IOP Conf. Series: Journal of Physics: Conf. Series **1037** (2018) 052021 doi:10.1088/1742-6596/1037/5/052021

Uncertainty quantification in the analyses of operational wind power plant performance

Anna Craig[1], **Mike Optis, Michael Jason Fields, Patrick Moriarty**
National Renewable Energy Laboratory, Golden, CO, USA
E-mail: [1]anna.craig@nrel.gov

Abstract. In the present work, we examine the variation introduced in the evaluation of an operating plant's wind power production as a result of the choices analysts make in the processing of the operational data. For this study, an idealized power production for individual turbines over an operational period was predicted by fitting power curves to the turbine production data collected during expected operation (that is, without curtailment or availability losses). A set of 240 possible methods were developed for (a) defining what data represented expected operation and (b) modeling the power curve. The spread in the idealized power production as predicted by the different methods was on average almost 3% for the 100 turbines considered. Such significant variation places a lower bound on the precision with which analysts may employ such data as benchmarks for calibration of their energy estimation processes and limits the potential for identification of refinements to the energy estimation models for improved accuracy.

1. Introduction

Data collected during the operation of wind plants are used in many diverse contexts, such as identification of 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. In the present work, we begin to build an understanding of how variations in an analyst-chosen data handling methodology can influence an analysis result. The operational analysis process was conceptualized as having four main stages: (1) data collection, (2) data preprocessing, (3) data processing, and (4) data analysis. A generalized scheme of these stages, possible analysis components within the stages, and the flows of data between the analysis components are presented in Figure 1.

Within this generalized schema, it is important to recognize that different analysts will focus on different data and employ different methods depending on their goals. For example, if the analysis goal is to validate and calibrate a preconstruction annual energy production model, concentrating on the plant-level metered production data and historical meteorological data may be appropriate. Conversely, if the goal is to identify underperforming turbines, analysis of

the individual turbine supervisory control and data acquisition (SCADA) and status code data may be a more pertinent approach.

The analysis we chose for the present work was completed at the level of individual turbines performance to predict an "ideal" power production (that is, with the turbine operating at all times without curtailment or availability losses) over the operational period. Following the collection of data, the goal of the data preprocessing stage was to remove erroneous measurements or measurements recorded during times of "abnormal" operational conditions, such that only operational data representing the "expected" performance of the turbine remain. In the processing stage, a model was then fit to this filtered power production data, using the measured nacelle wind speeds as the input. The model was then evaluated using a (turbine-specific) time series of wind-speed inputs to yield the ideal power production over the full operational time period. Section 2 describes two key preprocessing and processing choices: (1) how to remove data representing abnormal operation so that the power curve model is fit to only expected operational data and (2) what power curve model to use. A full operational evaluation was performed following each possible choice, for a total of 240 possible methods. The results of the analysis are presented in Section 3, with a discussion of the implications in Section 4.

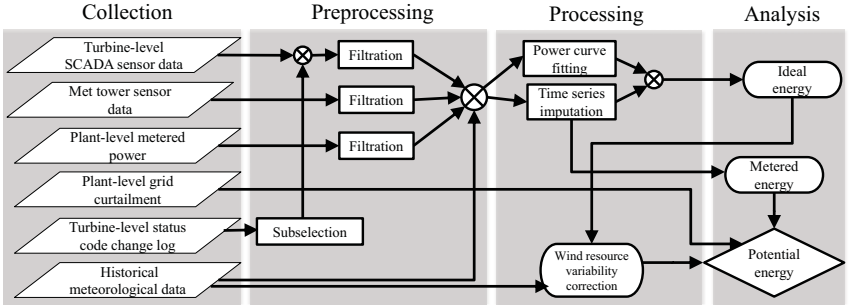

Figure 1. Generalized flow for operational analyses, indicating four main stages in the process and a selection of potential analysis components and possible data flow between analysis components.

2. Methods

2.1. Data collection

The data used for the present analysis include SCADA measurements from 100 turbines in an operational wind power plant and environmental measurements from four permanent meteorological towers, both reported as averages over 10-minute intervals. These data are supplemented by time-stamped status code changes for each turbine and hourly external historical meteorological data (from the National Aeronautics and Space Administration Modern Era-Retrospective Analysis for Research and Applications, MERRA-2, data set [6].) A 3-year operational period was considered. For reasons of confidentiality, details concerning the data set (e.g., turbine rated capacity, time period of operational data) have been omitted and results are presented on normalized bases.

2.2. Data preprocessing

As shown in Figure 1, the data preprocessing stage involved flagging and filtration of data that did not represent normal turbine operating behavior. For this process, we developed flags based on both the status code event logs and the sensor-based data flagging, in the latter case developing flags at both the level of basic quality control of the data and at the level of statistics-based outlier detection.

The status code change log provided a per-turbine record of the times at which a certain status code change occurred, with a code either becoming active or inactive. The codes represent a detailed accounting of turbine operational status encompassing many aspects of the physical, mechanical, and electrical condition of the turbine, with multiple codes typically being active

simultaneously. In order for a data point to be considered representative of expected operation, it must have occurred during a time frame when all "normal operation" status codes were active and all selected "abnormal operation" status codes were inactive. Note that the functional meaning, triggering, and recording of such codes is specific not only to the turbine manufacturer, but also to the turbine model and potentially even to the SCADA software version under use at a given time. As the number of codes can be on the order of several hundred, the task of evaluating which codes are important for including expected operational data and excluding abnormal operational data is an individual choice for the analyst.

To determine the variation due to status code use, we identified five possible sets of status codes as representing a range of possible analyst choices plus the extreme case of no status codes being applied because the status code change log was unavailable as a data source. We assumed that if status codes were available, any analyst would correctly identify all "normal operation" codes. Beyond that, however, codes indicating an abnormality in operational parameters may or may not be considered by an analyst, depending on whether the analyst was more focused on retaining only a very clean data set or on retaining as much data as possible. To model this, we identified four "abnormal operation" status codes explicitly representing reduced power performance (as a result of curtailment, pitch stall, or temperature limits). These codes were incrementally added to the set of flags applied during the data filtration: the two associated with curtailment we deemed to be the most obvious and therefore most likely applied, followed by the pitch stall code, and finally the temperature limit code, which we deemed to be relatively nonobvious. The resulting sets of status codes are summarized in Table 1.

Table 1. Sets of status codes used in flagging data for abnormal vs. expected operation.

Set	Normal operation status codes	Abnormal operation status codes
0	*None*	*None*
1	'load status', 'system status', 'error status'	*None*
2	'load status', 'system status', 'error status'	'external curtailment'
3	'load status', 'system status', 'error status'	'external curtailment', 'internal curtailment'
4	'load status', 'system status','error status'	'external curtailment', 'internal curtailment', 'pitch stall'
5	'load status', 'system status', 'error status'	'external curtailment','internal curtailment', 'pitch stall','temp limit'

Sensor-measurement-based data flagging may also be performed given a physical understanding of the system; that is, expectations for system operation based on prior knowledge (e.g., the classical wind-speed power relation). Similar to the status code data, analyst choice can play a key role in determining algorithms and parameters to flag abnormal operational data, although the algorithms would likely be generalizable to operational data collected across turbine manufacturers, models, and SCADA software version. Looking forward, this could potentially allow for a more standardized analysis across power plants, especially with regard to more detailed classification of abnormal operational data. Examples of the choices faced by the analyst include which measurement channels to consider (e.g., incorporating external measurements from neighboring turbines, proximate meteorological towers), whether or not to assume time dependence of the measurements (e.g., accounting for sensor drift, recalibration events, mechanical degradation), what flags to apply to each measurement channel considered and with what parameters, and what order to apply the flags in, to identify just a few.

To make the problem tractable, for this work we developed a method that is believed to be representative of the analysis complexity currently used in industry [7, 8] while still remaining relatively basic, thus reducing the number of analyst choices required. The measurement channels considered were the output power, nacelle wind speed, and the pitch for one of the blades. For this preprocessing stage, only data from the individual turbine were considered and it was assumed that all measurements were independent of time history. Therefore, we focused on the development and application of different flagging algorithms, with the variations between methods arising from variation in which flags were applied for data filtration.

The first sensor-measurement-based flags were quality control on the data. These flags entailed removing out-of-range measurements (e.g., negative or unrealistically high wind speeds or power output below a certain threshold) or periods of sensor irresponsiveness (i.e., extended time intervals for which the measurement did not change as much as would be expected under normal operation; e.g., as when an anemometer was iced over or otherwise stuck). The next level of complexity in sensor-measurement-based, quality-control flagging employed an understanding of how the system should physically behave. The first such flag focused on blade pitch: under expected operation, the blade pitch will lie within a small range around 0 degrees over wind speeds up to the threshold when the turbine begins generating at full capacity, after which the blade pitch is adjusted to regulate power. The constructed filter therefore removed data for which the blade pitch was too high over a window of wind speeds up to the maximum generation wind speed. Similarly, considering the wind speed and power data, if the power was below some threshold at wind speeds above what was required for maximum generation, the data were removed as likely representing a condition of curtailment and therefore not to be considered for modeling expected operation.

For the present analysis of variation introduced by the analyst and overall method, we assumed that all reasonable analysts would perform some sort of quality control of the data, yielding at the end a similar set of flagged data. Variation due to choice of algorithm parameters employed within each quality control flagging algorithm were considered in a preliminary analysis using Monte Carlo resampling for a single overall method. Although the variation was found to be non-negligible, the computational cost of employing a Monte Carlo resampling for each of the overall methods considered here was prohibitive. It remains an area of continuing study to employ more computationally efficient algorithms to include this source of analysis variation.

Depending on the experience of the analyst, they may perform further sensor-measurement-based data flagging based on the statistics of the measurements themselves. It is important to note that while each of the status-code and quality-control flags could be determined from the full, unfiltered data set, data that were previously flagged by the status-code and quality control flags should be removed from consideration prior to computing the statistics-based flagging algorithms so that the statistics are not skewed by the inclusion of outliers. For the present work, we developed three statistics-based flagging algorithms for the wind-speed power relation: (1) deviation of the power within wind-speed bins [7, 8, 9], (2) deviation of the wind-speed within power bins [7, 8], and (3) Mahalanobis deviation within wind-speed power K-means clusters [10, 7, 8]. To simplify the analysis, we applied each flag separately on the data (i.e., they were not applied in an ordered cascade, with the data flagged by one removed before consideration by the next.) To determine the variation due to analyst choice in statistical-based flagging, the eight possible subsets of the three statistical filters were each considered, as enumerated in Table 2. Each of the eight statistical flags sets were computed for each of the six status code sets, yielding 48 possible variations introduced by the analyst in the preprocessing stage of the data.

2.3. Data processing

In the processing stage of the operational analysis, a model for this expected power production was then fit to the filtered data remaining after the preprocessing method, taking as input

Table 2. Sets of statistics flags used in flagging data for abnormal vs. expected performance

Set	Flags included	Set	Flags included
0	*None*	4	Wind-speed-binned power, power-binned wind speed
1	Wind-speed-binned power	5	Wind-speed-binned power, clustered Mahalanobis
2	Power-binned wind speed	6	Power-binned wind-speed, clustered Mahalanobis
3	Clustered Mahalanobis	7	Wind-speed-binned power, power-binned wind-speed, clustered Mahalanobis

concurrent measurements characterizing the wind resource; in this case, the nacelle-measured wind speed. For the present analysis, we have considered five models: a logistics five-parameter (L5P) function, a linear-segmented (LS) model, a wind-speed binned averaged model (as described by the International Electrotechnical Commission standard 61400-12-1 and hereafter denoted "IEC"), a nonparametric spline (n-spline) model, and an ensemble extra trees regression (ETR) model [11, 12, 9, 7, 8, 13]. Fitting all power curve models to each of the 48 possible sets of fully flagged data yielded 240 possible variations in the operational analysis.

2.4. Data analysis

The "expected" performance-fit power curves, $g_t(\mathbf{x})$, were then evaluated using a standard set of input wind speeds (spanning the operational period), $\mathbf{x}'_t$, to determine an "operational idealized power" production for each turbine over the full operational period: $OIP_t = \sum_{\substack{operational \\ period}} g_t(\mathbf{x}'_t)$. Again, for confidentiality, this value is expressed normalized by the median value as computed across all 240 evaluations: $OIP^*_t = \frac{OIP_t}{Md(OIP_t)}$.

For an individual turbine, the spread in the OIP^*_t, defined to be the maximum difference between the 240 OIP^*_t evaluations, represents the "worst-case scenario" of the largest disagreement between two analysts. The interquartile range of the OIP^*_t is defined to be the center 50% of OIP^*_t results and may be interpreted as the typical disagreement between two analysts for that turbine.

Aggregating these metrics across all considered turbines, the median OIP^*_t spread is therefore an indicator of how much disagreement we would expect as being possible between the highest and lowest power estimates for any given turbine, whereas the median OIP^*_t interquartile range is how much disagreement we have to expect between different power estimates for any given turbine. The median OIP^*_t spread in the array is therefore a relatively conservative estimate of the uncertainty in the OIP^*_t of a turbine, whereas the median OIP^*_t interquartile range is a lower bound to the uncertainty in the OIP^*_t of a turbine.

3. Results

3.1. Data preprocessing and processing

Presented in Figure 2 are examples for two turbines of wind-speed power data remaining after data preprocessing flags were applied and the resulting power curve fits. Figure 2(a, d) represent the minimum removal of data considered: status codes were not used (or unavailable) and no statistical flags were employed; thus only quality-control flagging of the data was completed. (As the removal of data in even this minimal case was already quite high, it is here noted that the flag for a minimum power production of 10 kW was alone sufficient to remove around 18% of the total data for a turbine.) Despite the high percentage of data removed, there is clearly persistence of data with low power at high wind speeds, representing unexpected operational

performance. Figure 2(b, e) represent the most highly filtered data set considered: all identified status codes were applied as flags and all statistical flags were applied. Although the application of this method to turbine A yielded a reasonable set of data, application of this method to turbine B yielded excessive data removal. Finally, Figure 2(c, f) show the raw, unfiltered data, overlaid with the five power curve fits made using each of the 48 filtering methods considered (240 curves total).

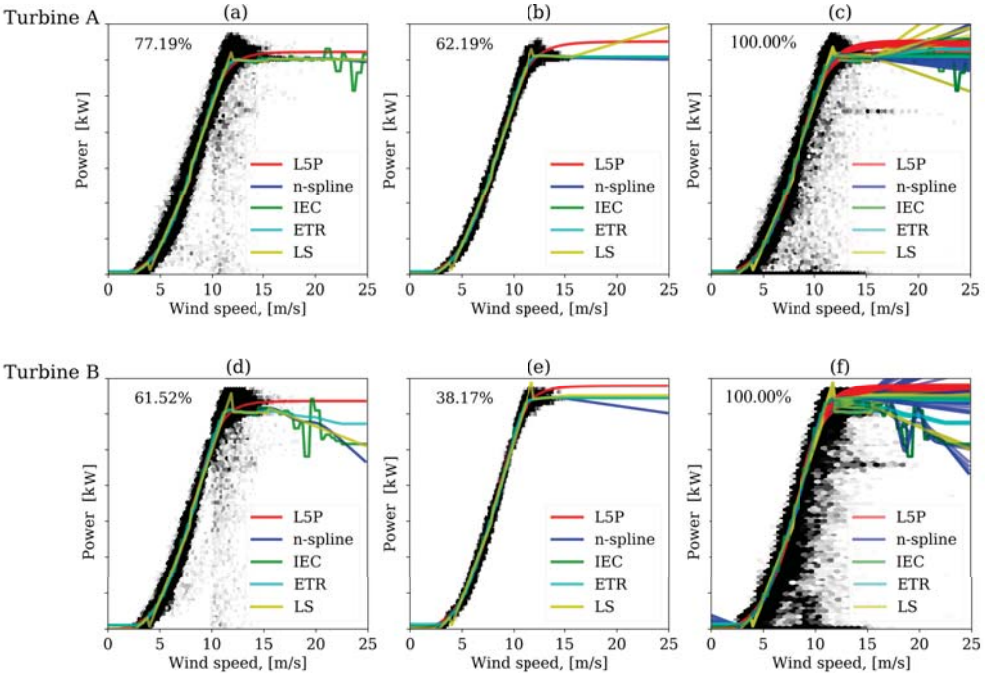

Figure 2. Examples for single turbines of (a,b) wind-speed power data remaining after data preprocessing flags were applied (numerically reported as a percentage in upper left of each subplot) and the resulting power curve fits and of (c) unfiltered wind-speed power data with all associated power curves. Wind speed and power data are shown in hexbin format, such that the intensity of the color map provides an indication of the number of data points.

It is clear that the power curve models were unreliable in reasonably capturing the expected behavior at higher wind speeds. This unreliability was likely a result of a relative sparsity of data in comparison to the data retained at lower wind speeds. The high-wind-speed data sparsity may be attributed to a compound effect: first, for this data set there was an initial uneven distribution of data over operating conditions, such that the data at the high wind speeds were already relatively sparse. Second, the statistical-based flagging algorithms did not take into account the relative densities of available data, thus flagging the expected operational data at high wind speeds as outliers. Even the L5P model, while allowing enforced asymptotic behavior at the high wind-speeds, also suffered from this high-wind-speed data sparsity by biasing the fit toward the over-rated-capacity generation that occurred near wind-speeds just at the threshold of full capacity generation and could represent a greater proportion of the data when the high-wind-speed data were too sparse. The net result was a power asymptote for high wind speeds that was higher than what is physically expected. A data set that contains more high-wind-speed production data (arising from longer duration data sets, more optimized plant control strategies, and so on) could potentially alleviate many of these concerns.

This sensitivity of power curve models to data sparsity in the high-wind-speed regime was

not revealed by examining the power curve fits for random selections of turbines, as we have done in developing these methodologies. From a methodological standpoint, therefore, we suggest developing more advanced flagging methods to account for uneven data distribution over operating conditions and avoid flagging sparse data points as outliers, developing more robust power curve models, or manually examining the power curve model for each individual turbine for reasonability. For the purpose of the present analysis, we emphasize that the variations reported here were meant to capture the variation between analyses, including the potential for analyst oversight, such as choosing a power curve model that does not appropriately fit the data for some subselection of the turbines. The development of additional, more robust flagging methods and power curve models is currently ongoing to more thoroughly characterize the extent of this variation between analysts.

3.2. Data analysis

For the two individual example turbines, the distributions of the OIP_t^* values from the 240 analyses are presented in Figure 3(a). For turbine A, the spread of the distribution was 2.3% (of the median OIP_t), whereas the interquartile range was 0.4%; for turbine B, the spread was 5.5%, whereas the interquartile range was 1.2%. The distributions of the spread and interquartile ranges over 99 turbines are shown in Figure 3(b) and (c), respectively. Over the full 100 turbines, the median spread of the OIP_t^* was 2.9% and the median interquartile range was 0.7%, whereas the maximum observed spread was 8.6% and the maximum interquartile range was 2.3%. It is noted that turbine A was selected to have statistics close to the median values for the array, whereas turbine B was selected to have one of the most extreme cases. The most extreme case, on examination, showed a spike in the wind-speed power data, such that for a very limited range of wind speeds, a large range of powers were reported. We believe that this represents a failure of the wind-speed sensor, which may or may not be caught by an analyst. When removed from consideration, the median spread and interquartile range of the OIP_t^* over the remaining 99 turbines remained 2.9% and 0.7%, respectively, while the maximum observed spread dropped to 5.5% and the maximum interquartile range dropped to 1.6%.

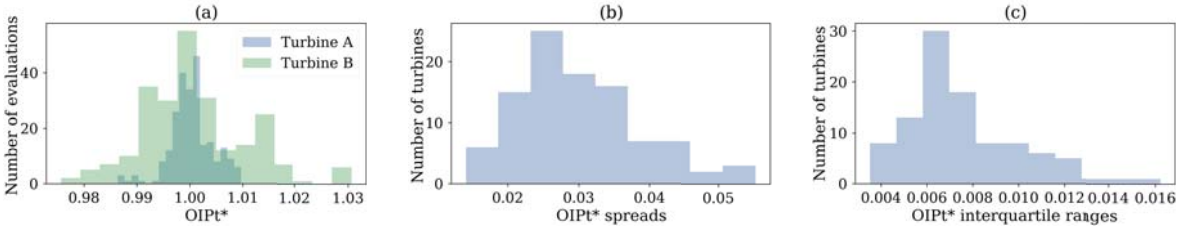

Figure 3. Distributions of (a) the OIP_t^* values from the 240 analyses for two individual turbines, (b) the OIP_t^* spreads across the 240 analyses for each of the 99 turbines, (c) the OIP_t^* interquartile ranges across the 240 analyses for each of the 99 turbines

Figure 4 shows the OIP_t^* for two example turbines within each power curve model, split by the status code set applied. Therefore, spread between bars within each column represents the variation due to the status code set choice of the analyst and the vertical extent of each bar represents the spread due to the statistics flag set choice of the analyst. There was a trend that the increasing number of status code flags applied increased the estimate of OIP_t^*; this is not surprising as the additional status codes applied would be expected to remove data for which the turbine was performing suboptimally. Furthermore, with an increasing number of status code flags applied (bars from left to right within each column), the sensitivity to which statistical algorithm flags were applied decreased (as indicated by the decreasing vertical extent of the

bars). This decrease in sensitivity is in keeping with the expectation that as additional status codes increasingly removed abnormal operational data, not only were those outliers that the statistical algorithms did not need to identify, but the statistics of the remaining data became better defined and flagging based on these statistics became more precise.

Figure 5 shows the same data, but in this case split by the status code set applied, such that the spread between bars within each column represents the variation due to the statistics algorithm set choice of the analyst and the vertical extent of each bar represents the spread due to the status code set choice of the analyst. Again, with the addition of more flagging algorithms (i.e., more stringent removal of data, color group gray to red to green to blue) there was the same trend that the estimate of OIP_t^* increased. More notably, however, is that while each statistical algorithm was based on the same wind-speed power relationship, the specific algorithms chosen significantly impacted the result. For the three algorithms considered here, the use of the clustered Mahalanobis algorithm yielded reduced estimates (set 3 compared to sets 1 and 2 in the red group and sets 5 and 6 compared to set 4 in the green group). The lack of using any statistical flags, however, introduced the most extreme estimates on the low side, indicating the importance of employing some level of data flagging beyond simply status code and quality-control flagging to remove the operational data representing below-expected performance.

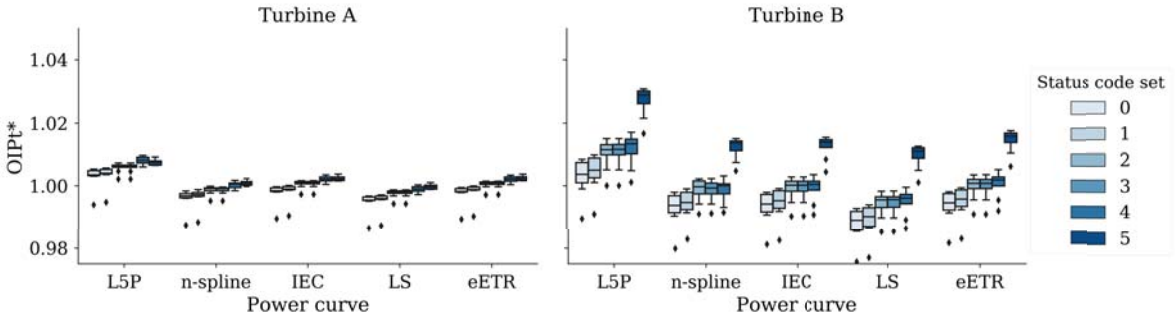

Figure 4. Characterization of OIP_t^* variation for two individual turbines. For each turbine, each column represents a power curve model, the bars represent the status code set, and the box-and-whisker statistics represent the spread over the statistical flag sets.

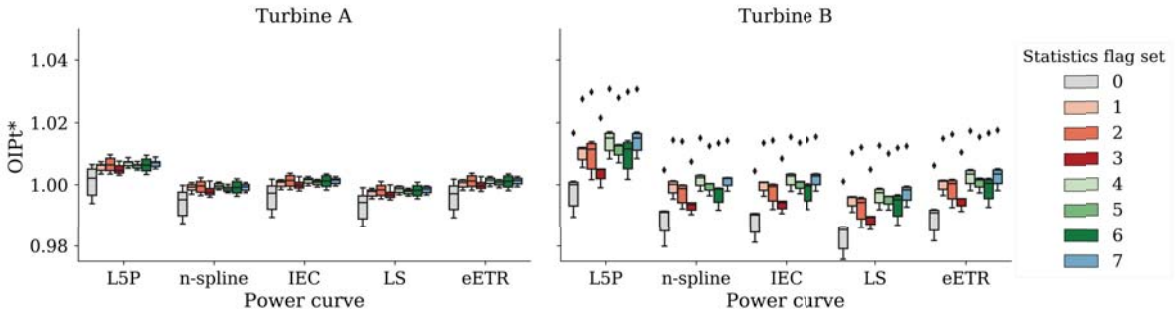

Figure 5. Characterization of OIP_t^* variation for two individual turbines. For each turbine, each column represents a power curve model, the bars represent the statistics flag set, and the box-and-whisker statistics represent the spread over the status code sets. The color group of the bars indicates the number of statistics flags in a set: grey is none, red is one, green is two, and blue is three.

At present, we believe that the increased sensitivity of turbine B to the analysis method, in comparison to turbine A, may be attributed to relatively more data being flagged and filtered, which in turn leads to less reliable power curve fits (as previously discussed). Returning to Figure 2, the relative fraction of data retained for fitting turbine B could be as little as 61% of the relative fraction of data retained for fitting turbine A in the same processing method. The physical reasons underlying the more extreme flagging of data (e.g., location of the turbine within the array, preferential curtailment of specific turbines, mechanical failures) remains under investigation. Understanding these underlying physical reasons will assist in the development of more robust flagging methods and power curve models.

4. Discussion and future work

In the present work, we examined the variation introduced in the evaluation of operational performance as a result of analyst choices in data flagging and power-curve fitting methods. Across a set of 240 possible methods, each representing what might be reasonably developed by an experienced and careful analyst, the typical spread of the estimated OIP_t^* for an individual turbine was 2.9% of OIP_t^* for that turbine. In terms of using operational data as a benchmark for production estimation models (e.g., wake loss models for individual turbines within an array), such significant variation places a lower bound on the precision with which such estimation models may be calibrated against and limits the potential for identifying model refinements for improved accuracy.

Yet, this estimate for variability is itself a lower bound on the variation in the overall data analysis method: only five possible power curve models were considered; variations in the parameters used for the quality-control sensor-measurement-based flagging algorithms were not included; and possible variations in the methods used to fill in missing or flagged data required as input to the power curve models. Furthermore, although these three potential sources of variation are the subject of the immediate ongoing work, it is again emphasized that many more variations in methodology remain unexplored: which measurement channels to consider, whether or not to assume time dependence of data, whether or not to consider the operational status of neighboring turbines, whether or not to incorporate additional external data sources, and so on.

Even within the limited method variations considered in the present work, however, the variation is non-negligible and by comparing across different methods, we have been able to identify several potential methodological sensitivities, which, when considering each method in isolation, had remained hidden. At present, we believe that much of the variation resulted from unreliable fitting of the power curve model, which may have been contributed to, in part, by extreme flagging of data; understanding the underlying physical reasons for the differences in flagging sensitivity will help guide the development of more robust flagging algorithms and power curve models.

From a methodological standpoint, having identified the flagging and power curve models as potential sources of significant variation has served to guide the direction for future work in improving the methods for increased robustness. More broadly, from the standpoint of understanding the variations between methods, the inclusion of additional, improved methods will allow for a more thorough characterization of the variations that can be introduced between different methodologies.

Although some level of standardization in operational data processing methods across the wind industry would certainly be beneficial (three important initiatives in this direction include the International Energy Commission International Standard 61400 for wind energy generation systems – Part 12-1: Power performance measurements of electricity producing wind turbines [9], the American Wind Energy Association Wind Standards Committee, TR-1 subcommittee on Wind Plant Power Performance [7], and the U.S. Department of Energy, Atmosphere-

to-Electrons-sponsored program Performance Risk, Uncertainty, and Finance [8]), complete standardization will likely be infeasible. 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. This work suggests, however, that it may be beneficial for multiple methods to be considered in an ensemble manner, as was done here. Not only can an ensemble approach illuminate potential shortcomings in the individual methods, it allows some characterization of the uncertainty introduced by the analysis methodology.

Acknowledgments

The authors would like to thank the other members of the Performance Risk, Uncertainty, and Finance team, including Shuangwen Sheng, John Meissner, Jordan Perr-Sauer, Caleb Phillips, Shreyas Ananthan, and Travis Kemper. The authors would also like to thank Katherine Dykes and Garrett Barter for their technical and editorial reviews of the present manuscript.

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. The Alliance for Sustainable Energy, LLC (Alliance) is the manager and operator of the National Renewable Energy Laboratory (NREL). NREL is a national laboratory of the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy. This work was authored by the Alliance and supported by the U. S. Department of Energy under Contract No. DE-AC36-08GO28308. Funding was provided by the U.S. Department of Wind Energy Technologies Office. The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy 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.

References

[1] Tocco M 2017 Energy performance benchmarking *AWEA Wind Resource and Project Energy Assessment Conference* (Natural Power)
[2] Bullard M 2017 Repowering wind projects *AWEA Wind Resource and Project Energy Assessment Conference* (AWS Truepower)
[3] Perry A 2017 Cross validation of operational energy assessments *AWEA Wind Resource and Project Energy Assessment Conference* (Vaisala)
[4] Johansen N 2017 High resolution wind farm energetic availability *AWEA Wind Resource and Project Energy Assessment Conference* (Avangrid Renewables)
[5] Parkhe V 2017 Can data science be used to predict turbine performance? *AWEA Wind Resource and Project Energy Assessment Conference* (DNV GL)
[6] National Aeronautics and Space Administration, Goddard Space Flight Center, Global Modeling and Assimilation Office; URL `https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/`
[7] *(discussions within)* AWEA Standards Development Wind Standards Committee, TR-1 subcommittee; URL `https://www.awea.org/standards-development` *Group chair:* Kramak B (AWS Truepower)
[8] *(discussions within)* Atmosphere to Electrons: Performance, Risk, Uncertainty, and Finance (PRUF) URL `https://a2e.energy.gov/about/pruf` *Project lead:* Fields, M J (NREL)
[9] International Energy Commission 2017 International standard: Wind energy generation systems – part 12-1: Power performance measurements of electricity producing wind turbines Tech. Rep. IEC 61400-12-1
[10] Yesilbudak M 2016 Partitional clustering-based outlier detection for power curve optimization of wind turbines *5th International Conference on Renewable Energy Research and Applications* ed IEEE
[11] Lydia M, Selvakumar A I, Kumar S S and Kumar G E P 2013 *IEEE Transactions on Sustainable Energy* **4** 827–835 ISSN 1949-3029
[12] Lydia M, Kumar S S, Selvakumar A I and Kumar G E P 2014 *Renewable and Sustainable Energy Reviews* **30** 452–460
[13] Power Curve Working Group URL `http://www.pcwg.org/` *Group chair:* Geer T (DNV-GL)