uncertainty-quantified expression for the expected production would be:

$$E(P^{sim}) \in [E(P_{worst}^{sim}), E(P_{best}^{sim})] \pm (\text{model uncertainty})$$

For the case illustrated in Fig. 7(a), this expression evaluates to: $E(P^{sim}) \in [0.2850, 0.2985] \pm 0.0127 = [0.2723, 0.3112]$ for a total uncertainty of 0.0389 or 13% the median $E(P)$ as computed over the ensembles of simulations.

Within this framework, using the predictable production estimate as a benchmark against which to calibrate the model used to generate the simulated production estimate is equivalent to tuning the simulation model to reduce the model uncertainty—that is, to reduce the gaps between the two *CDF* estimates. However, it is important to realize that because the model uncertainty is defined as being the gap between the best/worst *CDFs* for the two production estimates, the simulation model calibration can only be completed to within the input uncertainties of each estimate. Therefore, the input uncertainties place lower bounds on the precision of the calibration.

In addition to examining the ability of the simulations to accurately and precisely model the production distributions, it may also be of interest to consider bias-correcting a single metric against the predicted production. In this case, the cumulative operational performance over the period of interest was computed, yet this method would also be applicable for the industry standard annual energy production (AEP) metric. Furthermore, it is emphasized that although here the simulations will still be used as the "preconstruction estimate," which is to be bias-corrected against the benchmark operational data, this method for uncertainty propagation in the bias correction of a metric would also be applicable to the fully developed preconstruction AEP estimation process (including resource modeling, plant production modeling, loss assessment, etc.), provided the input uncertainty for that estimate has been quantified (e.g., by considering ensembles of methods, as was done here for the simulations.)

To illustrate **metric uncertainty** propagation in the context of bias correction, the predictable operational performance was summed over the full operational time frame and normalized by the full nameplate capacity production over the same time period. The simulated performance over all simulations in a sample set was summed, divided by the temporal coverage of the sample set, and multiplied by the temporal duration of the full operational time period. This projected value for the production during the operational time period was then similarly normalized by the associated nameplate capacity. Bias correction of the simulated operational production (*OP*) to match the predicted *OP* may only be done to within the uncertainties associated with each value. These classifications of uncertainty are illustrated in Fig. 7(b). In using the simulation model to predict the *OP* of an arbitrary plant (e.g., in generating a preconstruction energy estimate),

Fig. 7 Illustrations of uncertainty propagation in different analysis contexts. (a) Example calculation of model uncertainty between predictable production estimate (for single input imputation method and power curve model; input uncertainty over preprocessing methods) and simulated production estimate (for a single wake loss model and sample set size; input uncertainty over different sample sets). (b) Example calculation of calibration uncertainty between predictable production estimate of OP (for single input imputation method and power curve model; input uncertainty over preprocessing methods) and simulated production estimate of OP (for a single wake loss model and sample set size; input uncertainty over different sample sets).

the uncertainty-quantified expression would be:

$$OP^{sim} = \overline{OP^{sim}} \pm [2 * \sigma(OP^{sim}) + \sigma(OP^{benchmark})]$$

Here taking as the uncertainty associated with the OP estimates as $2 * (\text{the standard deviation over the } OP \text{ estimates of the methods in the ensembles})$, although this is a somewhat less conservative estimate than the full spread of estimates, as used above for the model uncertainty. For the case illustrated in Fig. 7(b), assuming that the bias has been perfectly corrected such that $\overline{OP^{sim}} = \overline{OP^{benchmark}}$, this expression evaluates to: $OP^{sim} = 0.361 \pm [2 * (0.002) + (0.002)] = 0.361 \pm 0.006$ or a total uncertainty of 3.5% (as a percentage of the mean OP).

A. Input uncertainty: predictable production distribution and expected value

From the results presented in Fig. 8, it is clear that different operational performances are predicted for different turbines in a given array. This is by no means surprising, as plants are often located at a site with dominant wind directions and therefore some turbines are more often operating under waked conditions than other turbines in the plant.

Fig. 9 presents a visual summary of the input uncertainty for the $E(P)$ value (i.e., the spread between best and worst estimates) when considering increasingly more specific ensembles of methods for the predictable performance for an individual turbine. Clearly, the most variation occurs when different input imputation methods, power curve models, and preprocessing methods are all considered as composing an ensemble (top level in figure). This level of variation could be representative of the differences in predictable production estimates between companies. Assuming any individual company will select a single method for the input imputation (middle level), variation over just the power curve model and preprocesses still remains relatively high, although the “full_NTF” method tends to be less sensitive than the other methods. Assuming further that any individual company will select both a method for the input imputation and the power curve model (bottom level), the variation drastically decreases, although some power curve models show higher sensitivity to the preprocessing methods than others (with the IEC, spline, and ensemble extra trees regression showing the highest sensitivities.) This level of variation could be considered representative of the differences in predictable

Fig. 8 $CDF(P)$ for four individual turbines in each project (by column). Shaded region indicates maximum and minimum values over all methods in ensemble, with the solid line indicating the median. Square markers indicate median $E(P)$ with the maximum and minimum values (input uncertainty) indicated by the horizontal bar.

production estimates as produced by different analysts, ranging from those with a tendency to favor allowing more noise in the data to retain greater fractions of the data to those with a tendency to favor retaining smaller fractions of the data to reduce noise in the data.

If an individual company does internally standardize the power curve and input imputation methods, that single method variation will be applied to all turbines in a plant and indeed all plants. It is therefore reasonable to consider the distribution of input uncertainties across turbines in single plants and in all plants. (To emphasize—the input uncertainties here arise only from different preprocessing by the individual analyst.)

For this summary evaluation, a more lenient definition of input uncertainty was adopted—methods that produced a significant fraction of outlier estimates for $E(P)$ within the considered ensembles were removed from consideration. To accomplish this, first, the $E(P)$ estimates for each specific turbine/imputation method/power curve across preprocessing methods were compared and outlier preprocessing methods removed. Next, the $E(P)$ estimates for each turbine were compared across imputation method/power curve and outliers identified. Looking across all turbines in all arrays, the imputation method/power curve method variations that resulted in a significant fraction of the evaluations being outliers were also removed. For the remaining imputation method, power curve method variations, the median input uncertainty in $E(P)$ could be computed across all remaining preprocessing methods for each turbine and (after normalizing by the median $E(P)$) across all turbines within a specific array, as shown in Fig. 10 or even across all turbines in all arrays, as shown in Fig. 11. (Note that the x-axis labels for each bar in these figures indicates the total number of $E(P)$ evaluations included in each estimate; for example, if all 60 preprocessing methods had been valid for each of the 137 turbines in project h21, there would have been 8,220 evaluations available for each imputation method/power curve model. As indicated, this full ensemble was never achieved as between 30 and 557 preprocessing methods were removed within each imputation method/power curve model method variation. Similarly, if all 60 preprocessing methods had been valid for each of the 338 turbines in all projects, 20,280 evaluations of $E(P)$ would have been available for each imputation method/power curve model method variation.)

To summarize Fig. 10, across method variations, plant h21 has input uncertainties ranging from $<1\%$ to $>3\%$ with typical values around 1.75% , plant o46 has input uncertainties ranging from $<0.5\%$ to $>1.5\%$ with typical values around 0.75% , and plant z01 has input uncertainties ranging from $<4\%$ to $>8\%$ with typical values around 4% . The wide variation between plants is likely due to differences in both the quantity and quality of the data available for each project. For example, plant h21 has almost 4 years of data, while the other two plants have 3 years. However, the data for plant o46 showed very little noise even in the raw state, while the data for plant z01 had very high variation as there was a change in the operational protocols during the reported time frame. (It is noted that in a preliminary study, more complex preprocessing methods highly tailored to plant z01 were considered and even in this case, the input uncertainties remained on the order of 3% [14]) The high uncertainties should not be interpreted as indicative of the preprocessing methods being unsuitably or insufficiently developed for the given data sets, but rather that data with high variations simply present more opportunity for variations in the preprocessing methodologies as developed by individual analysts, with no one variation being more or less “correct” than another.

If a company is going to adopt a single standardized imputation method and power curve model to process operational data to then use as a benchmark for calibration of production prediction models, it would be advisable to select the

Fig. 9 Heat map showing input uncertainty in $E(P)$ (as percentage of median $E(P)$ as computed for each ensemble) for an individual turbine in each project (by row).

Fig. 10 Outlier-removed input uncertainties in $E(P)$, normalized by the median $E(P)$, as computed across preprocessing methods for each individual turbine/imputation method/power curve model. Bar height shows median value, black bars show 95% confidence interval, as computed by comparing across all turbines within an array. The x-axis labels indicate the total number of $E(P)$ evaluations included in the estimate.

Fig. 11 Outlier-removed input uncertainties in $E(P)$, normalized by the median $E(P)$, as computed across preprocessing methods for each individual turbine/imputation method/power curve model. Bar height shows median value, black bars show 95% confidence interval, as computed by comparing across all turbines within all arrays. The x-axis labels indicate the total number of $E(P)$ evaluations included in the estimate.

imputation method and power curve model that shows the least sensitivity to the choices of the individual analyst's preprocessing methods. As shown in Fig. 11, the linear 5 parameter and the asymptoted linear 5 parameter power curve models are the least sensitive to the input data preprocessing, while using the meteorological tower data transformed to nacelle-measurement equivalents provides the least sensitive input to the power curve model. The quality-control filtered, gap-filled input series also provides relatively low sensitivity, although it is noted that the cross-correlation gap-filling method was highly computationally intensive in comparison to the NTF gap-filling method.

B. Input uncertainty: simulated production distribution and expected value

Figures 12 and 13 show the $CDF(P)$ as determined by the different wake models and using different input sets of different sizes. As expected, when fewer samples are considered in a given set, the predicted output shows more dependency on the individual samples, thus yielding higher variability between the different sets of samples (i.e., higher input uncertainty). Fig. 14 shows this trend more clearly, plotting the input uncertainty as a percentage of the median $E(P)$ (as computed over the full time series for each wake model) as a function of the temporal coverage of the annual observed wind resource. From this figure, it is clear that, for these simple engineering models, to capture the yearly performance of the array to within 1% uncertainty, the input sample must have roughly 50% annual coverage of the wind resource.

If it is assumed that an individual company internally standardizes the wake velocity model and the percent annual coverage of the simulations, the distribution of input uncertainties across turbines in single plants and in all plants may again be evaluated.

To summarize Fig. 15, for plants h21 and z01 it is possible to achieve ~1% uncertainty if enough simulations are run to represent more than a full year of the wind resource, while ~2% uncertainty is achieved if only enough simulations are run to represent just over half a year of the wind resource. For plant o46, it is not possible to achieve <2% uncertainty even if more than a full year of the wind resource is simulated. This suggests a very high sensitivity to the exact wind resource conditions, possibly as a result of very close turbine spatial placements and therefore high sensitivity to wind direction for wake position calculations.

As shown in Fig. 16, across all plants the three velocity models showed very similar input uncertainties and, as would be expected, the input uncertainty will be reduced as the number of simulations run is increased, albeit with decreasing returns, suggesting that simulating just ~60% of a year's predicted wind resource could be sufficient.

C. Model uncertainty: predictable and simulated productions

To reiterate from the introduction to this section, the **model uncertainty** is defined as the absolute area between the best/worst CDFs for two production estimates and accounts for the possibility of getting the "right answer" for the "wrong reasons." For the present study, very simple engineering wake loss models were used in the simulations and the simulation parameters were not calibrated beyond the values reported in the literature or as reported by the turbine

Fig. 12 $CDF(P)$ for a turbine in plant h21. Shaded region indicates maximum and minimum values over all sample sets of the given size (by column), with the solid line indicating the median. Square markers indicate median $E(P)$ with the maximum and minimum values indicated by the horizontal bar. Purple indicates the Gauss model, blue indicates the Jensen model, and green indicates the Multizone model.

Fig. 13 Detail from Fig. 12, showing $E(p)$ more clearly, with value and input uncertainties as indicated.

Fig. 14 Input uncertainty as a percentage of the median $E(P)$ (as computed over the full time series for each wake model) for a randomly selected turbines for each project (by column). The red dashed line indicates the 1% uncertainty level. Purple indicates the Gauss model, blue indicates the Jensen model, and green indicates the Multizone model.

manufacturers in online documentation. Fig. 17 presents the model uncertainties as determined from the simulations in the “precalibration” state. Ideally, these model uncertainties could be reduced to zero with parameter calibration and model refinement, but it is by no means guaranteed.

As expected, the model uncertainty *increases* as the number of samples in each simulation set increases, because the input uncertainties (boundaries as defined by best/worst CDF) associated with the simulations decreases and therefore (assuming the median simulation CDF remains approximately the same) the gap between the best/worst CDFs for the two production estimates increases.

In the uncalibrated state presented here, it is notable that the “full_NTF” imputation method disagrees significantly more with the simulations than the other imputation methods. Examining this more closely, in Fig. 18, the simulated production CDFs for a single wake loss model and 124.7% annual resource coverage is compared against the predictable production CDFs for a single power curve model and each of the four different imputation methods. The “full_NTF” imputed input tends to predict more production in the intermediate range (roughly 0.5–0.75 nameplate capacity) than either the other imputation methods or the simulation methods. Although intuitively it is tempting to predict that the other imputation methods may allow more precise calibration of the simulation methods, because the distributions are already more similar, it is important to emphasize that any calibration of the simulation model will likely result in nonlinear changes in the final production distributions and therefore it is difficult to predict which predictable production curve would provide a benchmark against which the simulation model can reasonably be calibrated.

D. Total uncertainty: predictable and simulated productions

Again, in using the simulation model to predict the production of an arbitrary plant (e.g., in generating a preconstruction energy estimate), the uncertainty-quantified expression for the expected production would be:

$$E(P^{sim}) \in [E(P_{worst}^{sim}), E(P_{best}^{sim})] \pm (\text{model uncertainty})$$

For convenience, the input uncertainties in $E(P)$ are visually summarized in Fig. 19, while Fig. 17 presents the model uncertainties. Combining the input and mean uncertainties, the total uncertainties are presented in Fig. 20.

As shown in the figure, the lowest total uncertainties are achieved with the simulated production being computed using the Multizone model and >30% annual resource coverage and with the predictable production being computed using the “QCgap_crossCorrFilled” imputation method and spline or ensemble extra trees regression power curve model. Even in this best case, however, the uncertainty remains over 15%, emphasizing the crucial need to reduce the input uncertainties for both simulated and predictable productions and then calibrate the simulation model against the predictable production.

Fig. 15 Input uncertainties in E(P), normalized by the median E(P), as computed across sample sets for each individual turbine/velocity model/sample set size. Bar height shows median value, black bars show 95% confidence interval, as computed by comparing across all turbines within a plant. The x-axis labels indicate the total number of E(P) evaluations included in the estimate. The green dashed line indicates the 1% uncertainty level; the red dashed line indicates the 2% uncertainty level.

Fig. 16 Input uncertainties in $E(P)$, normalized by the median $E(P)$, as computed across sample sets for each individual turbine/velocity model/sample set size. Bar height shows median value, black bars show 95% confidence interval, as computed by comparing across all turbines within all plants. The x-axis labels indicate the total number of $E(P)$ evaluations included in the estimate.

E. Metric uncertainty: predictable and simulated operational production bias correction

Continuing to assume that a given company will standardize the imputation method and power curve model used in determining the benchmark predictable OP (input uncertainty is therefore computed over the ensemble of 60 preprocessing methods) and the number of simulations (percent annual wind resource coverage) and wake loss models in the simulated OP (input uncertainty is therefore computed over the ensemble of different samples of the wind resource), the total uncertainty in OP as a percentage of the mean (predictable) OP may be computed for each turbine. Taking the mean of this total uncertainty over all turbines in all plants yields an estimate of the metric uncertainty for each combination of predictable OP processing method and simulation OP estimation method, as shown in Fig. 21.

As shown in the figure, the lowest total uncertainties in OP are achieved with the higher number of simulations (greater annual resource coverage) with the Gauss wake model showing some benefit over the Jensen or Multizone wake models. For processing the benchmark OP , there is little difference between the any of the nacelle-based input series or any of the power curve models. Given that differences were observed in previous examinations of the production distributions, this result indicates the danger of single metric analysis as here, the “right” answer is being achieved for the “wrong” reasons and therefore hides the potential failure of these models to generalize to other plants/operational conditions.

More important than identifying more or less sensitive methods, however, is the magnitude of the total uncertainties—at best, the estimate can only be bias-corrected to the benchmark with an uncertainty of 2% of the mean benchmark value, given the input uncertainties in the benchmark and the estimation methods.

IV. Conclusion

The present work has examined the variation due to analyst choice introduced into the use of operational wind plant data as a benchmark against which to calibrate wake loss simulation models. Specifically, 60 methods for preprocessing the operational data, 8 power curve models, and 4 methods of imputing the input to the power curve models were considered for a total of 1920 estimates of the predictable power production by each turbine in each of three plants. Simulations taking as input samples of the observed wind resource were then run using 3 wake loss models in the simulations. The simulation input sample sets ranged in size from 4% to 125% of a full year of the observed wind resource.

If a single standardized imputation method and power curve model are used to process the operational data for all turbines in all plants, the uncertainty in the predictable expected production is estimated to be between 1% and 3% the median predictable expected production. Similarly, if a single standardized number of input samples and single wake model are used in the analogous simulations, the uncertainty in the simulated expected production remains at best on the order of 2%. These input uncertainties place lower bounds on the precision of the calibration between the simulation model estimates and operational benchmark data.

These estimates for input uncertainties due to analyst choices are arguably lower bounds. For example, variations in

Fig. 17 Heat map showing model uncertainty as computed for each combination of predictable processing method (by row) and simulation estimation method (by column), averaged over all turbines in all plants.

Fig. 18 Example CDFs for the predictable production estimates for different input imputation methods and a single power curve model (input uncertainty over preprocessing methods) and simulated production estimate (for a single wake loss model and sample set size; input uncertainty over different sample sets) for randomly selected turbines (by row) within each plant (by column).

Fig. 19 Heat maps showing input uncertainties (a) as computed for each combination of predictable processing method (by row) and (b) simulation estimation method (by column), averaged over all turbines in all plants.

the initial quality-control algorithms were not considered, different methods for deriving the “observed” wind resource for input to the simulations were not considered, and, perhaps most significantly, only wind speed was considered as an input to the power production model, while more advanced models taking as input wind direction, turbulence intensity, wind shear, wind veer, time history, operational status of neighboring turbines, or any combination of these data sources are not only possible but potentially very beneficial.

Although some level of standardization in benchmarking methods across the wind industry would certainly be beneficial (three important initiatives in this direction include the International Energy Commission International

Fig. 20 Heat map showing total uncertainty as computed for each combination of predictable processing method (by row) and simulation estimation method (by column), averaged over all turbines in all plants.

Fig. 21 Heat map showing total uncertainty in the *OP* metric as computed for each combination of predictable *OP* processing method (by row) and simulation *OP* estimation method (by column), averaged over all turbines in all plants.

Standard 61400 for wind energy generation systems – Part 12-1: Power performance measurements of electricity producing wind turbines [8]; the American Wind Energy Association Wind Standards Committee, TR-1 subcommittee on Wind Plant Power Performance [6]; and the U.S. Department of Energy, Atmosphere-to-Electrons-sponsored program Performance Risk, Uncertainty, and Finance [7]), complete standardization is not only infeasible, but undesirable. Given the inherent differences in wind power plant operational conditions and the unlimited potential for refinement of analysis methods, the role of an individual, experienced analyst will likely always remain important. Furthermore, as data sets become more complex, analysis algorithms become more refined, and computational costs decrease, adherence to a single standardized method would retard the ability of the industry to exploit the data to the fullest extent for increased accuracy and precision of models and methods.

To maintain confidence in the increasingly complex models and methods, however, it is imperative to quantify the uncertainty in an analysis due to the analyst choices. As demonstrated here, this can be accomplished by considering ensembles of analyses with reasonable variations in the choices made. Although the present work focused on developing ensembles of methodologies for the specific goal of using operational wind plant data to refine wake loss models, the framework of using ensembles of reasonable methods to quantify the uncertainty in an analysis is broadly generalizable not only to different analyses within the wind industry, but within the renewable energy sector more broadly.

Appendix

A. Operational data flagging algorithms

Please see Table 1 for the parameters used to define the preprocessing algorithms. Examples of the data flagged and the remaining cleaned data as determined by random selections of the preprocessing methods are given in Figs. 22 and 23, respectively.

Fig. 22 Data flagged by quality control and statistical preprocessing for a randomly selected turbine within each plant, with a random preprocessing method applied. The black hexbin plots indicate the raw data (with darker shades indicating more data) and the red ‘x’ markers indicate the flagged data. The percentages in the upper left indicate the data flagged as a percentage of the total available data for the turbine.

Table 1 Statistical flag parameters

Flag	Parameters	Set	First-pass value	Second-pass value
Wind speed binned power	Bin width	loosest	0.07	0.07
		base	0.045	0.045
		tightest	0.02	0.02
	σ threshold	loosest	4.0	3.0
		base	3.0	2.0
		tightest	2.0	1.0
	scalar threshold	loosest	0.2	0.18
		base	0.15	0.14
		tightest	0.1	0.08
Power binned wind speed	Bin width	loosest	0.05	0.05
		base	0.03	0.03
		tightest	0.015	0.015
	Bin minima	loosest	0.1	0.1
		base	0.025	0.025
		tightest	0.01	0.01
	Bin maxima	loosest	0.85	0.85
		base	0.9	0.9
		tightest	0.95	0.95
	σ threshold	loosest	4.0	3.0
		base	3.0	2.0
		tightest	2.0	1.0
	scalar threshold	loosest	0.2	0.175
		base	0.15	0.09
		tightest	0.1	0.075

Fig. 23 Data remaining after noise removal via quality control and statistical preprocessing, for a randomly selected turbine within each plant, with a random preprocessing method applied (note: not the same preprocessing methods as indicated in Fig. 22). The blue hexbin plots indicate the cleaned data (with darker shades indicating more data). The percentages in the upper left indicate the data remaining as a percentage of the total available data for the turbine.

B. Power curve models

1. International Electrotechnical Commission

Inspired by the International Electrotechnical Commission (IEC) standard 61400-12-1 for power curve reporting (IEC 61400-12-1), a wind speed-binned step function was implemented. In this method, the power data was binned by wind speed (bin width of 0.045) and the mean power within each bin was computed. At evaluation of this power “curve” an input wind speed was matched to the appropriate bin and the mean power value for that bin was reported as the predicted production for the input wind speed.

2. Logistics five parameter

The logistics five parameter function as a model for wind speed-power curves was proposed and characterized by Lydia et al. [9] and compared to other models by Lydia et al. [10]. The function is described by the following:

$$P(u) = d + \frac{a - d}{\left(1 + \left(\frac{u}{c}\right)^b\right)^g}$$

P : the predicted power

u : the wind speed

a : the minimum asymptote

b : Hill's slope; effectively the steepness of the curve

c : the inflection point of the curve

d : the maximum asymptote

g : asymmetry factor (around the inflection point)

In keeping with the results reported by Lydia et al. [9], a differential evolution algorithm was employed to find the optimal values for the parameters a , b , c , d , and g .

3. Generalized additive (spline) model

A generalized additive model is described by the following:

$$P(\mathbf{x}) = \beta_0 + f_1(x_1) + f_2(x_2) + \dots + f_p(x_p)$$

P : the predicted power

$\mathbf{x}$: a vector of input measurements

$f_i()$: feature functions composed of n penalized B splines

The implementation of the algorithm provided in the Python package “pyGAM” was employed in the present work [15]. It is noted that by using only one input parameter (wind speed in this case), the GAM is equivalent to a spline fit model. The default number of splines for the fitting was set to 35.

4. Ensemble extra trees regression

To explore the potential of a more advanced machine-learning model, a decision-tree model was selected after comparison with several other machine learned models, such as support vector machines, Gaussian process, and logistic regression. The choice of a decision tree-based algorithm was motivated by the need to choose a highly efficient model for training on a very large data set, a robustness to noise in both the input and the training data, and, looking forward to the integration of wind direction and other wind resource characteristics as input features for power curve prediction, by the convenience of using an algorithm robust to non-normally distributed input. To ensure the model was not overfitting the data, the ensemble extra trees regression model was selected (although it is noted that other possible choices include a random forest regressor with bootstrapping, bagging regressors, and AdaBoost.)

Details of the algorithm may be found on the Sci-kit learn documentation page for “`sklearn.ensemble.ExtraTreesRegressor`.” During development, a grid search with cross-validation was used to optimize the algorithm's hyperparameters, selecting the number of estimators to be 1000, all samples being using for each tree, and the minimum number of samples required to split an internal node to be 20% of the total number of samples.

C. Imputed wind resource distributions

The distributions of the full time-series inputs as computed by the four input imputation methods are compared in Fig. 24.

Fig. 24 The kernel density estimate (KDE) distributions of the full time-series of inputs to the power curve models as computed by the four imputation methods are compared for a randomly selected turbine within each plant, with a random preprocessing method applied (for the “PPgap_NTFfilled” method).

Funding Sources

This work was funded by the U.S. Department of Energy within the Atmosphere-to-Electrons research program and by a National Renewable Energy Laboratory Director’s Postdoctoral Fellowship awarded to Anna Craig through the Laboratory Directed Research and Development program. This work was authored by the National Renewable Energy Laboratory, operated by the 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.

Acknowledgments

The author would like to thank the other members of the Performance Risk, Uncertainty, and Finance team at the National Renewable Energy Laboratory, including M. Jason Fields, John Meissner, Mike Optis, Shuangwen Sheng, Jordan Perr-Sauer, Caleb Phillips, and Monte Lunacek. The author would also like to thank Patrick Moriarty for his technical and editorial reviews of the present manuscript.

References

- [1] Tocco, M., “Energy Performance Benchmarking,” *AWEA Wind Resource and Project Energy Assessment Conference*, Natural Power, 2017.
- [2] Bullard, M., “Repowering wind projects,” *AWEA Wind Resource and Project Energy Assessment Conference*, AWS Truepower, 2017.
- [3] Perry, A., “Cross Validation of Operational Energy Assessments,” *AWEA Wind Resource and Project Energy Assessment Conference*, Vaisala, 2017.
- [4] Johansen, N., “High Resolution Wind Farm Energetic Availability,” *AWEA Wind Resource and Project Energy Assessment Conference*, Avangrid Renewables, 2017.
- [5] Parkhe, V., “Can Data Science Be Used To Predict Turbine Performance?” *AWEA Wind Resource and Project Energy Assessment Conference*, DNV-GL, 2017.
- [6] AWEA Standards Development, Wind Standards Committee, TR-1 subcommittee (on-going). URL <https://www.awea.org/standards-development>, (Group chair: Brian Kramak, AWS Truepower).

- [7] Atmosphere to Electrons: Performance, Uncertainty, Risk, and Finance (PRUF), (on-going). URL <https://a2e.energy.gov/about/pruf>, (Principle investigator: Michael Jason Fields, NREL).
- [8] Commission, I. E., “International Standard: Wind energy generation systems – Part 12-1: Power performance measurements of electricity producing wind turbines,” Tech. Rep. IEC 61400-12-1, 2017.
- [9] Lydia, M., Selvakumar, A. I., Kumar, S. S., and Kumar, G. E. P., “Advanced Algorithms for Wind Turbine Power Curve Modeling,” *IEEE Transactions on Sustainable Energy*, Vol. 4, No. 3, 2013, pp. 827–835. doi:10.1109/TSTE.2013.2247641.
- [10] Lydia, M., Kumar, S. S., Selvakumar, A. I., and Kumar, G. E. P., “A comprehensive review on wind turbine power curve modeling techniques,” *Renewable and Sustainable Energy Reviews*, Vol. 30, 2014, pp. 452–460.
- [11] Power Curve Working Group, (on-going). URL <http://www.pcwg.org/>, (Group chair: Taylor Geer DNV-GL).
- [12] NIST e-Handbook of Statistical Methods, N., 2018 URL <https://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm>.
- [13] Annoni, J., Fleming, P., Scholbrock, A., Roadman, J., Dana, S., Adcock, C., Porte-Agel, F., Raach, S., Haizmann, F., and Schlipf, D., “Analysis of Control-Oriented Wake Modeling Tools Using Lidar Field Results,” *Wind Energy Science Discussions*, 2018, pp. 1–17.
- [14] Craig, A., Optis, M., Fields, M. J., and Moriarty, P., “Uncertainty quantification in the analyses of operational wind power plant performance,” *The Science of Making Torque from Wind*, edited by European Association of Wind Energy, 2018.
- [15] Serven, D., “pyGAM,” GitHub repository, January 2018. URL <https://github.com/dswah/pyGAM>.