Received: 29 December 2020 | Revised: 5 April 2021 | Accepted: 23 April 2021

DOI: 10.1002/we.2645

RESEARCH ARTICLE

WILEY

Lowering post-construction yield assessment uncertainty through better wind plant power curves

Nicola Bodini[1] | Mike Optis[1] | Jordan Perr-Sauer[2] | Eric Simley[1] | M. Jason Fields[1]

[1]National Wind Technology Center, National Renewable Energy Laboratory, Golden, Colorado, USA

[2]Computational Science Center, National Renewable Energy Laboratory, Golden, Colorado, USA

Correspondence
Nicola Bodini, National Wind Technology Center, National Renewable Energy Laboratory, Golden, Colorado, USA.
Email: nicola.bodini@nrel.gov

Funding information
Wind Energy Technologies Office; U.S. Department of Energy (DOE), Grant/Award Number: DE-AC36-08GO28308

Abstract

Many operational analyses of wind power plants require a statistical relationship, which can be called the wind plant power curve, to be developed between wind plant energy production and concurrent atmospheric variables. Currently, a univariate linear regression at monthly resolution is the industry standard for post-construction yield assessments. Here, we evaluate the benefits in augmenting this conventional approach by testing alternative regressions performed with multiple inputs, at a finer time resolution, and using nonlinear machine-learning algorithms. We utilize the National Renewable Energy Laboratory's open-source software package OpenOA to assess wind plant power curves for 10 wind plants. When a univariate generalized additive model at daily or hourly resolution is used, regression uncertainty is reduced, in absolute terms, by up to 1.0% and 1.2% (corresponding to a −59% and −80% relative change), respectively, compared to a univariate linear regression at monthly resolution; also, a more accurate assessment of the mean long-term wind plant production is achieved. Additional input variables also reduce the regression uncertainty: when temperature is added as an input to the conventional monthly linear regression, the operational analysis uncertainty connected to regression is reduced, in absolute terms, by up to 0.5% (−43% relative change) for wind power plants with strong seasonal variability. Adding input variables to the machine-learning model at daily resolution can further reduce regression uncertainty, with up to a −10% relative change. Based on these results, we conclude that a multivariate nonlinear regression at daily or hourly resolution should be recommended for assessing wind plant power curves.

KEYWORDS

machine learning, operational analysis, post constriction yield assessment, uncertainty

1 | INTRODUCTION

Wind power plant financial transactions, both in the pre-construction and operational stages, require an accurate assessment of the wind plant annual energy production (AEP) through a regression relationship between the power produced by the whole wind plant and concurrent atmospheric variables, to determine a relationship that can be called a wind plant power curve. While wind turbine manufacturers generally certify the power curve of each single wind turbine based on the method of bins described in the IEC 61400-12-1 and IEC 61400-12-2 standards,[1,2] the relationship to calculate the power output of the whole wind plant cannot be easily determined as the sum of the power curves of the single turbines, due to the complex interaction between wake effects, topographic variability, and other losses.[3] Surprisingly, despite the major economic importance of the wind plant long-term annual energy production assessment, no standards on wind plant post-construction yield assessment (PCYA) exist, and the available literature on the topic is limited to a peer-reviewed journal,[4] a consultant report,[5] an academic thesis,[6] and limited conference proceedings.[7,8]

Conversations with major wind energy consultants who represent most of the operational market share in North America revealed that simple mathematical approaches are used to determine wind plant power curves for PCYA. Specifically, all the analysts we consulted confirmed that in the PCYA, a univariate linear regression at a monthly resolution between wind speed and wind plant power production is the current industry standard to derive wind plant power curves. In more detail, density-adjusted wind speed[1,2] is generally used to strengthen its correlation with wind plant power production:

$$U_{\mathrm{dens,corr}} = U\left(\frac{\rho}{\rho_{\mathrm{mean}}}\right)^{1/3} \tag{1}$$

where $U_{\mathrm{dens,corr}}$ is the density-corrected wind speed, U is the wind speed, ρ is the air density (at the same height as wind speed), ρ_{mean} is the mean density over the entire period of record, and the exponent 1/3 is obtained from the cubic relationship between wind power and wind speed.[9]

While applying such a simple regression model leads to a quick operational analysis implementation with easily interpretable results, this approach has several limitations. First, performing a regression at a monthly resolution limits the ability to accurately capture the full variability of the atmospheric properties and transfer it into the long-term wind power plant performance assessment. For example, the wind resource is known to experience strong seasonal[10,11] and diurnal[12,13] cycles in many locations, which would translate to seasonal and diurnal variability of the wind plant energy production,[14,15] and could only be captured by performing operational analysis at a finer time resolution. When using a finer time resolution, however, the relationship between density-adjusted wind speed and energy production strongly deviates from linear.[3] Moreover, wind speed is not the only atmospheric feature driving wind plant power production, which can be influenced by other variables and processes, such as, but not limited to, temperature, wind direction, and turbulence.[16,17]

Therefore, adopting a multivariate and nonlinear regression model could capture the impact of additional physical variables on wind plant power production, thus leading to a more accurate PCYA. We expect that these improvements over the current industry standard could ultimately lead to a reduction of the uncertainty component connected to the regression applied in the development of wind plant power curves. Reducing the uncertainty associated with these relationships would result in better terms on transactions for operational wind plants, such as refinancing, buying, and selling. Halberg[18] estimated how a 1% absolute reduction in AEP uncertainty would translate to about $500,000 improvements in net present value for a 50-MW wind power plant. In a macroeconomic perspective, savings of this type would become even more substantial in the near future, given the impressive growth that wind energy has been experiencing, with the global installed wind capacity increasing by 91% from 2012 to 2017, with an additional 56% expected increase by 2022.[19]

Given the large amount of data and the complexity of the nonlinear relationships between multiple variables that can be expected when wind plant power curves are assessed at a submonthly time resolution, the algorithms developed and popularized in the machine-learning literature seem natural candidates for implementing such an augmented power calculation. Machine-learning algorithms have successfully been applied to several wind-energy-related problems, with applications including wind power forecasting,[17,20,21] turbine power curve modeling,[22] turbine faults and controls,[23] and turbine blade management.[24] Recently, a few studies have applied machine-learning approaches to the determination of wind plant power curves. After demonstrating the limits, at high wind speeds, of the conventional method of bins[1,2] when applied to whole wind plants, Wan et al.[25] proposed a neural-network-based approach for wind plant power curve determination, using wind speed and wind direction data as inputs, and tested it at one operational wind plant. Marvuglia et al.[26,27] evaluated a similar approach, using wind speed as only input to the model. Velázquez et al.[28] considered two wind plants, and explored the benefits of including wind speed and wind direction data at various locations in the vicinity of the plants as additional inputs to a neural network.

Given this vibrant research context, the wind energy industry is also actively developing literature to better document and support the use of machine-learning-based regression relationships to support a range of operational analysis applications, including PCYA. Specifically, the American Wind Energy Association (AWEA) Technical Report (TR)-1 working group is actively working on recommending best practices for the use of machine-learning regression models for wind power plant energy assessment purposes. The study presented here aims to support the AWEA TR-1 effort in exploring the value of moving beyond univariate monthly based regression for wind plant power curves in the context of PCYA.

1.1 | Study goal

Here, we aim to demonstrate how applying multivariate machine-learning regression models at a daily and an hourly time resolution to determine wind plant power curves for PCYA can improve regression accuracy and reduce uncertainty compared to the current industry standard of a univariate linear regression at monthly resolution.

We adopt an open-source, augmented operational analysis toolkit and apply it to assess wind plant power curves using operational data from 10 wind plants in the United States, as detailed in Section 2. We then discuss and quantify the benefits (in terms of regression uncertainty and regression mean error) introduced when including additional inputs to the regression model, performing the analysis at a daily or an hourly resolution, and adopting a machine-learning regression model (Section 3). Finally, we highlight and summarize the main findings of our analysis and suggest future work in Section 4.

2 | DATA AND METHODS

2.1 | Wind plant operational data: The Wind Plant Performance Prediction Benchmark initiative

The Wind Plant Performance Prediction (WP3) Benchmark initiative[29] is a joint industry project aimed at sharing data and performing preconstruction and operational performance reconciliation at scale. The project has unprecedented support from all corners of the industry and represents the largest validation and data sharing initiative of its kind. Here, we perform PCYA for the 10 existing wind plants that have been selected in WP3 Phase 1. These wind plants have been chosen to mirror expected near-term future deployments in the US Central Plains. All wind power plants have a capacity of at least 100M˙W, with wind turbines with a rated power of least 1.5 MW and a hub height at 80 m above ground level or higher.

To preserve the confidential nature of proprietary wind plant data, we assigned project codes to each of the 10 wind plants. If the operational data provided by the project owner included data from the first 12 months of operation of the wind power plant (spin-up period), we excluded them from the analysis, thereby keeping only data that represent the mature operation of the plant. Moreover, for some projects we excluded additional months at the beginning of the valid period of record to obtain (near) full annual cycles and even out the potential impacts of seasonal variability that could skew the performed regression between the atmospheric variables and plant production (e.g., see Figure 3 and relative description in Section 3.1). The project codes, together with the period of record used for the PCYA, are shown in Table 1. The table also shows which wind plants have operational data at a submonthly time resolution, which allows for the wind plant power curves to be determined at a daily and an hourly resolution. All 10 wind plants have data for the power curves to be assessed at a monthly resolution.

2.2 | Reanalysis data

To assess operational wind plant power curves, atmospheric data are required to both perform the period-of-record regression with the wind plant power production data and then apply the regression to the long-term atmospheric conditions, with the so-called windiness correction. For this correction, we use 20 years of atmospheric data, from January 1997 through December 2017. We derive the atmospheric data from the

TABLE 1 Details of the 10 WP3 Phase 1 wind power plants analyzed in our study

Project code	Period of record used in analysis	Monthly resolution possible?	Daily resolution possible?	Hourly resolution possible?
c43	1 January 2015 to 31 December 2018	Yes	No	No
h21	1 May 2015 to 30 April 2018	Yes	No	No
n23	1 September 2015 to 31 August 2018	Yes	Yes	Yes
o46	1 March 2016 to 28 February 2018	Yes	Yes	Yes
q56	1 January 2015 to 30 October 2017	Yes	Yes	Yes
r47	1 September 2015 to 31 August 2017	Yes	No	No
t75	1 September 2015 to 31 August 2017	Yes	No	No
u15	1 September 2015 to 31 August 2018	Yes	Yes	No
w34	1 January 2015 to 30 October 2017	Yes	Yes	Yes
x41	1 May 2016 to 30 April 2019	Yes	Yes	No

TABLE 2 Coefficients of determination between density-corrected monthly average wind speed from four reanalysis products and the energy production data from the 10 wind plants considered in the analysis

Project Code	R^2 ERA5	R^2 MERRA-2	R^2 ERA-Interim	R^2 NCEP-2
c43	**0.74**	0.67	0.71	0.28
h21	**0.92**	**0.92**	0.81	0.82
n23	**0.78**	0.73	0.72	0.64
o46	**0.93**	0.90	**0.93**	0.80
q56	**0.84**	**0.84**	0.63	0.62
r47	0.90	**0.92**	0.87	0.84
t75	**0.84**	0.79	0.76	0.65
u15	**0.95**	**0.95**	0.92	0.89
w34	**0.93**	**0.93**	0.92	0.90
x41	**0.94**	0.85	0.86	0.80

Note: The largest coefficient(s) of determination for each wind farm is highlighted in bold.

ERA5 reanalysis product, which provides global atmospheric data at a $0.25^\circ \times 0.25^\circ$ spatial resolution.[30] ERA5 was chosen as it is been shown to provide the best correlation with near-surface winds.[31] Also, it provides the largest correlation between density-corrected wind speed and the energy production data from all 10 wind plants compared to a set of other reanalysis products (Modern Era Retrospective analysis for Research and Applications, Version 2 [MERRA-2],[32] NCEP-2,[33] and ERA-Interim[34]), as shown in Table 2.

Specifically, we use the diagnosed hourly 100-m wind components (to derive both wind speed and direction), 2-m temperature, and surface pressure. We then calculate 100-m air density using the surface pressure and 2-m temperature fields and the hydrostatic assumption, and calculate the density-corrected 100-m wind speed using Equation (1) at the hourly time resolution native to this dataset.

2.3 | Wind plant power curve approach

The WP3 initiative has led to the development of OpenOA,[1,35,36] an open-source Python package for wind power plant data operational analysis. As mentioned in the introduction, industry standard algorithms for wind plant power curves currently perform the calculation at a monthly resolution, using density-corrected wind speed as the only input to a linear regression model to correlate with wind plant power production. To overcome the limitations of such an approach and possibly reduce the uncertainty connected to the PCYA, we implement the following set of features in OpenOA:

- The operational analysis can be performed at a monthly, daily or hourly resolution.
- The regression between wind speed and wind-plant-produced power can be augmented by incorporating temperature and/or wind direction (more specifically, its sine and cosine, to preserve the cyclical nature of this variable) as additional inputs.
- The type of regression algorithm can be chosen between the conventional linear regression and (when enough training data are available, that is, for plants with operational data at a daily or hourly resolution) more sophisticated machine-learning algorithms, which are capable of capturing the nonlinearity in the relationship between wind plant energy and atmospheric variables at submonthly resolution.

The OpenOA pipeline we follow for wind plant power curve assessment can be described as follows:

1. Long-term wind speed data from the reanalysis product are density-corrected with Equation (1).
2. The atmospheric inputs (here: density-corrected wind speed, wind direction, and temperature) from a long-term reanalysis product (here: ERA5) are averaged at the time resolution that will be used for the PCYA (here: monthly, daily or hourly).
3. Revenue meter data, production-based availability, and curtailment losses are summed at the time resolution that will be used for the operational analysis (here: monthly, daily or hourly), to derive gross energy data:

$$\text{Gross energy} = \text{Revenue meter energy} + \text{Curtailment losses} + \text{Availability losses} \tag{2}$$

4. When the operational analysis is performed at a monthly resolution, gross energy data are normalized to 30-day months (e.g., for January, the gross energy values are multiplied by 30/31).

[1]https://github.com/NREL/OpenOA

5. A regression between monthly/daily/hourly gross energy production and concurrent reanalysis atmospheric data (density-corrected wind speed and, when desired, wind direction and/or temperature) is performed.
6. Long-term monthly/daily/hourly average atmospheric properties (density-corrected wind speed and, when desired, wind direction and/or temperature) are then calculated, using 20 years of the available long-term reanalysis data.
7. A long-term data set of gross energy production is obtained by applying the fitted regression relationship to the long-term monthly/daily/hourly atmospheric data from the reanalysis product, with the long-term (or windiness) correction.
8. When the operational analysis is performed at a monthly resolution, the resulting long-term monthly gross energy, which is based on 30-day months, is then denormalized to the actual number of days in each calendar month (e.g., for January, the value is multiplied by 31/30).

To then assess long-term annual net energy production, long-term estimates of availability and curtailment losses would need to be applied to the gross energy estimates. This final step is omitted in our analysis, as we are only focusing on the uncertainty connected to the regression performed in the process. Figure 1 shows a diagram of the main steps of this methodological pipeline.

OpenOA applies a set of filters during the pre-processing stage of its PCYA calculations to discard time stamps (i.e., hours, days or months) in which at least one of the following situations occurs:

- the sum of percent availability and percent curtailment losses is greater than 15%;
- more than 1% of the raw high-resolution reported plant data (either energy production, availability or curtailment) are missing;
- wind speed is less than 0 m/s or greater than 40 m/s;
- wind speed is between 5 and 40 m/s, and the reported plant power is less than 2% or more than 120% of the nominal plant capacity.

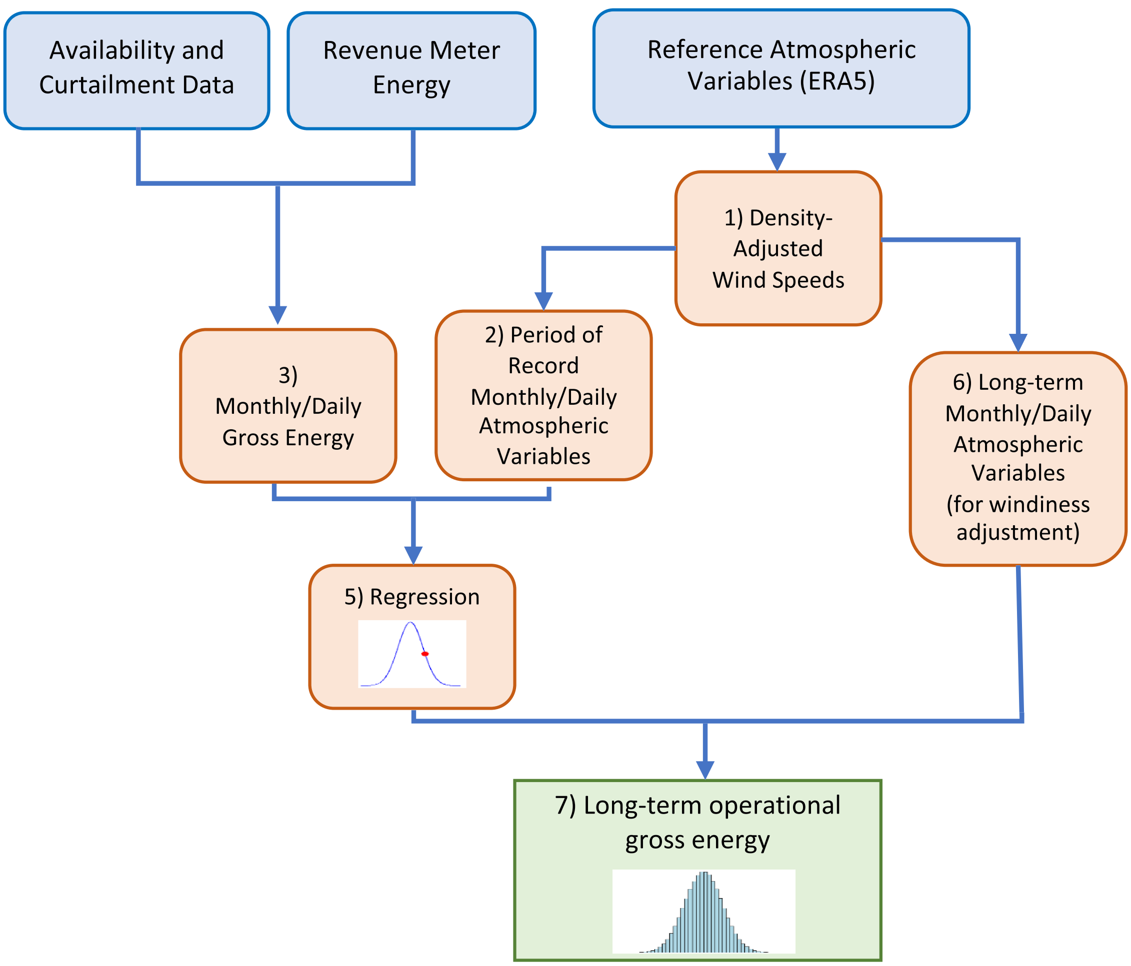

FIGURE 1 Diagram showing the main steps of the PCYA performed in our study. The numbers in the boxes tie the process with the numbered steps in Section 2.3. The distributions highlight the phases of the process that are affected by the Monte Carlo approach for uncertainty quantification, as detailed in Section 2.4

A more detail discussion on the application of these filters is given in Section 3.2.

OpenOA allows the user to choose between three machine-learning regression models: a generalized additive model (GAM[37]), a gradient boosting model,[38] and an extremely randomized trees regressor.[39] In this study, as a proof of concept of the machine-learning capabilities, we use the GAM, which has previously been demonstrated to provide accurate fits to turbine power curves.[40] We note how choosing one of the two other machine-learning regression models would not significantly alter the conclusions of our analysis. The number of splines in the GAM model is sampled from 5 to 40, with 10 values tested, using a randomized search with five-fold cross-validation. The mean square error between observed and machine-learning-predicted power values is used as a performance metric to train the learning algorithm. An example of a GAM fitted to daily density-corrected wind speed data for project w34 is shown in Figure 2. The nonlinearity of the relationship at fine time resolution clearly emerges.

2.4 | Uncertainty in operational wind plant power curve assessment

Within OpenOA, uncertainty quantification in the PCYA is performed with a Monte Carlo approach. Inputs to the analysis and intermediate calculations are randomly sampled from distributions derived from imposed or calculated uncertainties. By repeating the calculation thousands of times (in our analysis, 20,000 for each regression setup), a distribution of gross energy is obtained, from which the PCYA uncertainty can be calculated, in terms of its coefficient of variation (i.e., the standard deviation of gross energy values divided by their mean). This long-term operational energy uncertainty is affected by several components,[4,41] which are connected to the error in the atmospheric data, the interannual variability of the wind resource, the regression model used, the uncertainty in the long-term correction, and others. All the augmented features analyzed in our study (i.e., additional input variables, finer time resolution, more sophisticated regression models) are strictly connected to the regression phase of the operational wind plant power curve assessment. Therefore, to assess the value in adding these augmented features, we will focus on the PCYA uncertainty component connected to the regression phase. Quantifying the relative importance of regression uncertainty to the total PCYA uncertainty is challenging given the scarcity of literature on the topic. Bodini et al.[4] have recently performed PCYA on over 470 wind plants in the U.S., and found that on average regression uncertainty accounts for about a third of the total uncertainty resulting from the sources considered in their analysis.

Several ways exist to quantify regression uncertainty within a Monte Carlo framework. Bodini et al.[4] have quantified regression uncertainty in a linear regression model for PCYA by sampling multiple values of the regression slope and regression intercept from a multivariate normal distribution centered around the regression best-fit parameters and the covariance matrix equal to one of the best-fit parameters. For a simple linear

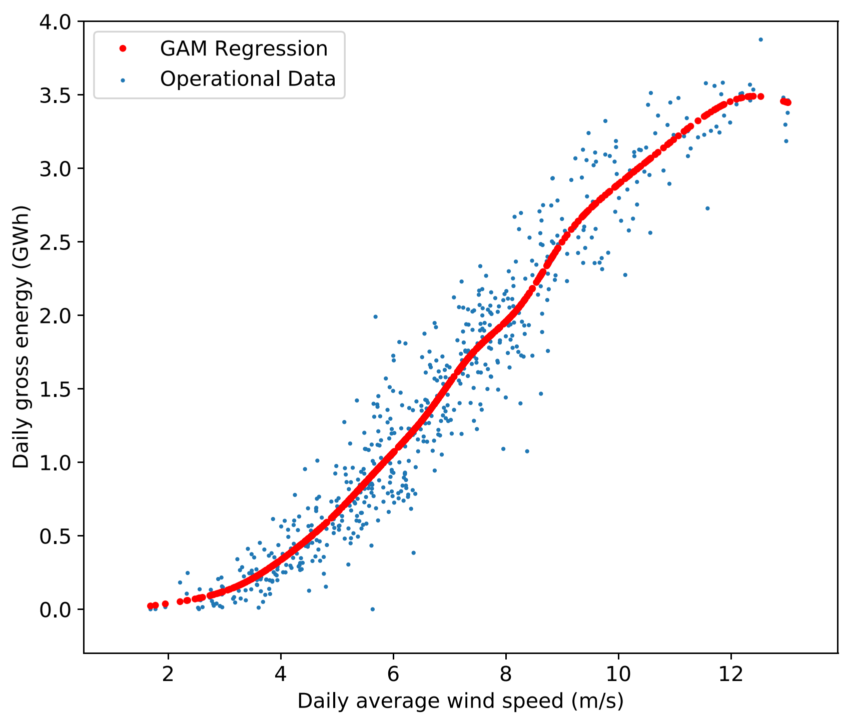

FIGURE 2 Example of the application of a GAM regression algorithm for the wind plant power curve assessment performed at a daily resolution (for project w34)

regression, an analytic equation that can derive such a covariance matrix can be found (e.g., in JCGM 100:2008[42]). However, such a technique is not directly applicable when a more complex machine-learning regression model is applied. Therefore, in our analysis we use a bootstrapping approach to quantify regression uncertainty.[43] At each Monte Carlo iteration, we perform a regression using a slightly different data set, which is obtained by randomly sampling with replacement (i.e., each data point can be included either zero, one or multiple times, and the sampled data set has the same size of the original data set) the reanalysis and concurrent wind plant power production within the wind plant period of record. We then use, as metric of this uncertainty component, the coefficient of variation of the resulting long-term operational gross energy distribution resulting from the multiple Monte Carlo iterations. Figure 3 shows an example of the application of the bootstrapping approach to quantify (monthly) linear regression uncertainty, for project w34.

On the left panel, the spread of the regression lines from 20,000 Monte Carlo iterations is shown, which results in the distribution of long-term operational gross energy shown in the panel on the right. The regression uncertainty is then estimated in terms of the coefficient of variation of this distribution (here: CoV = 1.70%).

3 | RESULTS

3.1 | Operational wind plant power curve assessment at a monthly resolution: Value of adding temperature for projects with large seasonal variability

We start our analysis of the benefits of augmented regression for wind plant power curves by assessing the value in adding input features to the conventional linear regression performed at the monthly time resolution. In this context, wind direction represents a problematic input for linear models because of both its cyclical nature (i.e., 0° and 360° are the same) and generally nonmonotonic relationship with wind plant power production. Multiple wind plant power curves for various wind direction sectors could be built as in Khalid and Savkin,[44] but this is beyond the scope of our current analysis. Here, we only focus on the value of adding temperature as a second input to the linear regression model. In fact, many wind plants experience seasonality in their wind power production, as it is possibly connected to different wind regimes or other environmental factors, so that the relationship between density-corrected wind speed and produced power can change significantly when considering different times of the year. In absence of direct turbulence metrics from the reanalysis products, which might better represent seasonal wind regimes, we use temperature as an easily available proxy for such variability.

Figure 4 shows an example of such a seasonal variability for one of the 10 wind plants considered in our analysis. The left panel shows the best-fit line when a single linear regression between density-corrected monthly average wind speed and monthly gross energy is performed without any data segregation by season: in this case, the coefficient of determination is $R^2 = 0.74$. However, when a different regression line is

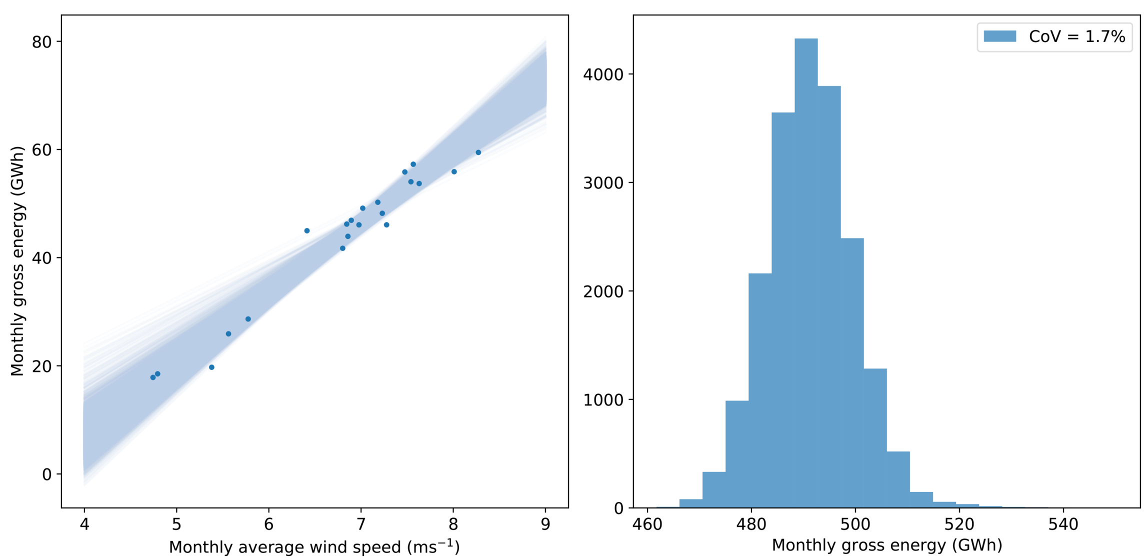

FIGURE 3 Example of applying the bootstrapping approach to quantify regression uncertainty, when a linear regression between density-adjusted wind speed and gross energy is applied at the monthly resolution (project w34). The left panel shows 20,000 regression lines obtained by randomly sampling (with replacement) the inputs to the monthly regression model. The histogram on the right shows the resulting distribution of long-term operational gross energy, with its coefficient of variation shown in the legend

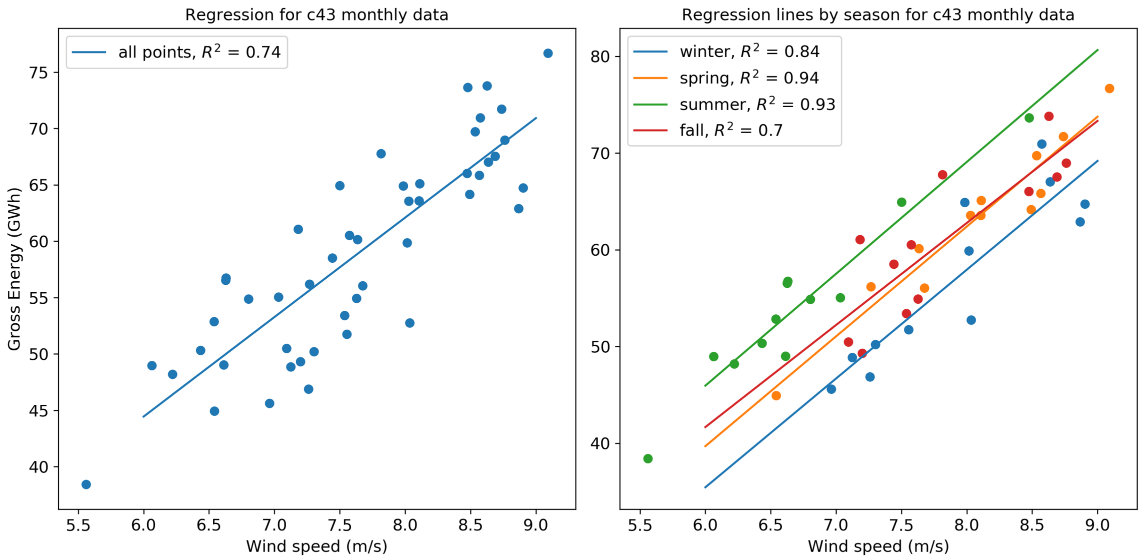

FIGURE 4 (left) Monthly linear regression between the density-adjusted monthly average wind speed and gross energy for project c43; (right) same, but with four linear regression applied, with monthly average data classified by season

calculated for each season (right panel), clear differences emerge: density-corrected wind speed being the same, the wind plant in summer generates up to 20% more power. By creating four best-fit regression models, the accuracy of the regression is generally larger, with the average of the four correlation coefficients $R^2 = 0.85$, representing a 15% increase over the R^2 obtained without any seasonal segregation, and peak values greater than 0.90 in summer and winter.

In order to capture the importance of the seasonal variability of the wind plant power production regimes in the framework of an augmented regression for wind plant power curves, we include temperature as additional input (together with density-corrected wind speed) in a monthly linear regression. The first metric we consider to quantify the benefits of such augmentation is the normalized percent root-mean-square error (NRMSE) of the regression, which we define as

$$\mathrm{NRMSE} = \frac{\mathrm{RMSE}}{N} \tag{3}$$

where N is the nominal energy (GWh) that the wind plant can produce in one month/day/hour, calculated by converting the wind plant nominal power capacity (GW) into GWh at the appropriate time resolution. The normalization allows for a direct comparison between wind power plants with significantly different sizes and, therefore, annual energy production, and between the different time resolutions used in the PCYA. Table 3 shows the change in the average (across the 20,000 Monte Carlo iterations) percent NRMSE when temperature is added as a second input to the regression for the 10 wind plants, obtained from the application of the operational analysis pipeline described in Section 2.3.

Adding temperature as an additional input to the monthly linear regression always decreases the mean regression error. For c43, which is the wind power plant with the largest seasonal variability in the set of 10, the mean NRMSE decreases by 45%, from 3.3% to 1.8% of the nominal energy that the wind plant can produce in one month. The distributions of long-term operational gross energy for wind plants c43 and o46, two of the projects that showed the largest decrease in NRMSE are shown in Figure 5. We note how for some of the wind plants the more accurate regression also leads to a slight shift (always <1%) of the mean long-term operational gross energy production.

Next, we can consider how adding temperature as an additional monthly linear regression input impacts the long-term operational gross energy uncertainty component connected to regression. Figure 6 shows the absolute and relative changes in uncertainty (expressed as the percent coefficient of variation of the long-term operational gross energy distribution) for the 10 wind plants.

While most of the projects do not see a significant change in regression uncertainty, c43, o46, and x41, which all experience a strong seasonal variability in plant power production, see a reduction in their regression uncertainty, in absolute terms, by up to 0.5% (corresponding to a 43% relative change in uncertainty), which could translate into major savings in the plants' financial transactions, such as refinancing, purchasing/selling, and mergers/acquisitions. This reduction in uncertainty can be qualitatively observed as a narrowing of the long-term gross energy distributions in Figure 5.

TABLE 3 Average normalized (by the nominal energy that each wind plant can produce in 1 month) percent RMSE for the PCYA setups at a monthly resolution

	Average normalized percent RMSE Monthly linear regression	
Project Code	Wind speed	Wind speed + Temperature
c43	3.3%	1.8%
h21	2.5%	2.4%
n23	3.4%	3.1%
o46	2.2%	1.5%
q56	1.8%	1.7%
r47	4.5%	4.1%
t75	2.3%	2.1%
u15	3.2%	3.1%
w34	2.4%	2.1%
x41	3.6%	2.2%

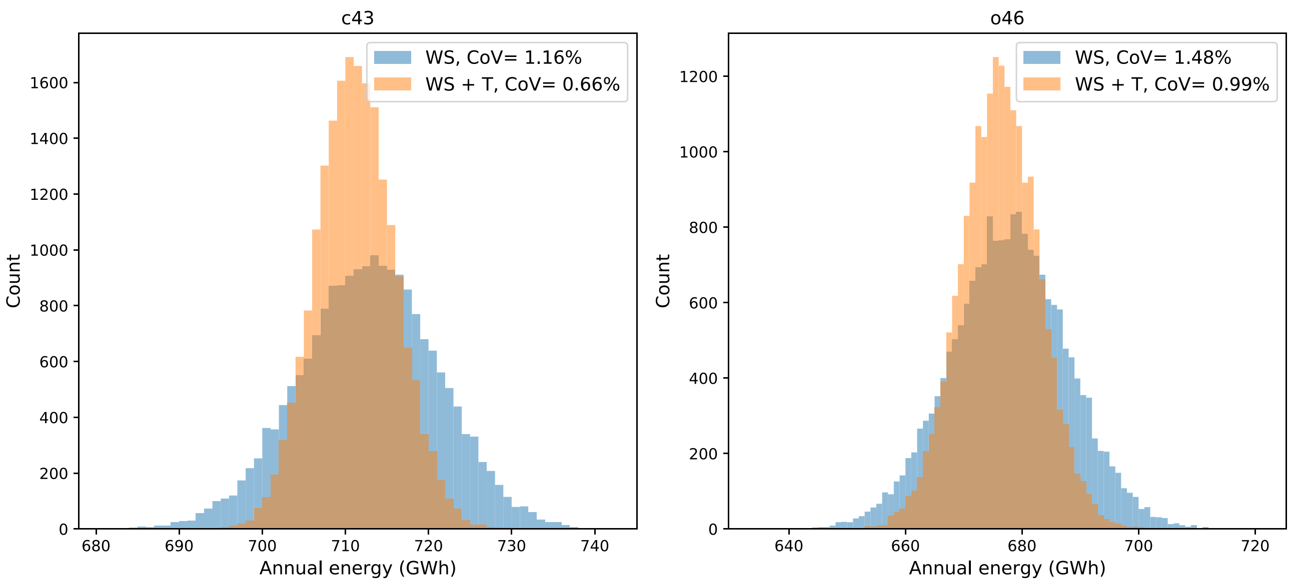

FIGURE 5 Long-term annual operational gross energy distributions obtained from linear regressions at a monthly resolution using density-corrected wind speed only and density-corrected wind speed together with temperature as inputs, for the two wind plants that show the largest reduction in regression uncertainty. For each setup, we ran 20,000 Monte Carlo iterations. The percent coefficient of variation for each case is shown in the legends

3.2 | Value in pushing toward finer time resolution and more sophisticated regression algorithms

The next augmentation we consider is the value in applying machine-learning regression algorithms and in performing the analysis at an even finer (here, daily, hourly) time resolution. In fact, adopting machine-learning models at a monthly resolution would not be feasible nor beneficial, as a large amount of data is needed to be able to successfully train a machine-learning model, and this would require an infeasibly long period of record. Also, the relationship between density-adjusted wind speed and wind plant power production at a monthly timescale is close to being linear[4], since cases with very low or very high wind speeds (where wind turbine power curves strongly deviate from being linear) are averaged out by the coarse time resolution. On the other hand, applying a linear model at a daily or hourly resolution would generally not be appropriate, as the relationship between density-corrected wind speed and wind plant energy production at daily and hourly resolution significantly deviates from being linear (e.g., Figure 2) as it includes samples over a wider range of wind speeds.

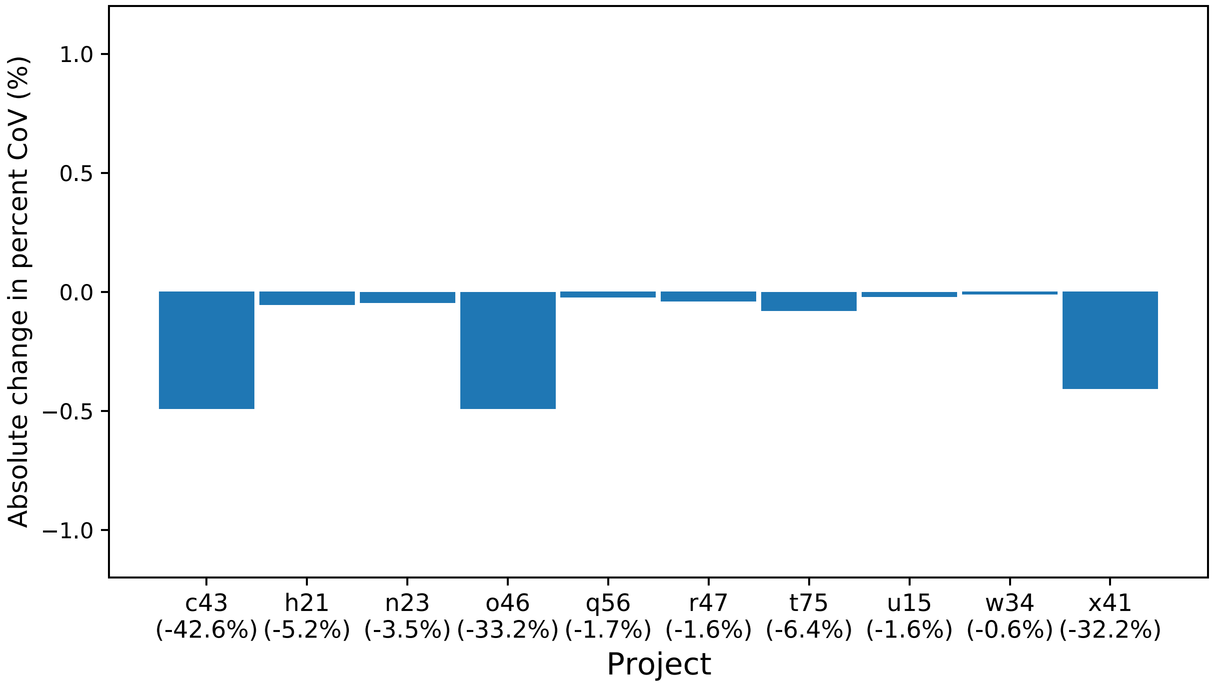

FIGURE 6 Change in magnitude in long-term annual operational gross energy regression uncertainty (expressed as the percent coefficient of variation) when temperature is added as an input to the linear regressions at the monthly resolution. The corresponding relative change is shown in the *x* axis label

Before being able to directly compare results obtained from regression models at different resolutions, a careful evaluation of the filters applied by OpenOA (see description in Section 2.3) is needed. The result of the application of the OpenOA filtering pipeline is dependent on the time resolution at which the PCYA is performed. More in detail, the use of such filters at a daily and hourly resolution is more time-specific: in fact, if a given month includes some bad data, the whole month would be excluded by the filtering algorithms when the operational analysis is performed at a monthly resolution, whereas only the specific days/hours with low-quality data would be filtered when applying the analysis at a daily/hourly resolution. As a consequence, we can expect the filtered wind speed averaged over the period of record to be different based on the time resolution used in the calculation, which in turn would lead to a shift (in either direction) in the predicted long-term energy production.

To prove this effect, Figure 7 shows the distributions of the density-corrected monthly (left) and daily (right) average wind speed in the period of record for wind plant w34, before and after the OpenOA filters are applied at the two different time resolutions.

By comparing the average wind speed values at the two time resolutions before and after the filters have been applied, it is clear how the filtered periods of record at the two resolutions do not fully overlap. Although the prefilter average wind speed is the same at the two time resolutions (as it should be), the postfilter average wind speed at a daily resolution is, for this wind power plant, about 2.5% lower than the postfilter average wind speed when the filters are applied at a monthly resolution. In fact, we see that at the monthly resolution the majority of the discarded months have relatively low average wind speeds, whereas at the daily resolution the data being discarded are more evenly distributed across the whole range of wind speeds. This difference in filtered mean wind speed between the two resolutions would directly translate in the mean long-term operational gross energy for w34 calculated at the daily resolution being lower than the mean value obtained at the monthly resolution. We have tested this for all the six wind plants with data available at daily resolution and the four wind plants with data available at hourly resolution and found that, when the OpenOA filters are applied at same resolution as the main PCYA, the mean long-term gross energy values obtained when applying a univariate GAM regression at daily or hourly resolution deviate from the ones obtained with a monthly linear regression, with differences up to ±2%. Therefore, we can conclude that the use of operational data at a daily or hourly resolution can also contribute to reducing the amount of data to discard in the PCYA regression, thus leading to more accurate assessments of the mean long-term operational power. However, dealing with different filtered periods of record in our specific analysis is not an optimal condition, as that would introduce some inconsistency in the direct comparison between the PCYA augmentations being tested. Therefore, in the rest of our analysis, we will always apply the OpenOA filters at the monthly resolution, even when the PCYA regression is then applied at a daily or hourly resolution, so that the filtered periods of record stay the same, and cannot affect any observed change in regression uncertainty. The postfilter data availability for each wind plant is shown in Table 4.

With this caveat in mind, we can proceed in our analysis by comparing the change in regression uncertainty between PCYA setups at different time resolutions. Figure 8 compares the Monte Carlo distributions of long-term operational gross energy obtained from a linear regression at

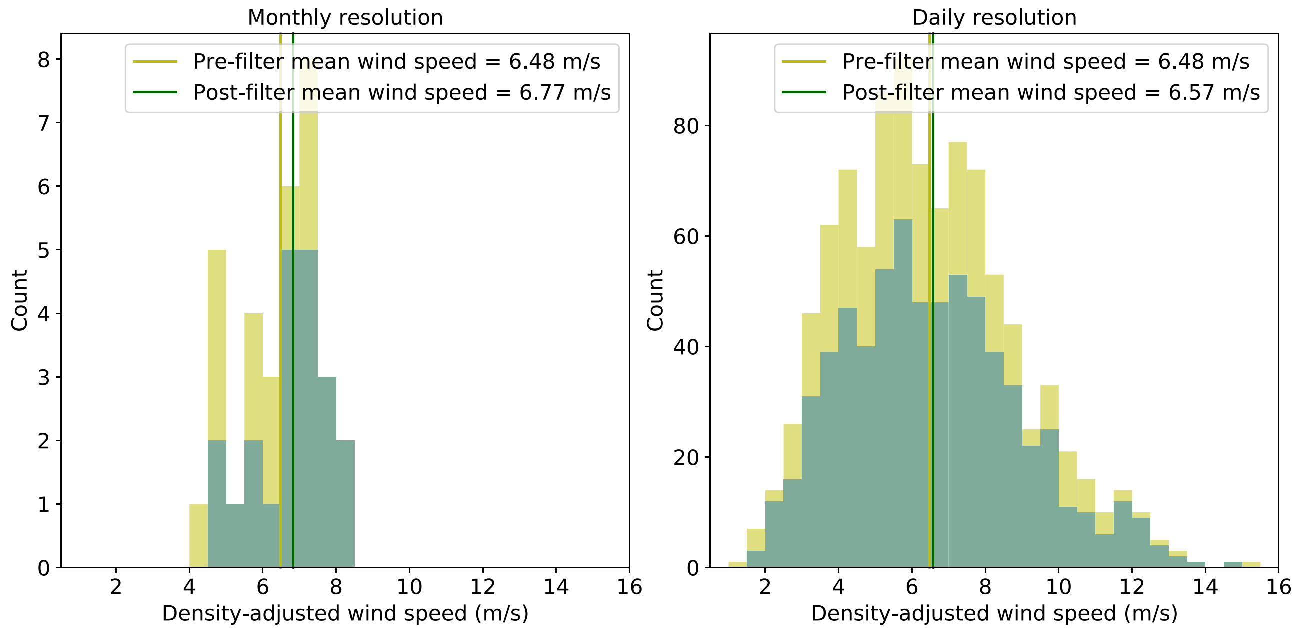

FIGURE 7 Histograms of monthly (left) and daily (right) average density-corrected wind speed in the period of record for wind plant w34 before and after the OpenOA filters have been applied. The vertical lines show the average density-corrected wind speed value for the various cases

TABLE 4 Monthly data availability after the OpenOA filtering pipeline is applied to the available period of record shown in Table 1

Project code	Postfilter data availability (months)
c43	47 out of 48
h21	36 out of 36
n23	34 out of 35
o46	22 out of 23
q56	17 out of 22
r47	24 out of 24
t75	24 out of 24
u15	33 out of 36
w34	21 out of 34
x41	35 out of 36

a monthly resolution, a GAM regression at a daily resolution, and a GAM regression at an hourly resolution, all using only density-corrected wind speed as input. Results are shown for the six wind plants that had operational data available at least at a daily resolution (see Table 1).

In all cases, we find that the application of the daily machine-learning regression greatly narrows the long-term annual operational energy distributions compared to the monthly linear regression, therefore reducing the operational analysis uncertainty component connected to regression. Switching to an hourly resolution further reduces the width of the distributions, thus indicating further reduction in uncertainty. These effects are quantitatively assessed in Figure 9, which shows the absolute and relative changes in uncertainty for the six considered wind plants, once again with uncertainty expressed as the percent coefficient of variation of the long-term gross energy distribution.

For all wind plants, we observe a significant reduction in regression uncertainty when finer time resolutions are used in the PCYA. When adopting the GAM regression at daily resolution, the change in magnitude in regression uncertainty, in terms of CoV, compared to a linear regression at monthly resolution range between −0.3% and −1.0%, which corresponds to a relative reduction between −24% and −57%. On the other hand, performing the analysis at an hourly resolution leads to even larger reductions in uncertainty, between −0.9% and −1.2%, corresponding to a relative reduction between −66% and −81%. Increasing the number of data points used in the wind plant power curve assessment and applying a nonlinear model at a daily or hourly time resolution significantly reduce the uncertainty in the PCYA.

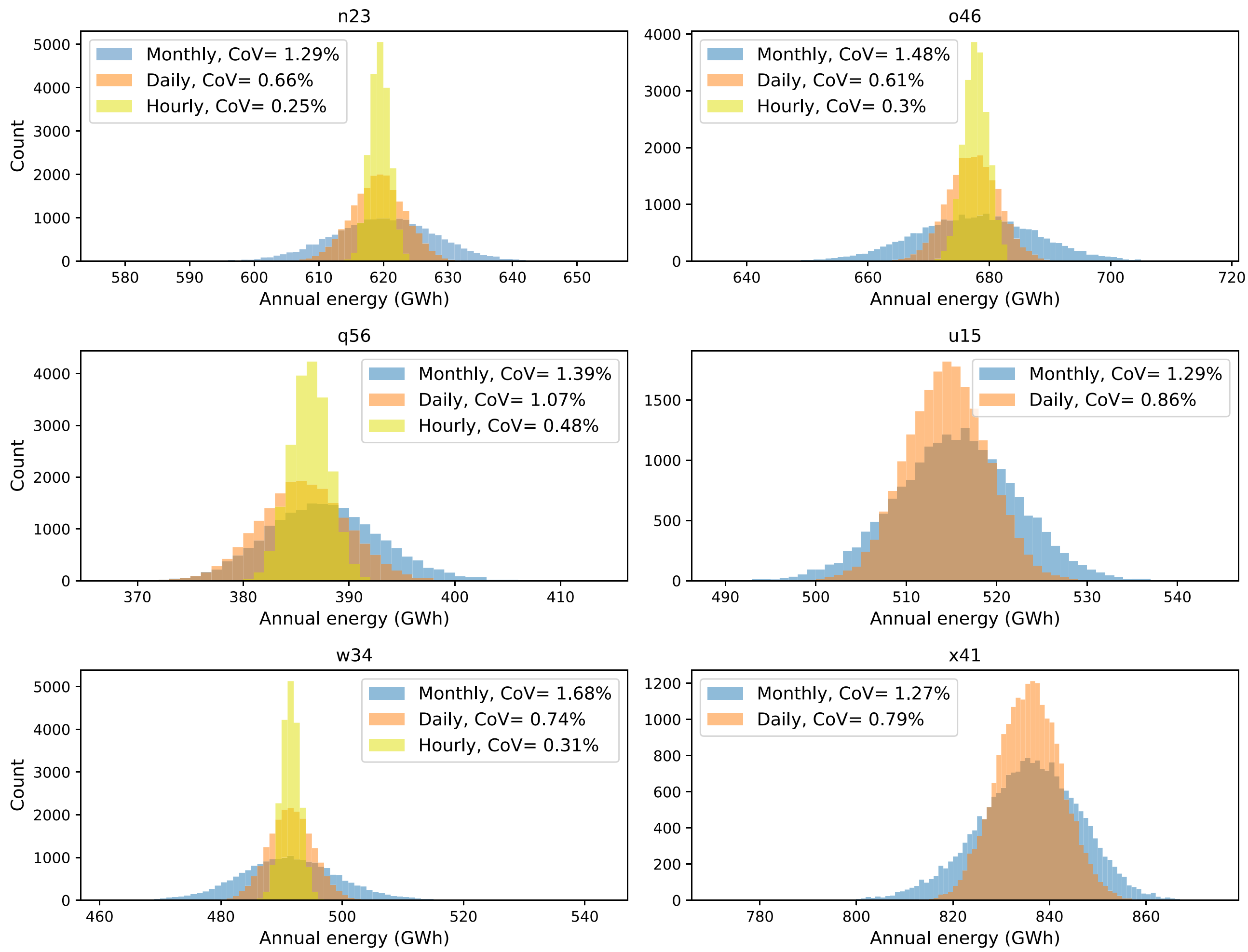

FIGURE 8 Long-term annual operational gross energy distributions obtained from a linear regression at the monthly resolution, from a GAM regression at the daily resolution, and from a GAM regression at the hourly resolution, all using only density-corrected wind speed as an input. For each approach, we ran 20,000 Monte Carlo iterations. The percent coefficient of variation for each case is shown in the legends

3.3 | Value in adding additional input variables for PCYA at daily and hourly resolution

Next, we circle back to the value of considering a multivariate regression, in the context of a machine-learning algorithm at a daily and hourly resolution. To do so, we compare the results from wind plant power curve assessment performed at a daily and hourly resolution using a GAM algorithm, but with different sets of input variables: density-adjusted wind speed only, density-adjusted wind speed and temperature, density-adjusted wind speed and wind direction (more specifically, its sine and cosine), and finally all three variables together.

First, we assess how the normalized percent RMSE varies among the considered regression setups (Tables 5 and 6).

At daily resolution, we find that the error of the regression significantly benefits from the additional input variables: for all six wind plants, it decreases when additional features are added as inputs to the GAM algorithm. In general, we find that adding wind direction as a second regression input has (for all but one wind plant) larger value compared to the case when temperature is considered as second input variable. In any case, including all three variables as inputs always led to the best results, with percent reductions in NRMSE as large as -17% compared to the case of a univariate GAM regression model. Such large error reductions would translate into stronger relationships between atmospheric properties and wind power plant energy production.

Similar considerations hold for the hourly resolution case (Table 6), with percent reductions in NRMSE as large as -10% when all three input variables are included compared to the case of a univariate GAM regression model. Finally, we note how the magnitude of the NRMSE increases as the time resolution gets finer, which is reasonable to expect given the larger scatter in the data without the smoothing effect of a coarse time resolution.

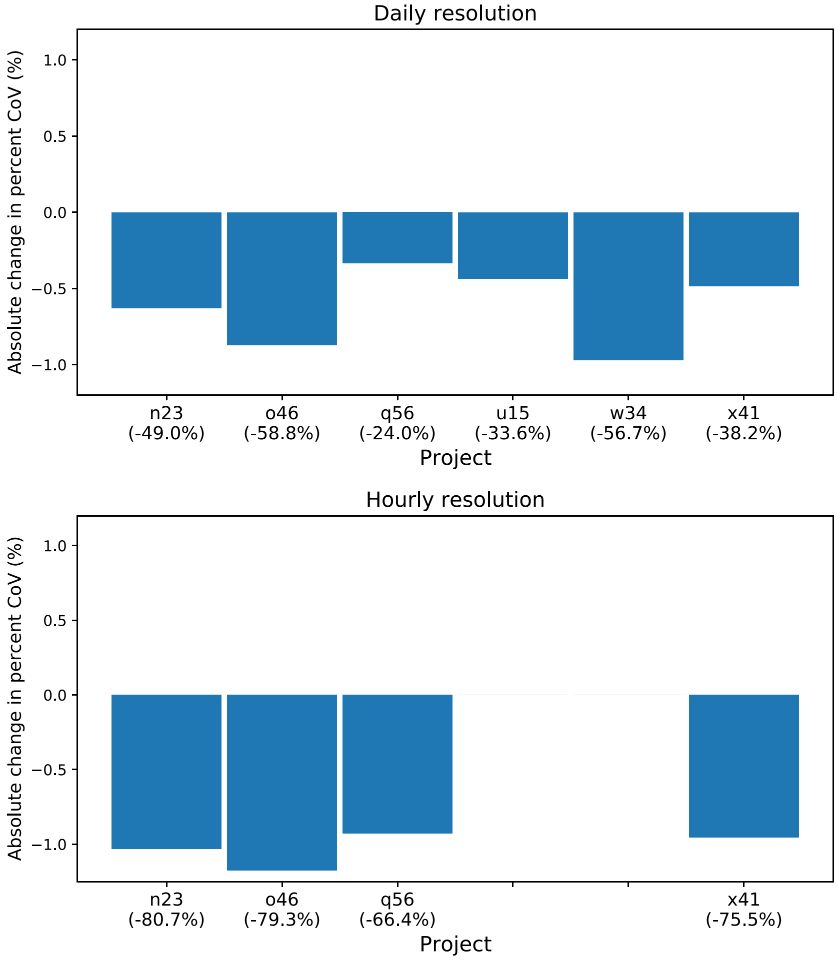

FIGURE 9 Change in magnitude in long-term operational gross energy regression uncertainty (expressed as percent coefficient of variation) when moving from a monthly linear regression to a daily (top) and hourly (bottom) GAM regression, using density-corrected wind speed only as input. The corresponding relative change is shown in the *x* axis labels

TABLE 5 Average normalized (by the nominal energy that each wind plant can produce in 1 day) percent RMSE for the PCYA setups at a daily resolution

	Average normalized percent RMSE Daily GAM regression			
Project Code	Wind Speed	Wind speed + Temperature	Wind speed + Wind Direction	Wind speed + Temperature + Wind Direction
n23	9.7%	9.2%	8.8%	8.6%
o46	5.7%	5.2%	5.2%	4.9%
q56	8.2%	8.0%	7.5%	7.1%
u15	11.6%	11.5%	11.2%	10.9%
w34	7.0%	6.7%	6.5%	6.1%
x41	13.2%	12.8%	11.3%	11.0%

Note: Performing the PCYA at a daily resolution was possible for only six wind plants because of limited data availability.

In terms of the effect of additional regression input variables to the PCYA uncertainty, Figure 10 shows how the uncertainty component connected to regression varies when both temperature and wind direction are added as additional inputs in the GAM regression algorithm, at both daily and hourly resolution.

Project Code	Average normalized percent RMSE Hourly GAM regression Wind speed	Wind speed + Temperature	Wind speed + Wind direction	Wind speed + Temperature + Wind direction
n23	18.3%	18.1%	17.6%	17.4%
o46	14.4%	13.4%	14.1%	13.2%
q56	18.1%	17.9%	17.6%	17.5%
w34	14.9%	14.8%	14.7%	14.5%

Note: Performing the PCYA at an hourly resolution was possible for only four wind plants because of limited data availability).

TABLE 6 Average normalized (by the nominal energy that each wind plant can produce in 1 h) percent RMSE for the PCYA setups at an hourly resolution

The reduction in regression uncertainty across the six wind plants, although present, is lower than what is achieved when moving from a univariate monthly linear regression to a univariate daily or hourly machine-learning model (Figure 9). Still, including additional input variables to the regression model can reduce the regression uncertainty, in relative terms, by up to 10% at a daily resolution, and up to 4% at an hourly resolution, among the sample of six wind plants considered here. The limited uncertainty reduction found on average can be explained by considering the strong relationship between density-corrected wind speed and gross energy at daily and hourly resolution, and by acknowledging that the use of regression uncertainty might not be capturing all the benefits associated with the additional features.

As an example of the more accurate relationship that can be achieved when multiple input variables are used, we show in Figure 11 the wind plant power curve calculated at a daily resolution using a GAM with density-corrected wind speed, temperature, and wind direction as inputs, for

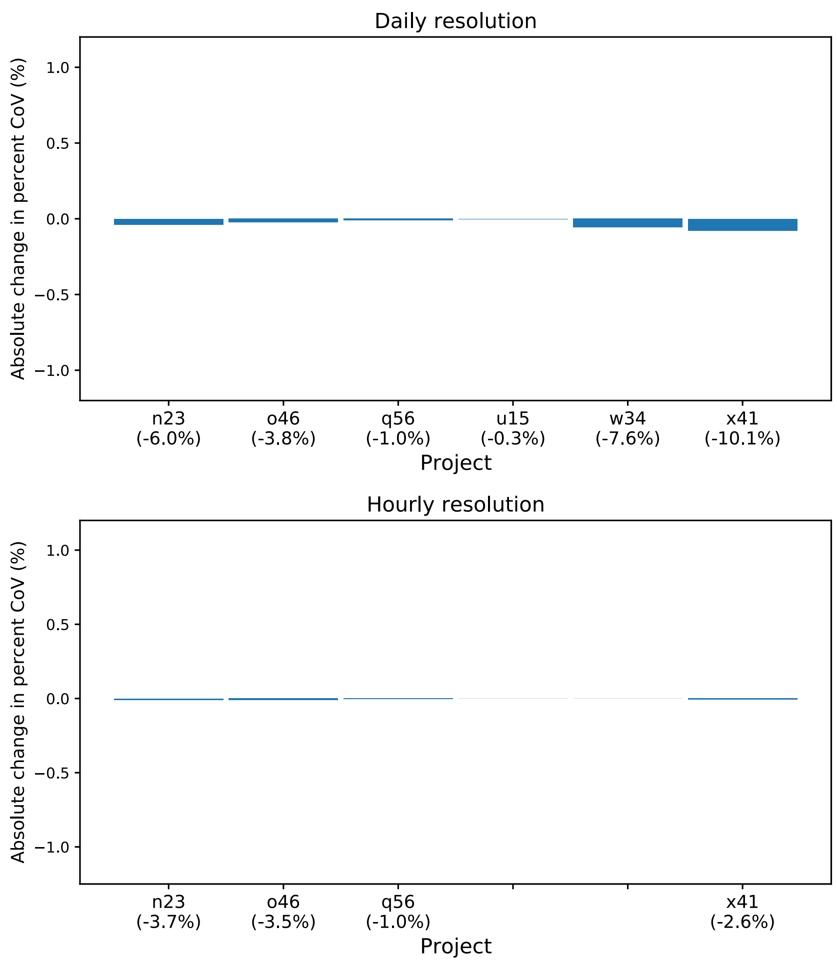

FIGURE 10 Change in magnitude in long-term operational gross energy regression uncertainty (expressed as the percent coefficient of variation) when adding temperature and wind direction as additional inputs to a daily (top) and hourly (bottom) GAM regression. The corresponding relative change is shown in the *x* axis labels

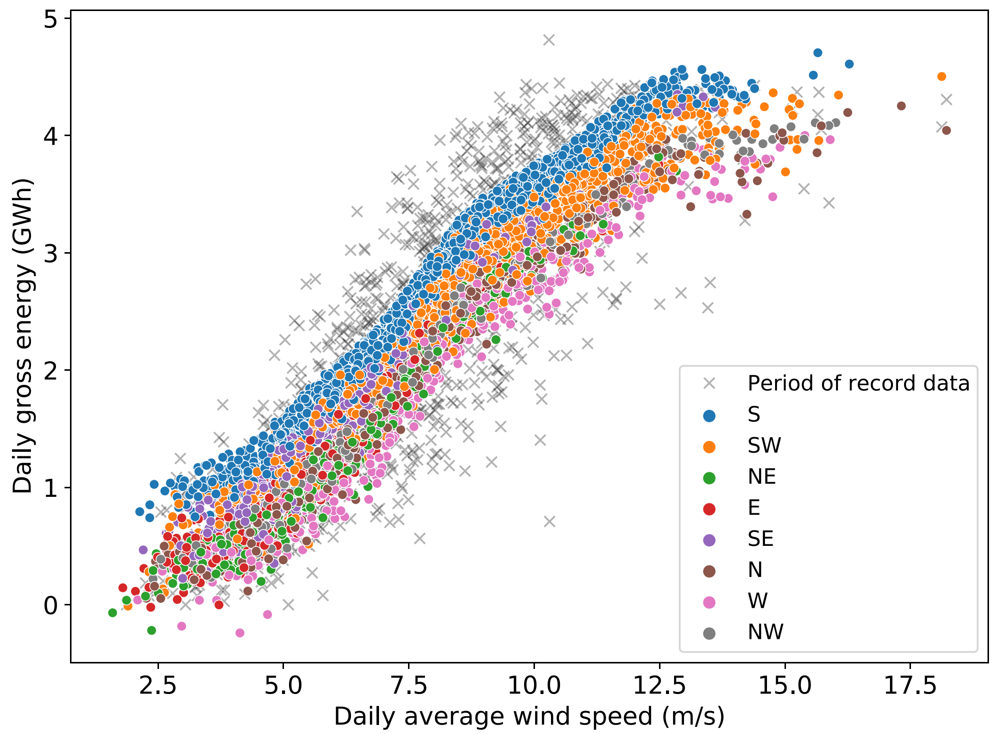

FIGURE 11 Wind plant power curve for project x41 obtained from a GAM regression at a daily resolution using wind speed, temperature, and (sine and cosine of) wind direction as inputs. Energy production data from the wind plant period of record are shown with gray crosses, and long-term predictions by circles are color-coded by wind direction bin

project x41. Data from the wind plant period of record are shown with black crosses, whereas long-term data are color-coded based on the wind direction sector.

When the regression model is fed with wind direction data, it is then able to clearly predict the impact of different wind direction regimes on the long-term wind plant power production. In fact, we see in this example how different wind direction regimes have a significant impact on the wind plant power curve, with southerly flow producing more energy than the other wind directions, wind speed being the same. This differentiation is consistent with the layout of the wind power plant, wherein the wind turbines experience significantly lower wake losses for southerly flow as a result of greater turbine spacing, and with the absence of other upwind plants directly south of x41. We conclude that the inclusion of wind direction as a regression input allows the wind plant power curve to adapt itself to the internal layout of the wind plant, to better represent the impact of wind turbine wakes. Therefore, we recommend using a multivariate regression whenever the data availability allows, even if this does not result in a direct reduction in operational analysis regression uncertainty as derived from a bootstrapping approach.

Finally, we quantify the generalization skills of the proposed multivariate nonlinear regression models at a daily and hourly resolution in terms of their accuracy when applied to wind plant operational data not considered in the machine-learning training phase. In fact, the error metrics in Tables 5 and 6 are calculated over the entire period of record, thus including the data on which the models were trained. To assess the ability of this model to generalize, we apply a nested cross-validation approach. For each wind plant, we divide the period of record in four (contiguous) parts. We train each model setup four times, each using randomized cross-validation on three fourths of the period of record, while the last fourth is kept as testing period. Then, we calculate the normalized RMSE for each trained algorithm (i.e., on each of the four different testing periods), and use their average as best estimate of the regression model generalization error. We find that, for all wind plants, at both considered time resolutions, and all the considered combinations of input variables, the generalization RMSE is no more than 11% larger than the values shown in Tables 5 and 6. Therefore, we can conclude that the proposed nonlinear approaches for wind plant power curves are capable of providing an accurate match with the real wind plant operational data.

4 | CONCLUSIONS

Accurately assessing the long-term operational energy production with a wind plant power curve is essential for wind plant operation and financial transactions. A 1% absolute reduction in PCYA uncertainty can amount to about $500,000 savings for a 50-MW wind power plant.[18]

Historically, a univariate linear regression at a monthly resolution has been used in the wind energy industry to relate density-corrected wind speed to wind plant power production. Here, we have supported the recommendations from the AWEA TR-1 working group and demonstrated the value in going beyond the current industry standard, and tested the benefits introduced when a multivariate machine-learning regression at a daily and hourly resolution is used for operational wind plant power curve assessment. To do so, we analyzed operational data from 10 wind plants in simple terrain, and quantified the change in regression uncertainty and regression normalized root-mean-square error when various augmentations of the standard operational analysis monthly linear regression are implemented. We used NREL's open-source OpenOA software package to run the analysis.

We discovered major benefits when adopting machine-learning, nonlinear regression models at a daily our hourly time resolution. When a generalized additive model at a daily resolution is used with density-corrected wind speed as the only input feature, all of the projects analyzed in this study experienced a significant reduction in regression uncertainty (up to a -1.0% absolute difference, corresponding to a -60% relative change) compared to the current industry standard approach. When an hourly resolution is used, the reduction in regression uncertainty further increases, up to a -1.2% absolute difference, corresponding to a -80% relative change. Increasing the number of input features used in the regression can also improve the accuracy of the PCYA. Even simple augmentations, such as adding temperature data to the operational analysis monthly linear regression, can reduce the regression uncertainty by up to 0.50% (-43% change relative to the univariate GAM at the same data frequency) and the regression NRMSE by up to 45% for projects that show a large seasonal variability. Adding temperature and wind direction as additional inputs to the learning model at a daily or hourly resolution further reduces the regression uncertainty (up to -10% relative change at a daily resolution, and up to -4% relative change at an hourly resolution), with reductions in regression NRMSE by up to 17% at a daily resolution and 10% at an hourly resolution. Finally, assessing the wind plant power curve at a daily or hourly resolution can also lead to a more accurate assessment of the average long-term energy production than an analysis at a monthly resolution, as it allows for a more targeted filtering of the periods with low-quality operational data.

Many opportunities exist to further test and expand the PCYA work proposed in here. First, the analysis could be replicated for wind power plants that are not located in simple terrain, and therefore consider how the results would vary in complex terrain or offshore. Moreover, different machine-learning algorithms could be considered to potentially further reduce the operational analysis regression uncertainty. In addition, additional variables should be tested to further assess the value in applying a multivariate regression. For example, turbulence regimes and atmospheric stability are known to have a strong impact on wind turbine and wind power plant production, and therefore would likely further strengthen the proposed operational analysis regression process. Since reanalysis products generally do not provide a direct quantification of turbulence intensity or similar variables, the use of mesoscale model based products should be considered as an alternative source of atmospheric data. Finally, future research should focus on assessing the impact of finer time resolution and multiple input features on potentially reducing the minimum length of the period of record required to achieve a given accuracy in the wind plant power curve predictions.

ACKNOWLEDGEMENTS

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

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

AUTHOR CONTRIBUTIONS

Mike Optis and Michael Jason Fields envisioned the analysis. Nicola Bodini implemented the augmented features in OpenOA, which were reviewed by Jordan Perr-Sauer and Eric Simley. Nicola Bodini ran the analysis for the 10 wind power plants, in close consultation with Mike Optis. Nicola Bodini wrote the manuscript, with major contributions from Mike Optis. All authors reviewed the manuscript.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/we.2645.

DATA AVAILABILITY STATEMENT

The wind power plant data used in this work are proprietary and cannot be shared with the public. The open-source software used to perform wind plant operational analysis is publicly available at https://github.com/NREL/OpenOA.

ORCID

Nicola Bodini https://orcid.org/0000-0002-2550-9853
Mike Optis https://orcid.org/0000-0001-5617-6134
M. Jason Fields https://orcid.org/0000-0002-8781-6138

REFERENCES

1. Commission IE. Wind turbines—part 12-1: Power performance measurements of electricity producing wind turbines. IEC 61400-12-1; 2017.
2. Commission IE. Wind turbines—part 12-2: Power performance of electricity-producing wind turbines based on nacelle anemometry. IEC 61400-12-2; 2013.
3. Brower M. *Wind resource assessment: a practical guide to developing a wind project*. Hoboken, New Jersey: John Wiley & Sons; 2012.
4. Bodini N, Optis M. Operational-based annual energy production uncertainty: are its components actually uncorrelated? *Wind Energy Sci*. 2020;5(4): 1435-1448.
5. Lindvall J, Hansson J, Undheim O, Vindteknikk J. Post-construction production assessment of wind farms. Tech. Rep. 2016:297, Energyforsk; 2016.
6. Khatab AM. Performance analysis of operating wind farms. *Master's thesis*: Department of Earth Sciences, Campus Gotland; 2017.
7. Cameron J. Post-construction yield analysis. European Wind Energy Association Technical Workshop; 2012.
8. Lunacek M, Fields MJ, Craig A, et al. Understanding biases in pre-construction estimates. *J Phys Conf Ser*. 2018;1037(6):62009.
9. Manwell JF, McGowan JG, Rogers AL. *Wind energy explained: theory, design and application*: John Wiley & Sons; 2010.
10. Klink K. Climatological mean and interannual variance of United States surface wind speed, direction and velocity 1. *Int J Climatol A J R Meteorol Soc*. 1999;19(5):471-488.
11. Monahan AH. The probability distribution of sea surface wind speeds. Part II: dataset intercomparison and seasonal variability. *J Clim*. 2006;19(4): 521-534.
12. Whiteman CD. *Mountain meteorology: fundamentals and applications*: Oxford University Press; 2000.
13. Barthelmie RJ, Grisogono B, Pryor SC. Observations and simulations of diurnal cycles of near-surface wind speeds over land and sea. *J Geophys Res Atmos*. 1996;101(D16):21,327-21,337.
14. Suomalainen K, Silva CA, Ferrão P, Connors S. Synthetic wind speed scenarios including diurnal effects: Implications for wind power dimensioning. *Energy*. 2012;37(1):41-50.
15. Kiviluoma J, Holttinen H, Weir D, et al. Variability in large-scale wind power generation. *Wind Energy*. 2016;19(9):1649-1665.
16. Callaway DS. Sequential reliability forecasting for wind energy: temperature dependence and probability distributions. *IEEE Trans Energy Convers*. 2010;25(2):577-585.
17. Optis M, Perr-Sauer J. The importance of atmospheric turbulence and stability in machine-learning models of wind farm power production. *Renew Sustain Energy Rev*. 2019;112:27-41.
18. Halberg E. A monetary comparison of remote sensing and tall towers. AWEA Wind Resource Workshop 2017 Salt Lake City; 2017.
19. Global Wind Energy Council. Global wind report—annual market update 2017, Global Wind Energy Council; 2018.
20. Giebel G, Brownsword R, Kariniotakis G, Denhard M, Draxl C. The state-of-the-art in short-term prediction of wind power: a literature overview. ANEMOS. plus; 2011.
21. Demolli H, Dokuz AS, Ecemis A, Gokcek M. Wind power forecasting based on daily wind speed data using machine learning algorithms. *Energy Convers Manag*. 2019;198:111823.
22. Clifton A, Kilcher L, Lundquist JK, Fleming P. Using machine learning to predict wind turbine power output. *Environ Res Lett*. 2013;8(2):24009.
23. Leahy K, Hu RL, Konstantakopoulos IC, Spanos CJ, Agogino AM. Diagnosing wind turbine faults using machine learning techniques applied to operational data. *2016 IEEE International Conference on Prognostics and Health Management (ICPHM)*: IEEE; 2016:1-8.
24. Arcos Jiménez A, Gómez Muñoz C, García Márquez F. Machine learning for wind turbine blades maintenance management. *Energ*. 2018;11(1):13.
25. Wan Y-H, Ela E, Orwig K. Development of an equivalent wind plant power-curve, Golden, CO (United States), National Renewable Energy Lab. (NREL); 2010.
26. Marvuglia A, Messineo A. Learning a wind farm power curve with a data-driven approach. *No. 057*: Linköping University Electronic Press; 2011: 4217-4224.
27. Marvuglia A, Messineo A. Monitoring of wind farms' power curves using machine learning techniques. *Appl Energy*. 2012;98:574-583.
28. Velázquez Medina S, Carta JA, Portero Ajenjo U. Performance sensitivity of a wind farm power curve model to different signals of the input layer of anns: case studies in the Canary Islands. *Complexity*. 2019;2019:2869149.
29. Fields MJ, Optis M, Lee CYJ, et al. Wind plant performance prediction (WP3) benchmark introduction; 2019.
30. Hersbach H, Bell B, Berrisford P, et al. The ERA5 global reanalysis. *Q J R Meteorol Soc*. 2020;146(730):1999-2049.
31. Ramon J, Lledo L, Torralba V, Soret A, Doblas-Reyes FJ. What global reanalysis best represents near-surface winds? *Q J R Meteorol Soc*. 2019;145 (724):3236-3251.
32. Gelaro R, McCarty W, Suárez MJ, et al. The Modern-Era Retrospective Analysis for Research and Applications, version 2 (merra-2). *J Clim*. 2017;30 (14):5419-5454.
33. Saha S, Moorthi S, Wu X, et al. The NCEP climate forecast system version 2. *J Clim*. 2014;27(6):2185-2208.
34. Dee DP, Uppala SM, Simmons AJ, et al. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q J R Meteorol Soc*. 2011;137(656):553-597.
35. Optis M, Perr-Sauer J, Philips C, et al. Openoa: an open-source code base for operational analysis of wind power plants. *Wind Energy Sci Discuss*. 2019:1-14.
36. Perr-Sauer J, Optis M, Fields JM, et al. Openoa: an open-source codebase for operational analysis of wind farms. *J Open Source Software*. 2021;6(58): 2171.
37. Hastie TJ, Tibshirani RJ. *Generalized additive models*, Vol. 43: CRC press; 1990.
38. Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal*. 2002;38(4):367-378.
39. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn*. 2006;63(1):3-42.

40. Craig A, Optis M, Fields MJ, Moriarty P. *Uncertainty quantification in the analyses of operational wind power plant performance*, Vol. 1037. The Science of Making Torque from Wind (TORQUE 2018), Milan, Italy. IOP Publishing; 2018;52021.
41. Assessment of site specific wind conditions for wind power stations. standard, International Electrotechnical Commission; in draft.
42. JCGM 100:2008. Evaluation of measurement data - guide to the expression of uncertainty in measurement. Joint Committee for Guides in Metrology; 2008.
43. Friedman J, Hastie T, Tibshirani R. *The elements of statistical learning*, Vol. 1: Springer series in statistics New York; 2001.
44. Khalid M, Savkin AV. *Direction dependent power curves for wind power prediction: A case study*: Springer; 2011;121-127.

How to cite this article: Bodini N, Optis M, Perr-Sauer J, Simley E, Fields MJ. Lowering post-construction yield assessment uncertainty through better wind plant power curves. *Wind Energy*. 2022;25(1):5–22. https://doi.org/10.1002/we.2645