Operational-Based Annual Energy Production Uncertainty: Are its Components Actually Uncorrelated?

Nicola Bodini [1] and Mike Optis [1]

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

Correspondence: Mike Optis (mike.optis@nrel.gov)

Abstract. Calculations of annual energy production (AEP) from a wind farm—whether based on pre-construction or operational data—are critical for wind farm financial transactions. The uncertainty in the AEP calculation is especially important in quantifying risk and is a key factor in determining financing terms. A popular industry practice is to assume that different uncertainty components within an AEP calculation are uncorrelated, and can therefore be combined as the sum of their squares. We assess the practical validity of this assumption for operational-based uncertainty, which is comprised of components associated with long-term correction and measurements, by performing operational AEP estimates for over 470 wind farms in the United States. We contrast the uncorrelated sum-of-squares method with a Monte Carlo approach, in which no assumptions of correlation between uncertainty components are made. Results show that several component pairs exhibit weak to moderate correlations: inter-annual variability and the linearized long-term correction (positive correlation); wind resource inter-annual variability and linear regression (negative); and reference wind speed uncertainty and linear regression (positive). The sources of these correlations are described and illustrated in detail in this paper, and the effect on the total AEP uncertainty calculation is investigated. Based on these results, we conclude that a Monte Carlo approach to operational AEP uncertainty quantification is more robust and accurate than the simple approach which neglects correlations between uncertainty components.

1 Introduction

Calculations of wind farm annual energy production (AEP)—whether based on pre-construction data before a wind power plant is built or on operational data after a wind farm has started its operations—are vital for wind farm financial transactions. Pre-construction estimates of AEP are needed to secure and set the terms for new project financing, whereas operational estimates of long-term AEP are required for important wind farm transactions, such as refinancing, purchasing/selling, and mergers/acquisitions. The need for AEP analyses of wind farms is increasing, as global wind capacity increased to 539 GW in 2017, representing 11% and 91% increases over 1-year and 5-year periods, respectively; and capacity is expected to increase by another 56% to 841 GW by 2022 (Global Wind Energy Council, 2018). In the United States, wind farms generated over 300,000 GWh in 2019, about 7.5% of the total US electricity generation from utility-scale facilities that year, with a 50% increase over a 6-year period (Energy Information Administration, 2020).

This rapid growth of the wind energy industry is putting an increased spotlight on the accuracy and consistency of AEP calculations. For pre-construction AEP estimates, there has been considerable movement towards standardization. The International Energy Commission (IEC) is currently developing a standard (IEC 61400-15:draft), and there have long been guidance and best practices available (Brower, 2012). By contrast, long-term operational AEP estimates do not have such extensive guidance or standards. Only limited standards covering operational analyses exist; IEC 61400-12-1:2017 addresses turbine power curve testing, and IEC 61400-26-3:2016 addresses the derivation and categorization of availability loss metrics. There are to our knowledge, however, no standards and very limited published guidance on calculating long-term AEP from operational data. Rather, documentation seems to be limited to a consultant report (Lindvall et al., 2016), an academic thesis (Khatab, 2017), and limited conference proceedings (Cameron, 2012; Lunacek et al., 2018).

Documentation and standards for pre-construction AEP methods are of limited use for operational-based AEP methods, given the many differences between the two approaches. In general, operational AEP calculations are simpler than pre-construction estimates because actual measurements of wind farm power production at the revenue meter replace the complicated pre-construction estimate process (e.g., meteorological measurements, wind and wake-flow modeling, turbine performance, estimates of wind farm losses). However, the two methods do share several similarities, including regression relationships between on-site measurements and a long-term wind speed reference, the associated long-term (windiness) correction applied to the on-site measurements, and estimates of uncertainty in the resulting AEP calculation. The uncertainty components for operational AEP calculations are simplified relative to those in a pre-construction estimate (IEC 61400-15:draft); shared components between the two methods are listed in Table 1.

The uncertainty values from each component listed in Table 1 must be combined to produce a total estimate of AEP uncertainty. We found no guidance in the literature for combining uncertainty components in an operational AEP estimate. However, considerable guidance exists for combining pre-construction uncertainties (Lackner et al., 2007; Brower, 2012; Vaisala, 2014; Kalkan, 2015; Clifton et al., 2016). In every case, recommended best practices assume that all uncertainties, σ_i, are uncorrelated and can therefore be combined using a sum of squares approach to give the total AEP uncertainty, $\sigma_{\text{tot,uncorr}}$:

$$\sigma_{\text{tot,uncorr}} = \sqrt{\sum_i \sigma_{\text{i}}^2} \tag{1}$$

To better understand how uncertainties are combined in long-term operational AEP calculations, we reached out to several wind energy consultants who regularly perform these analyses. These conversations revealed that uncertainties in a long-term operational AEP calculation are also assumed uncorrelated and combined using Equation 1.

Uncertainty component	Description
On-site measurements	Measurement error in met mast wind speeds (pre-construction) or power at the revenue meter (operational)
Reference wind speed data	Measurement or modeling error in long-term reference measured or modeled wind speed data
Losses	Error in estimated or reported availability and curtailment losses
Regression	Sensitivity in the regression relationship between on-site measurements and reference wind speeds
Long-term (windiness) correction	Sensitivity in the long-term correction applied to the regression relationship between on-site measurements and reference wind speeds
Inter-annual variability of resource	Sensitivity in future energy production because of resource variability

Table 1. Main Sources of Uncertainty in a Long-Term Operational AEP Estimate.

1.1 Goal of Study

The purpose of this study is to examine the extent to which the assumption of uncorrelated uncertainties—and therefore the combination of those uncertainties through a sum of squares approach—is accurate and appropriate for operational AEP calculations. Specifically, this study aims to identify potential correlations between AEP uncertainty components and propose a Monte Carlo approach to capture such correlations when combining individual uncertainty components. Monte Carlo methods have been used in different applications for uncertainty quantification within the wind energy industry, ranging from the prediction of extreme wind speed events (Ishihara and Yamaguchi, 2015), to offshore fatigue design (Müller and Cheng, 2018), to economic analysis of the benefits of wind energy projects (Williams et al., 2008). Here, the focus is on operational AEP uncertainty, using publicly available wind farm operational data. While in the analysis we focus on operational AEP calculation, we expect that the results from this analysis—namely the potential identification of correlated uncertainty components—can be equally relevant for informing and improving pre-construction AEP methods.

In Section 2, we first describe the data sources used in this analysis—namely wind farm operational data and reanalysis products—as well as the Monte Carlo approach to calculate operational AEP and quantify its uncertainty. Section 3 presents the main results of our analysis, in terms of uncertainty contributions and correlation among the different components. We conclude and suggest future work in Section 4.

2 Data and Methods

2.1 Wind Farm Operational Data and Reanalysis Products

Operational wind farm energy production data for this analysis are obtained from the publicly available Energy Information Administration (EIA) 923 database (EIA, 2018). This database provides reporting of monthly net energy production from all power plants in the United States, including wind farms. A total of over 670 unique wind farms are available from this data set.

Long-term wind speed data (needed to perform the long-term or windiness correction in an AEP estimate) are used from three reanalysis products over the period of January 1997 through December 2017:

- The Modern-Era Retrospective analysis for Research and Applications v2 (MERRA-2) (Gelaro et al., 2017). We specifically use the M2T1NXSLV data product, which provides diagnostic wind speed at 50 m above ground level (AGL), interpolated from the lowest model level output (on average about 32 m AGL), using Monin Obukhov similarity theory. Data are provided at an hourly time resolution.

- The European Reanalysis Interim (ERA-interim) data set (Dee et al., 2011). We specifically use output at the 58th model level, which on average corresponds to a height of about 72 m AGL. Data are provided at 6-hourly time resolution.

- The National Centers for Environmental Prediction v2 (NCEP-2) data set (Saha et al., 2014). We specifically use diagnostic wind speed data at 10 m AGL. Data are provided at a 6-hourly time resolution.

The wind speed data are density-corrected at their native time resolutions to correlate more strongly with wind farm power production (i.e., higher density air in winter produces more power than lower density air in summer, wind speed being the same):

$$U_{\text{dens,corr}} = U \left(\frac{\rho}{\rho_{\text{mean}}} \right)^{1/3} \tag{2}$$

where $U_{\text{dens,corr}}$ is the density-corrected wind speed, U is the wind speed, ρ is air density (calculated at the same height as wind speed), ρ_{mean} is the mean density over the entire period of record of the reanalysis product, and the exponent $1/3$ is derived from the basic relationship between wind power and wind speed cubed (Manwell et al., 2010). To calculate air density at the same height as wind speed, we first extrapolate the reported surface pressure to the wind speed measurement height, assuming hydrostatic equilibrium:

$$p = p_{\text{surf}} \exp\left[\frac{gz}{RT_{\text{avg}}} \right] \tag{3}$$

where p is the pressure at the wind speed measurement height, p_{surf} is the surface pressure, g is the acceleration caused by gravity, z is the wind speed measurement height, R is the gas constant, and T_{avg} is the average temperature between the reported value at 2 m AGL and at the wind speed measurement height.

To lessen the impact of limited and/or poor-quality data on the results of our analysis, we filter for wind farms with at least 8 months of data and with a moderate-to-strong correlation with all three reanalysis products ($R^2 > 0.6$). A threshold of 8

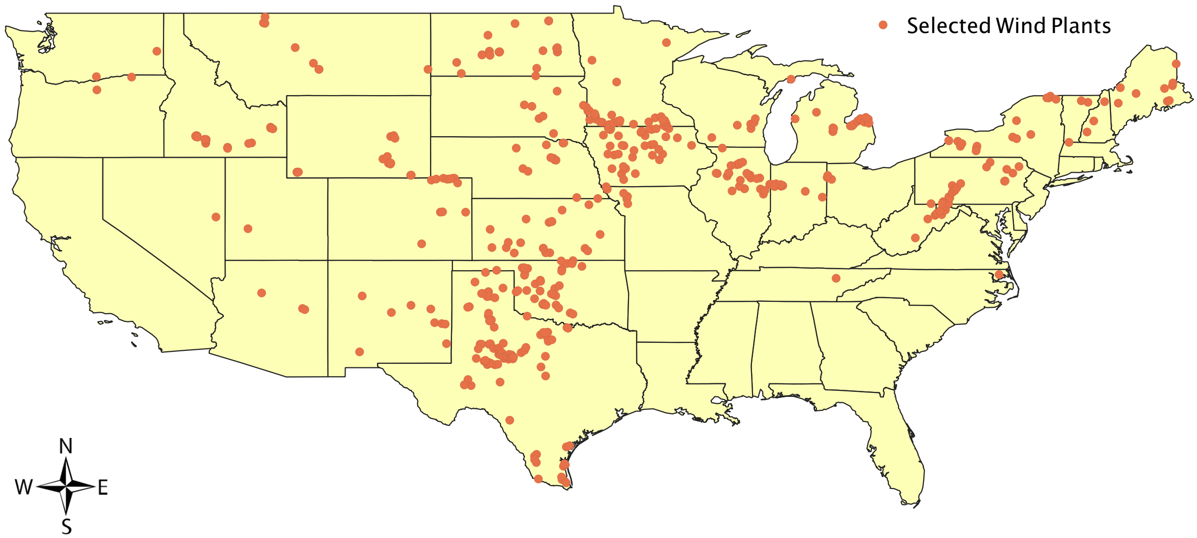

Figure 1. Map of the 472 wind farms that were considered in this study.

months is selected in order to investigate uncertainty as it relates to a low number of data points but not so low as to make the use of a regression relationship questionable. A total of 472 wind farms are kept for the final analysis, and their locations are shown in Figure 1. Because obtaining an accurate representation of wind data in complex terrain by reanalysis products is challenging (Shravan Kumar and Anandan, 2009), most of the selected wind plants are located in the Midwest and Southern Plains. Notably, no wind farms in California pass the filtering criteria, because they are predominately located in areas with thermally driven wind regimes such as Tehachapi Pass, where coarse-resolution reanalysis products are poor predictors of wind energy production.

The fundamental step in an AEP calculation involves a regression between wind speed (here, from the reanalysis products) and energy production (here, from the EIA 923 database). To investigate whether a simple linear function can be assumed to express the relationship between wind speed and wind farm energy production when considering monthly data, we show a scatterplot between MERRA-2 monthly wind speed and monthly energy production across all 472 sites in Figure 2. For each site, data have been normalized by the respective site mean. We show best-fits using a linear, quadratic, and cubic function, and calculate the mean absolute error of each fit. We find that the difference between the normalized MAE values from the considered functions is less than 0.7%. Therefore, the uncertainty connected with the choice of using a linear regression in the operational AEP methodology at monthly time resolution appears minimal. Moreover, through conversations with wind industry professionals, we found that a linear regression based on monthly data is the standard industry approach when performing bankable[1] operational AEP analyses.

[1] Results are accepted by banks, investors, and so on for use in financing, buying/selling, and acquiring wind farms.

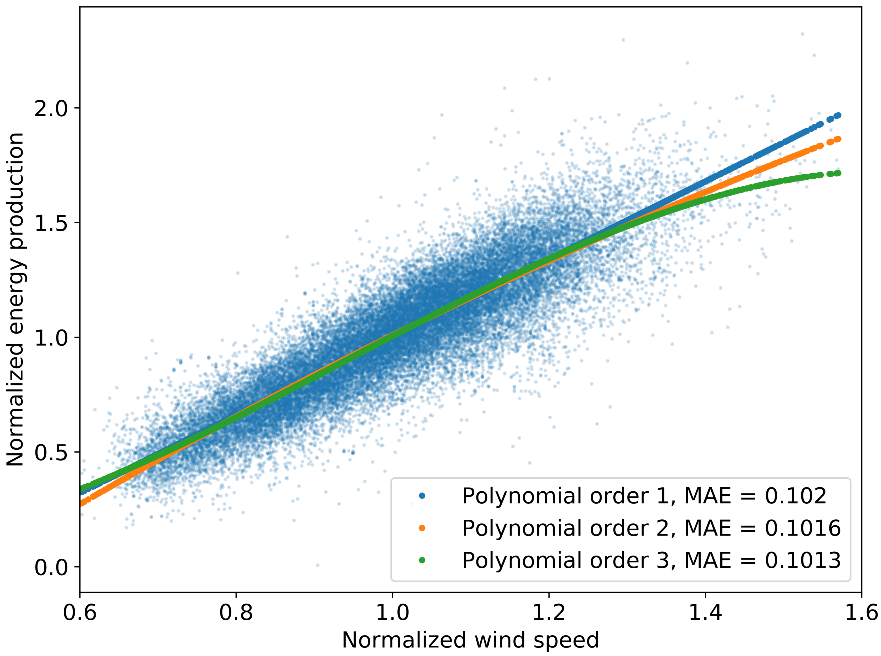

Figure 2. Scatterplot between normalized MERRA-2 monthly wind speed and monthly energy production across all 472 selected sites, and linear, quadratic and cubic best-fit lines.

2.2 Operational AEP Methodology

Given the lack of existing guidelines for a standard approach for operational AEP calculations, we instead base our methodology on conversations with several wind energy consultants. These conversations overwhelmingly revealed the following characteristics of an industry standard and bankable operational AEP analysis:

1. Wind speed data (measured or modeled) are density-corrected at their native time resolution, using equation 2.
2. Monthly revenue meter data, monthly average availability and curtailment losses, and monthly average wind speeds from a long-term wind resource product are calculated.
3. Monthly revenue meter data are normalized to 30-day months (e.g. for January, the revenue meter values are multiplied by 30/31).
4. Monthly revenue meter data are corrected for monthly availability and curtailment (i.e., monthly gross energy data are calculated).
5. A linear regression between monthly gross energy production and concurrent monthly average wind speeds is performed.

6. Long-term monthly average wind speed is then calculated for each calendar month (i.e., average January wind speed, average February wind speed, and so forth) with a hindcast approach, using 10–20 years of the available long-term reference monthly wind resource data (reanalysis products, long-term reference measurements, ...).

7. Slope and intercept values from the regression relationship are then applied to the long-term monthly average wind speed data, with the long-term or so-called windiness correction. A long-term data set of monthly (January, February, ...) gross energy production is obtained.

8. The resulting long-term monthly gross energy estimates, which are based on 30-day months, are then denormalized to the actual number of days in each calendar month (e.g. for January, the obtained value is multiplied by 31/30).

9. Long-term estimates of availability and curtailment losses are finally applied to the denormalized long-term monthly gross energy data, leading to a long-term calculation of operational AEP.

In the EIA-923 database, availability and curtailment data are not available. Therefore, in our analysis we omit steps 4 and 9 of the list, and only perform calculations on net energy data. A diagram outlining the resulting general process of the operational AEP analysis adopted in our study is shown in Figure 3.

2.3 Monte Carlo Analysis

To quantify the uncertainty of the long-term operational AEP estimate obtained using the methodology described in the previous section, we implement a Monte Carlo approach. In general, a Monte Carlo method involves the randomized sampling of inputs to or calculations within a method which, when repeated many times, results in a distribution of possible outcomes from which uncertainty can be deduced, usually calculated as the standard deviation or the coefficient of variation of the resulting distribution (ISO and OIML, 1995; Dimitrov et al., 2018). Here, we apply this approach to derive a distribution of long-term operational AEP values, from which its uncertainty can be calculated. To do so, we consider and include in the Monte Carlo approach five operational-based uncertainty components, so that five different samplings are performed at each Monte Carlo iteration. The following uncertainty components are included in our proposed Monte Carlo methodology for long-term operational AEP:

– Revenue meter measurement error. We incorporate this uncertainty component in the Monte Carlo simulation by sampling monthly revenue meter data from a normal distribution centered on the reported value, and 0.5% standard deviation. In fact, a value of 0.5% is coherent with what is typically assumed in the wind energy community as revenue meter uncertainty (IEC 60688:2012; ANSI C12.1-2014).

– Reference wind speed data modeling error. Quantifying the uncertainty of the long-term wind resource data used in the operational AEP assessment is challenging, as it can vary based on the location, long-term wind speed product used, or instrument from which reference observations are taken. To include this uncertainty component in a systematic way across the 472 locations considered in our analysis, we incorporate it in the Monte Carlo simulation by randomly selecting, at each iteration at each site, wind resource data from one of the three considered reanalysis products.

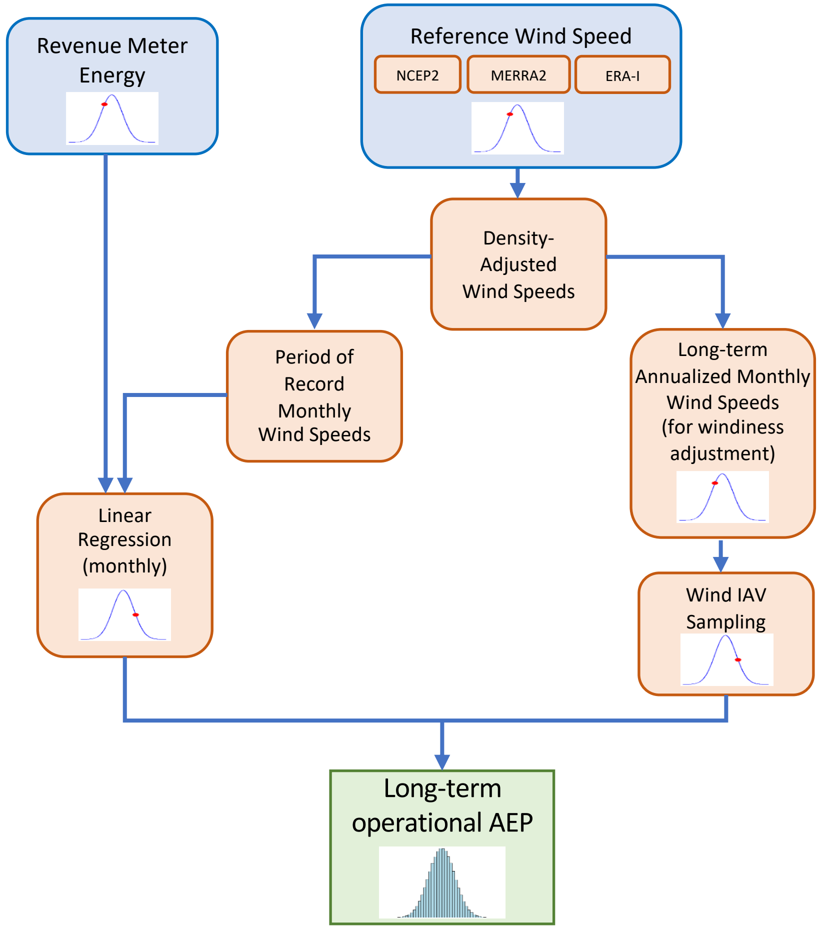

Figure 3. Long-term annual energy production (AEP) estimation process using operational data under a Monte Carlo approach; sources of uncertainty and points of Monte Carlo sampling are denoted by probability distribution images. Note: IAV denotes inter-annual variability.

- Linear regression model uncertainty. This component is incorporated in the Monte Carlo method by sampling the regression slope and intercept values from a multivariate normal distribution centered on their best-fit values and covariance matrix equal to the one of the best-fit parameters. The diagonal terms in the covariance matrix are given by the square of the slope and intercept standard errors. For a regression model between an independent variable x and a dependent variable y the standard error of the regression is defined as

$$e_y = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}, \tag{4}$$

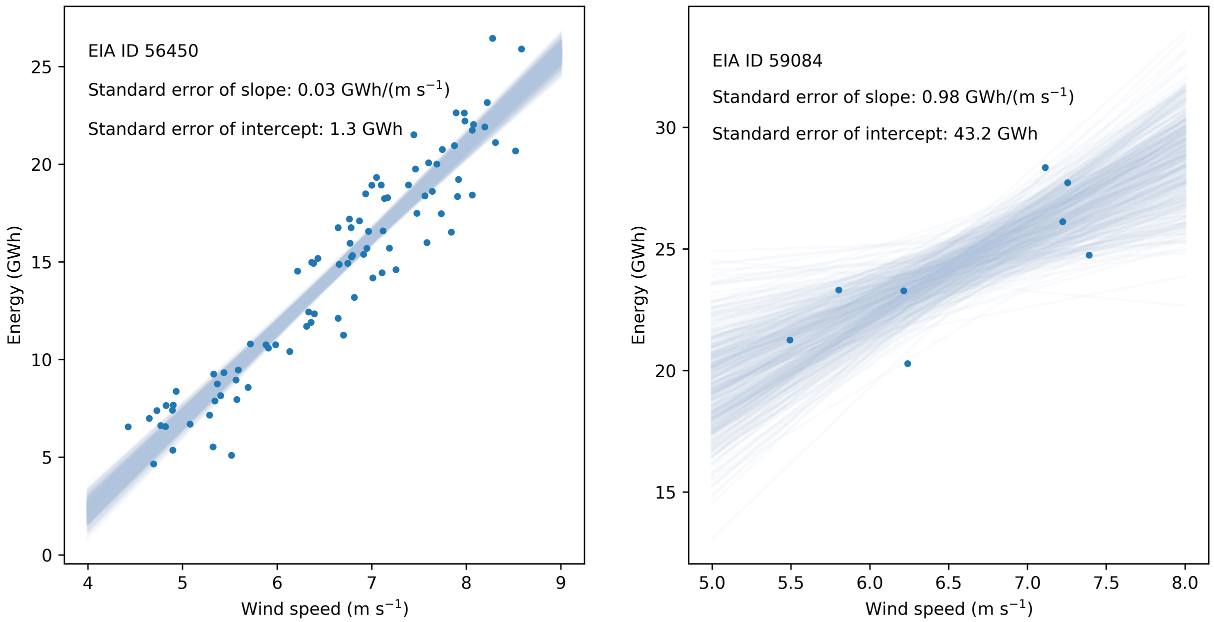

Figure 4. Sampling set of regression lines corresponding to the slope and intercept values derived from their standard errors in the Monte Carlo approach, for two stations in the EIA data set.

where $\hat{y_i}$ is the regression-predicted value for y_i, and n is the number of data points used in the regression. The standard error of the regression slope:

$$e_a = \frac{e_y}{\sum (x_i - \overline{x_i})^2}, \tag{5}$$

and the standard error of the intercept:

$$e_b = e_y\, e_a \sqrt{\frac{\sum x_i^2}{n}}. \tag{6}$$

e_a^2 and e_b^2 are the diagonal terms in the covariance matrix of the multivariate normal distribution of regression slope and intercept, from which Monte Carlo values are drawn. Slope and intercept values are strongly negatively correlated, which is captured by their covariance when performing the linear regression. The off-diagonal terms in the covariance matrix of the multivariate normal distribution constrain the random sampling of slope and intercept values, to avoid sampling unrealistic combinations. An example of this sampling is shown in Figure 4 for two projects of different regression strengths. We sample 500 slope and intercept values from a multivariate normal distribution centered around the best-fit parameters, and with covariance matrix derived from the standard errors of slope and intercept and their covariance. As shown in the Figure, the low standard errors found for the leftmost regression relationship constrain the possible slope and intercept values that can be sampled while the high standard errors in the rightmost regression relationship allow for a much wider sampling.

– Long-term (windiness) correction uncertainty. We incorporate this component by sampling the number of years (between 10 and 20) to use as the long-term wind resource data to which the regression coefficients are applied to derive long-term energy production data (the so-called windiness correction).

– Wind resource inter-annual variability (IAV) uncertainty. We incorporate this uncertainty component in the Monte Carlo method by sampling the long-term (reanalysis) average calendar monthly wind speeds (i.e., average January, average February) used to calculate long-term monthly energy production data. The sampling distribution is normal, centered on the calculated long-term average calendar monthly wind speed, and with a standard deviation equal to the 20-year standard deviation of the long-term average monthly wind speed for each calendar month.

Each of the listed sources of uncertainty corresponds to a Monte Carlo sampling, and is highlighted by a probability distribution in the flowchart in Figure 3. Note that uncertainty components related to availability and curtailment losses are not considered in our approach because the EIA 923 database does not include measurements of these losses.

For each wind farm, we estimate the total operational AEP uncertainty by running a Monte Carlo simulation 10,000 times. At each iteration, all five samplings, corresponding to the five considered uncertainty components (revenue meter, reference wind speed data, wind resource IAV, linear regression, and windiness correction), are simultaneously performed. The total uncertainty in long-term operational AEP is then estimated as the coefficient of variation of its resulting distribution. Convergence of the AEP distribution within 0.5% of the true mean after the 10,000 Monte Carlo runs was verified for all projects, with a 95% confidence.

To understand the impact of the single uncertainty components and study their correlation, we also run, at each site, the Monte Carlo simulation with only a single sampling performed (i.e. either revenue meter, reference wind speed data, IAV, linear regression, or windiness correction). At each wind farm, we run the Monte Carlo simulation 10,000 time for each of the five single operational uncertainty components considered. We quantify the impact of each single uncertainty component on the long-term operational AEP in terms of the coefficient of variation of the distribution of operational AEP resulting from the Monte Carlo simulation run when sampling only that single uncertainty component.

The code used to perform the AEP calculations is published and documented in NREL's open-source operational assessment software, OpenOA.[2] Calculations were performed on Eagle, NREL's high-performance computing cluster. Specifically, each wind farm was assigned a different processor and run in parallel. Given the general simplicity of the AEP method used here, computational requirements were moderate despite the 60,000 simulations (10,000 runs x 6 uncertainty setups) required for each wind farm.

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

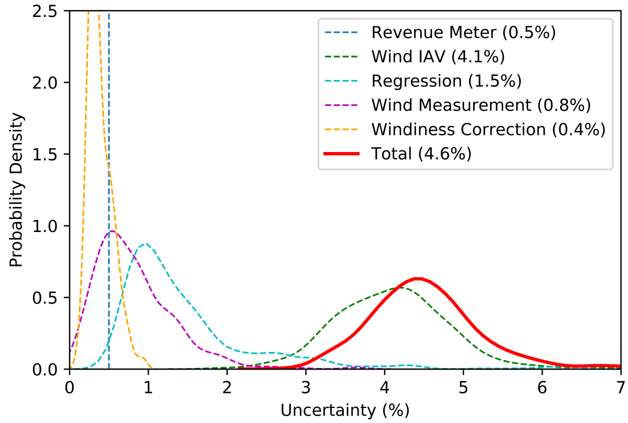

Figure 5. Operational-based AEP uncertainty distributions across projects for the different uncertainty components; mean values across projects are shown in the legend. Uncertainty values are quantified as the percent coefficient of variation of the long-term operational AEP distribution. Note that the sum of squares of the average values of the single components does not add up to the average of the total uncertainty.

3 Results

3.1 Operational-Based AEP Uncertainty Contributions

The application of the different setups of the Monte Carlo approach first allows for an assessment of the distributions of the total operational-based AEP uncertainty and of its single components across the 472 wind farms, expressed in terms of the percent coefficient of variation of the resulting AEP distributions (Figure 5). Uncertainty connected to wind resource IAV is found to contribute the most (average 4.1% across all wind farms). The uncertainty in the linear regression model has the second largest contribution (1.5%), followed by the uncertainty of the reference wind speed data (0.8%; here, of the reanalysis products), and revenue meter data (here, imposed at 0.5%). The long-term windiness correction has the smallest uncertainty component (0.4%). Therefore, the number of years used for the long-term windiness correction does not have a large impact on the overall uncertainty in operational AEP, at least for the sampled range of 10–20 years. Using as few as 10 years seems sufficient to give stability to the long-term AEP estimate, and adding additional years does not provide a significant reduction in the uncertainty connected with the long-term estimate.

The proposed Monte Carlo approach does not require any assumption on the correlation between the different uncertainty components; on the other hand, the conventional sum of squares approach assumes the uncertainty components are all uncorrelated. Therefore, we compare the total operational AEP uncertainty from the Monte Carlo method with all the five simultaneous samplings ($\mathrm{CoV}_{\mathrm{Monte\ Carlo}}$) with the total uncertainty calculated using the conventional sum of squares approach ($\mathrm{CoV}_{\mathrm{uncorrelated}}$). For the latter approach, we quantify each of the five uncertainty components as the coefficient of variation of the corresponding operational AEP distribution obtained by running the Monte Carlo simulation with a single sampling performed. We then combine the five uncertainty components into the overall AEP uncertainty using Eq. 1. Figure 6 shows the results of this com-

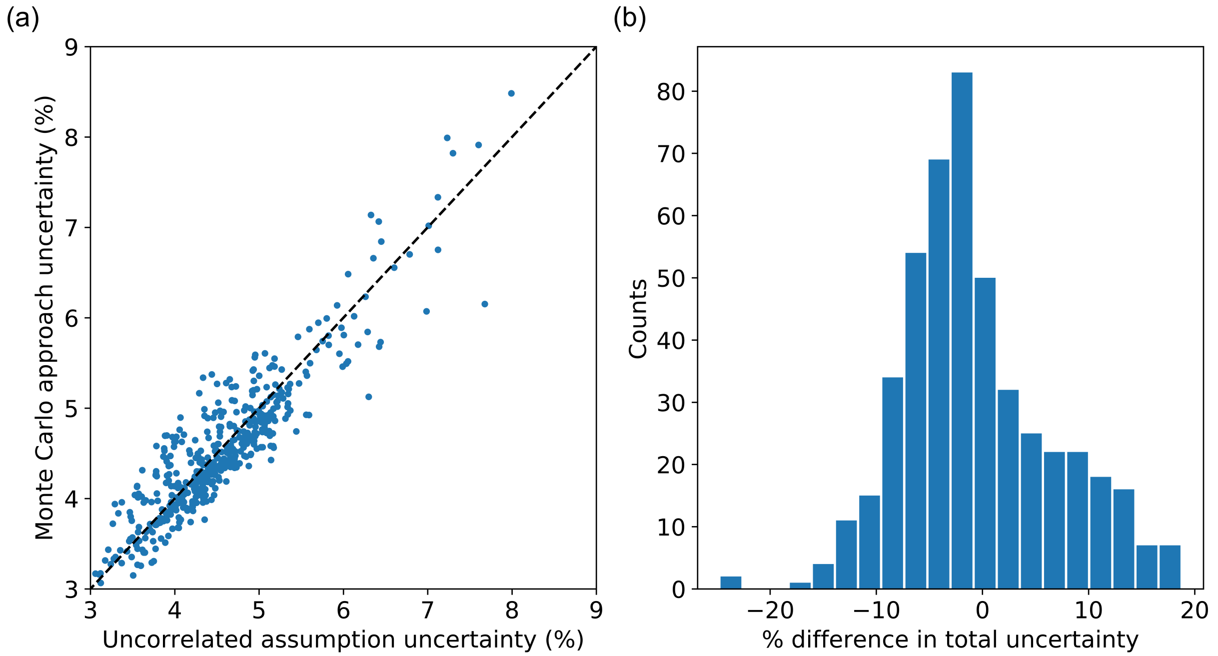

Figure 6. (a) Scatterplot of total operational AEP uncertainty values calculated with the proposed Monte Carlo approach and assuming uncorrelated uncertainty components for the 472 wind farms considered. Uncertainty is quantified as the percent coefficient of variation of the resulting long-term AEP distribution. (b) Histogram of percentage differences (Eq. 7) between the AEP uncertainties calculated using the two different approaches.

parison for the 472 wind farms considered, as a scatterplot and also as a histogram of the percentage difference between the two versions of the total AEP uncertainty:

$$\Delta_{\mathrm{CoV}} = \frac{\mathrm{CoV}_{\text{Monte Carlo}} - \mathrm{CoV}_{\text{uncorrelated}}}{0.5 \cdot (\mathrm{CoV}_{\text{Monte Carlo}} + \mathrm{CoV}_{\text{uncorrelated}})} \cdot 100 \tag{7}$$

A weak bias can be observed, with a median value of -2% in uncertainty percentage difference (which corresponds to a -0.25% median difference in the actual total uncertainty value). In other words, if correlations between the different uncertainty components are allowed and taken into account in the calculation method, the whole AEP uncertainty is then, on average, slightly reduced. This difference can be explained by considering that the two biggest sources of uncertainty (linear regression and IAV) are slightly negatively correlated (as will be shown in detail in the next section), thus making the Monte-Carlo-based total uncertainty lower, on average, than the one derived with the uncorrelated assumption. Moreover, assuming that all the uncertainty components are uncorrelated can introduce significant errors in the assessment of the AEP uncertainty for the single projects, with about 47% (16%) of the considered wind farms showing a $\pm 5\%$ (10%) uncertainty difference compared to the values from the Monte-Carlo-based approach. The mean absolute error of the distribution of uncertainty percentage differences is approximately 6% (Figure shown in the Supplement).

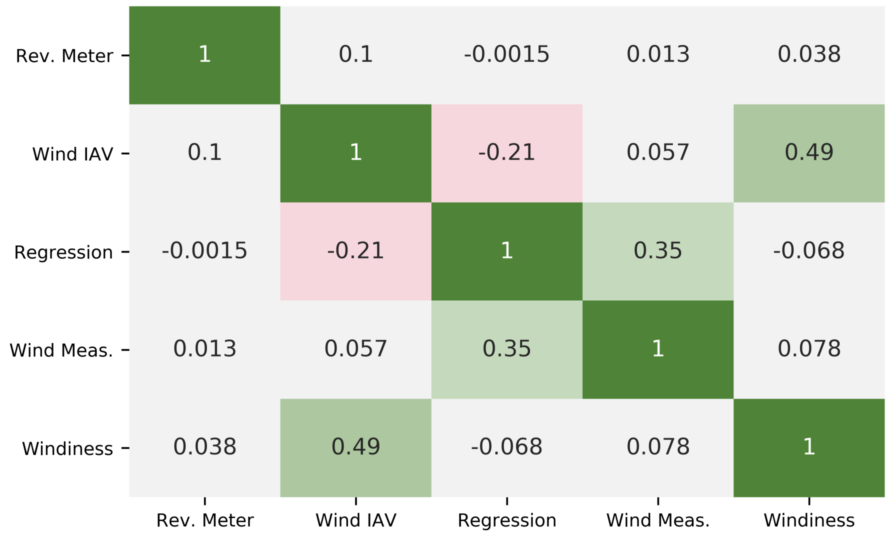

Figure 7. Correlation coefficient heat map between operational AEP uncertainty components, as calculated from the results of the Monte Carlo approach applied at the 472 wind farms considered in the analysis. Note: “Rev.” denotes “Revenue".

3.2 Correlation Between Operational-Based AEP Uncertainty Components

Because operational AEP uncertainty calculated by assuming a lack of correlation among its different components can greatly differ from the uncertainty values obtained when allowing for potential correlations, it is worth exploring the correlation between uncertainty components which are responsible for this difference. We leverage the results of the Monte Carlo analysis at the 472 wind farms considered to reveal the correlation between the single operational AEP uncertainty components, in terms of their Pearson correlation coefficient. As a result, we obtain the average correlation matrix in Figure 7. To assess which of the obtained correlations have statistical significance, we calculate the $p-$value (Westfall and Young, 1993) associated with the ten correlation coefficients. The test reveals that for three pairs of uncertainty components the probability of finding the *observed* not-zero correlation coefficients if the *actual* correlation coefficient were in fact zero ($p-$value) is less than 10^{-5}. Therefore, the following three correlations have strong statistical significance:

- The wind resource IAV and the long-term windiness correction uncertainties are moderately correlated ($R = 0.49$, $p = 1.9 \cdot 10^{-29}$).
- The linear regression and reference wind speed data uncertainties are weakly correlated ($R = 0.35$, $p = 2.5 \cdot 10^{-15}$).
- The wind resource IAV and the linear regression uncertainties appear weakly negatively correlated ($R = -0.21$, $p = 2.6 \cdot 10^{-6}$).

The first correlation noted earlier (wind resource IAV and long-term windiness correction) is explained simply by the fact that both uncertainty components are driven by wind resource variability. At a site with large wind variability, IAV will be large

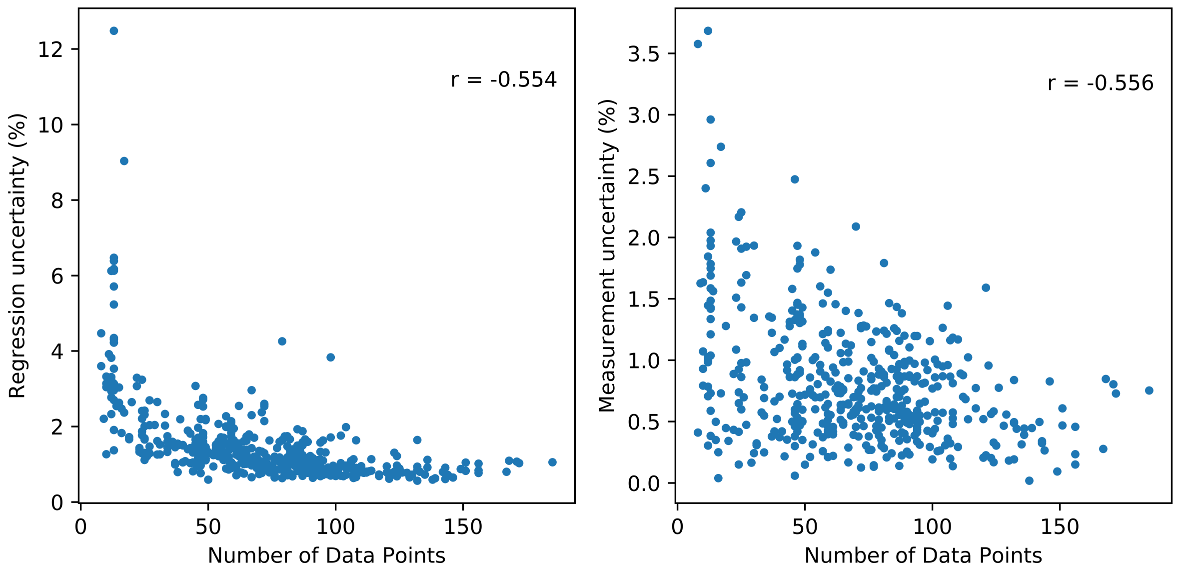

Figure 8. Dependence of linear regression uncertainty and reference wind speed data uncertainty on the number of data points in the period of record, for the 471 projects considered in the analysis.

by definition, and so will the uncertainty introduced by different lengths of time series used for the long-term AEP calculation.

The correlation between linear regression and reference wind speed data uncertainties can be justified given the dependence of both these uncertainty components on the number of data points used in the regression between energy production data and concurrent wind speed data (Figure 8).

Both the slope and intercept errors (Equations 5 and 6), from which the linear regression uncertainty depends (as described in Section 2.3), are inversely proportional to the number of data points, so that when a regression is performed on few data points, its uncertainty increases. This dependence is exemplified in Figure 4, where we have compared the sampling sets of regression lines for two stations in the EIA data set: for these two cases, the standard errors of regression slope and intercept for the station with 8 data points (on the right) are 30-50 times larger than what is found for the station with 90 data points (on the left).

The number of data points used for the regression has also an impact on the reference wind speed data uncertainty. In fact, short periods of wind plant operation record can lead to different interpretations from the reference wind resource data sets used as to whether that short period of record was above, equal to, or below the long-term average resource. Over a longer period of record, these potential discrepancies between different wind resource data sets (in our case, reanalysis products) tend to average out, therefore leading to a reduced uncertainty. We illustrate this phenomenon by exploring the long-term trend of the reanalysis products for the wind farm with one of the highest reported reference wind speed data uncertainties (EIA ID 60502,

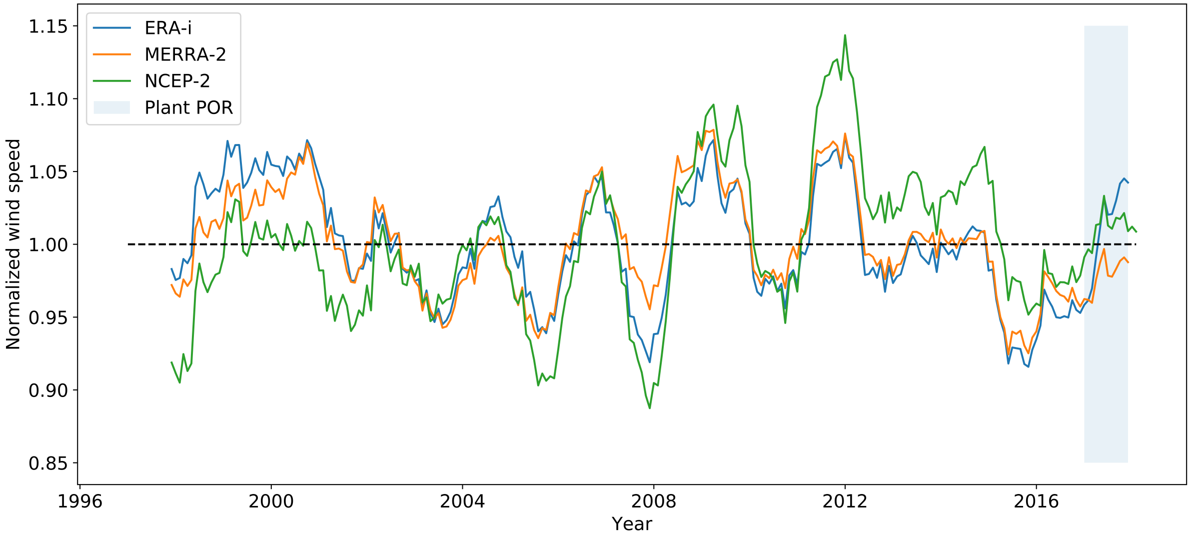

Figure 9. Long-term time series of normalized wind speed for EIA station ID 60502 from the three reanalysis products used in the study. The period of record (POR) for the wind farm is highlighted in light blue.

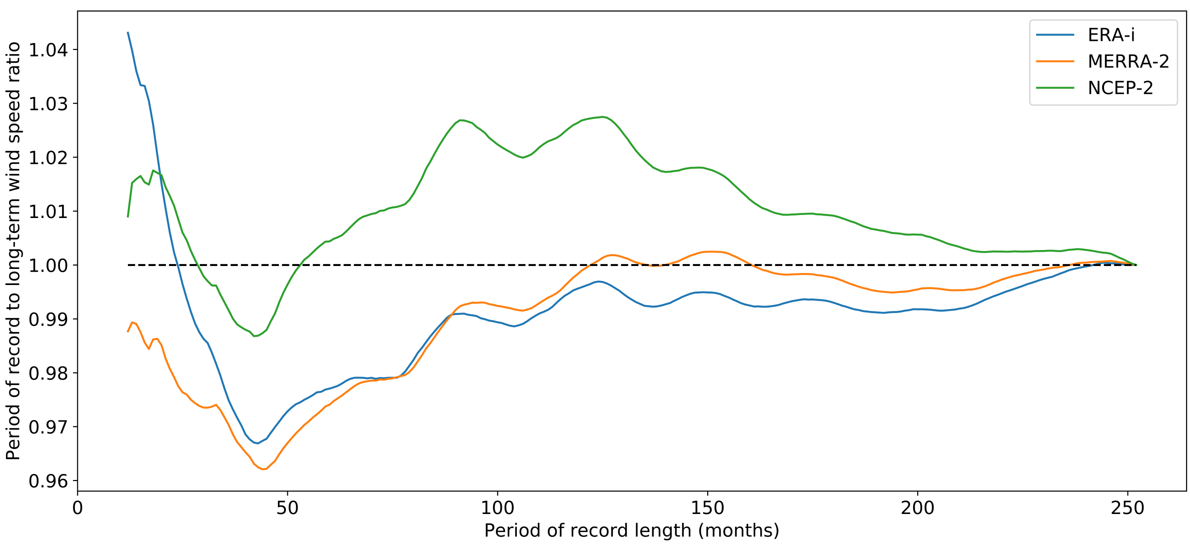

Figure 10. Ratio of wind speed to the long-term, 20-year average for periods of record of different lengths (all ending in December 2017), for EIA station ID 60502 using data from the three reanalysis products in the study.

reported 3.7% reference wind speed data uncertainty). Figure 9 shows the result. The period of record for wind farm operation (shown by a shaded blue area in Figure 9) was only 12 months. As shown in the figure, the various reanalysis products have very different interpretations of the wind resource in the short period of record, relative to the long-term (ERA-i: 4% above

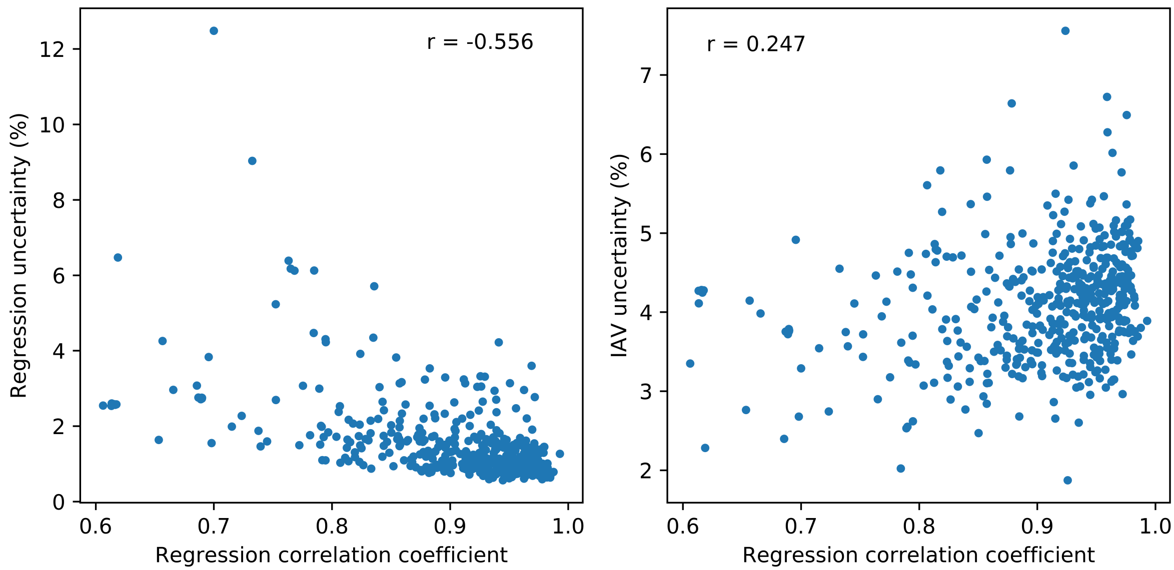

Figure 11. Dependence of linear regression uncertainty and IAV uncertainty on the R^2 of the regression between reanalysis wind speed and energy production data

average, MERRA-2: 1% below average; NCEP-2: 1% above average). Consequently, the use of each reanalysis product will lead to different magnitudes (both positive and negative) in the long-term windiness corrections, leading to high uncertainty in the resulting operational AEP calculation. By increasing the period of record (i.e., increasing the number of data points used in the regression), such discrepancies tend to average out. This is illustrated in Figure 10, where we show how the period of record to long-term wind speed ratio varies as we extend the period of record by increasing the number of months while keeping December 2017 as fixed ending time. For short periods of record, there is considerable deviation of this ratio among the different reanalysis products (i.e., the reference wind speed data uncertainty is high). As the length of the period of record increases, this ratio tends to converge to 1.0, and the spread between the three reanalysis products decreases (i.e., the reference wind speed data uncertainty is low).

Finally, the (weak) negative correlation between linear regression and wind resource IAV uncertainties is linked to the fact they respond differently to the R^2 coefficient between the reanalysis wind speed and the energy production data (Figure 11). Predictably, the linear regression uncertainty is inversely proportional to the coefficient of determination because a stronger correlation between winds and energy production will lead to a reduced uncertainty of the regression between the two variables. On the other hand, wind resource IAV uncertainty shows a direct correlation with the regression R^2 coefficient. This dependency can be explained as both quantities are directly correlated with the total variance of wind speed or, equivalently, produced energy. Figure 12 shows the relationship between IAV uncertainty and the total sum of squares $SS_{\text{tot, WS}}$ of reanalysis

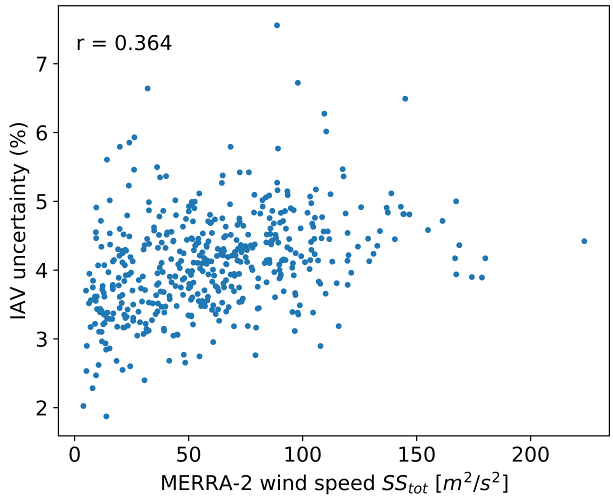

Figure 12. Dependence of IAV uncertainty on the total sum of squares $SS_{\text{tot, WS}}$ of MERRA-2 wind speed data, for the 472 projects considered.

wind speed (here, using MERRA-2 monthly data), which is proportional to the variance of the data:

$$SS_{\text{tot, WS}} = \sum_i (WS_i - \overline{WS})^2 \quad (8)$$

A direct correlation between IAV uncertainty and $SS_{\text{tot, WS}}$ emerges. At the same time, the linear regression R^2 coefficient also depends on the variance of the produced energy (and, equivalently, of wind speed) as it is defined as

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} \quad (9)$$

where SS_{res} is the total sum of the residuals from the linear regression. Equation 9 shows that when the total sum of squares SS_{tot} increases, so does R^2, thus confirming the direct correlation between R^2 and the variance in the data.

Finally, we note that although the sites selected for this analysis are primarily in simple terrain (Figure 1), we do not expect more complex topography to impact the correlations revealed from the Monte Carlo analysis, as all the underlying relationships would also be applicable to more complex sites.

4 Conclusions

Financial operations related to wind farms require accurate calculations of the annual energy production (AEP) and its uncertainty prior to the construction of the plant and, often, in the context of its operational analysis. As wind energy penetration increases globally, the need for techniques to accurately assess AEP uncertainty is a priority for the wind energy industry. Typically, current industry practice assumes that uncertainty components in AEP estimates are uncorrelated. However, we have shown that this assumption is not valid for the six components which comprise an operational-based uncertainty, using the EIA data set.

In this study we used a Monte Carlo approach to assess annual energy production; this not only accounts for correlations between uncertainty components, but also provides quantitative insight into aspects of the AEP calculation that drive its uncertainty. We have applied this approach using operational data from 472 wind farms across the United States in the EIA-923 database.

Our results show that ignoring correlations between uncertainty components causes a mean absolute difference of 6% compared to the uncertainty calculated with the Monte-Carlo-based approach, with larger deviations (up to 20%) for specific sites. Moreover, three pairs of uncertainty components reveal a statistically significant correlation: wind resource inter-annual variability (IAV) and long-term windiness correction (positive correlation); wind resource IAV and linear regression (negative); and reference wind speed data and linear regression (positive). Wind resource IAV and long-term windiness correction uncertainties are correlated because they both depend on wind resource variability. Wind resource IAV uncertainty is correlated with linear regression uncertainty because they are both inversely proportional to the number of data points in the period of record. Finally, reference wind speed data uncertainty and linear regression uncertainty show a negative correlation because they respond oppositely to the R^2 coefficient between the (reanalysis) wind speed and energy production data. Therefore, our results suggest that a Monte Carlo approach should be preferred to take into account these correlations between uncertainty components to lead to more accurate results, compared to the current industry standard approach. For all the projects considered in this study, the Monte Carlo simulation reached convergence within 10,000 runs. To facilitate the transition towards this proposed new industry standard, NREL's open-source OpenOA software[3] already supports the recommended Monte Carlo approach to assess AEP. In addition, the benefit of this technique will be further described in a guideline document in preparation for publication by the AWEA TR-1 working group.

Additional components of uncertainty in an operational AEP were not considered in our study because of limited reporting in the EIA-923 database. These components include reported availability, curtailment uncertainty, and various uncertainties introduced through analyst decision-making (e.g., filtering high-loss months from analysis and regression outlier detection). Future studies could include the impact of these additional sources of uncertainty on the operational AEP assessment. Finally, this study focused on correlations between operational AEP uncertainty components. Future work could explore correlations between pre-construction AEP uncertainty components. Given the numerous components (e.g., wake loss, wind speed extrapolation, wind flow model) and their intercomplexities, a Monte Carlo approach could reveal correlations that are at present not considered.

Code and data availability. EIA data used in this study are accessible from https://www.eia.gov/electricity/data/eia923/. Geographical data of the EIA wind farms are available at https://www.eia.gov/maps/layer_info-m.php. Software used to assess operational AEP is available from https://github.com/NREL/OpenOA.

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

Author contributions. NB and MO are equal contributors to this work. MO performed the AEP estimates on the wind farms considered in the study. NB and MO analyzed the processed data. NB wrote the manuscript, with significant contributions by MO.

Competing interests. The authors declare that they have no conflict of interest.

References

ANSI C12.1-2014: Electric Meters - Code For Electricity Metering, Standard, National Electrical Manufacturers Association, 2014.

Brower, M.: Wind resource assessment: a practical guide to developing a wind project, John Wiley & Sons, Hoboken, New Jersey, https://doi.org/10.1002/9781118249864, 2012.

Cameron, J.: Post-construction Yield Analysis, http://www.ewea.org/events/workshops/wp-content/uploads/proceedings/Analysis_of_Operating_Wind_farms/EWEA%20Workshop%20Lyon%20-%202-3%20Jessica%20Cameron%20Natural%20Power.pdf, European Wind Energy Association Technical Workshop, 2012.

Clifton, A., Smith, A., and Field, M.: Wind Plant Preconstruction Energy Estimates: Current Practice and Opportunities, Tech. rep., https://www.nrel.gov/docs/fy16osti/64735.pdf, 2016.

Dee, D. P., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d. P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the royal meteorological society, 137, 553–597, 2011.

Dimitrov, N. K., Kelly, M. C., Vignaroli, A., and Berg, J.: From wind to loads: wind turbine site-specific load estimation with surrogate models trained on high-fidelity load databases, Wind Energy Science, 3, 767–790, 2018.

EIA: A Guide to EIA Electric Power Data, Standard, Energy Information Administration, https://www.eia.gov/electricity/data/guide/pdf/guide.pdf, 2018.

Energy Information Administration: Monthly Energy Review - March 2020, Tech. rep., US Department of Energy, 2020.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., et al.: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), Journal of Climate, 30, 5419–5454, 2017.

Global Wind Energy Council: Global Wind Report - Annual Market Update 2017, Tech. rep., Global Wind Energy Council, 2018.

IEC 60688:2012: Electrical measuring transducers for converting A.C. and D.C. electrical quantities to analogue or digital signals, Standard, International Electrotechnical Commission, 2012.

IEC 61400-12-1:2017: Wind energy generation systems - Part 12-1: Power performance measurements of electricity producing wind turbines, Standard, International Electrotechnical Commission, 2017.

IEC 61400-15:draft: Assessment of site specific wind conditions for wind power stations, Standard, International Electrotechnical Commission, in draft.

IEC 61400-26-3:2016: Wind energy generation systems - Part 26-3: Availability for wind power stations, Standard, International Electrotechnical Commission, 2016.

Ishihara, T. and Yamaguchi, A.: Prediction of the extreme wind speed in the mixed climate region by using Monte Carlo simulation and measure-correlate-predict method, Wind Energy, 18, 171–186, 2015.

ISO, I. and OIML, B.: Guide to the Expression of Uncertainty in Measurement, Geneva, Switzerland, 122, 1995.

Kalkan, A.: Uncertainty in Wind Energy Assessment, http://www.windsim.com/documentation/UM2015/1506_WindSim_UM_Inores_Akgun_Kalkan.pdf, 2015.

Khatab, A. M.: Performance Analysis of Operating Wind Farms, Master's thesis, Uppsala University, Department of Earth Sciences, Campus Gotland, 2017.

Lackner, M., Rogers, A., and Manwell, J.: Uncertainty Analysis in Wind Resource Assessment and Wind Energy Production Estimation, https://doi.org/10.2514/6.2007-1222, https://arc.aiaa.org/doi/abs/10.2514/6.2007-1222, 2007.

Lindvall, J., Hansson, J., Undheim, O., and Vindteknikk, J.: Post-construction production assessment of wind farms, Tech. Rep. 2016:297, Energyforsk, 2016.

Lunacek, M., Fields, M. J., Craig, A., Lee, J. C. Y., Meissner, J., Philips, C., Sheng, S., and King, R.: Understanding Biases in Pre-Construction Estimates, Journal of Physics: Conference Series, 1037, 062 009, http://stacks.iop.org/1742-6596/1037/i=6/a=062009, 2018.

Manwell, J. F., McGowan, J. G., and Rogers, A. L.: Wind energy explained: theory, design and application, John Wiley & Sons, 2010.

Müller, K. and Cheng, P. W.: Application of a Monte Carlo procedure for probabilistic fatigue design of floating offshore wind turbines, 2018.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-y., Iredell, M., et al.: The NCEP climate forecast system version 2, Journal of Climate, 27, 2185–2208, 2014.

Shravan Kumar, M. and Anandan, V.: Comparision of the NCEP/NCAR Reanalysis II winds with those observed over a complex terrain in lower atmospheric boundary layer, Geophysical research letters, 36, 2009.

Vaisala: Reducing Uncertainty in Wind Project Energy Estimates, Tech. rep., https://www.vaisala.com/sites/default/files/documents/Triton-DNV-White-Paper.pdf, 2014.

Westfall, P. H. and Young, S. S.: Resampling-based multiple testing: Examples and methods for p-value adjustment, vol. 279, John Wiley & Sons, 1993.

Williams, S. K., Acker, T., Goldberg, M., and Greve, M.: Estimating the economic benefits of wind energy projects using Monte Carlo simulation with economic input/output analysis, Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology, 11, 397–414, 2008.